250 BOSTON STUDIES IN THE PHILOSOPHY OF SCIENCE

The Genesis of General Relativity
Edited by Jürgen Renn

Volume 4

Gravitation in the Twilight of Classical Physics

The Promise of Mathematics

Edited by
Jürgen Renn and Matthias Schemmel

Springer

The Genesis of General Relativity

BOSTON STUDIES IN THE PHILOSOPHY OF SCIENCE

VOLUME 250

The Genesis of General Relativity
Edited by Jürgen Renn

Volume 1

EINSTEIN'S ZURICH NOTEBOOK: INTRODUCTION AND SOURCE

Michel Janssen
University of Minnesota, U.S.A.

John D. Norton
University of Pittsburgh, U.S.A.

Jürgen Renn
Max Planck Institute for the History of Science, Germany

Tilman Sauer
Einstein Papers Project, Caltech, U.S.A.

John Stachel
Boston University, U.S.A.

and

Assistant Editor

Lindy Divarci
Max Planck Institute for the History of Science, Germany

Springer

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN-10 1-4020-3999-9 (HB)
ISBN-13 978-1-4020-3999-7 (HB)
ISBN-10 1-4020-4000-8 (e-book)
ISBN-13 978-1-4020-4000-9 (e-book)

As a complete set for the 4 volumes

Published by Springer,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

www.springer.com

Printed on acid-free paper

TABLE OF CONTENTS

Volume 1

Volume 2

(parallel volume)

JÜRGEN RENN

PREFACE

The transition from classical to modern physics in the first half of the twentieth century by quantum and relativity theories affected some of the most fundamental notions of physical thinking, such as matter, radiation, space, and time. This transition thus represents a challenge for any attempt to understand the structures of a scientific revolution. The present four-volume work aims at a comprehensive account of the way in which the work of Albert Einstein and his contemporaries changed our understanding of space, time, and gravitation. The conceptual framework of classical nineteenth-century physics had to be fundamentally restructured and reinterpreted in order to arrive at a theory of gravitation compatible with the new notions of space and time established in 1905 by Einstein's special theory of relativity.

Whereas the classical theory of gravitation postulated an instantaneous action at a distance, Einstein's new relativistic kinematics rather suggested an analogy between the gravitational field and the electromagnetic field, propagating with a finite speed. It is therefore not surprising that Einstein was not alone in addressing the problem of formulating a theory of gravitation that complies with the kinematics of relativity theory. The analysis of these alternative approaches, as well as of earlier alternative approaches to gravitation within classical physics, turns out to be crucial for identifying the necessities and contingencies in the actual historical development.

It is the profound conceptual transformation associated with the establishment in 1915 of a relativistic theory of gravitation that shows that the genesis of this theory, Einstein's general theory of relativity, was a genuine scientific revolution in its own right. The restructuring and reinterpretation of the fundamental concepts of classical physics involved in the development of the theory was a long and complicated process with far-reaching consequences. First of all, the new concepts had to be presented and transmitted to a wider scientific community. The new theory also created the need to reconsider all branches of physics in light of its new concepts and to look for possible experimental confirmation. It gave rise to new conceptual problems and new research programs, such as finding a unified field theory and integrating general relativity with quantum theory. Finally, the revision of the concepts of space and time by general relativity had a considerable impact on epistemological and philosophical discussions, attracting the attention of a non-specialized public and placing relativity theory and its creator at the focus of public discussion.

Einstein's path towards establishing the general theory of relativity has been an important topic in the history of twentieth-century physics. Our reconstruction of his

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 1
Einstein's Zurich Notebook: *Introduction and Source*.

unpublished research notes and our examination of the broader intellectual context of the relativity revolution has led to a reassessment and a deeper understanding of this process as a transformation of a comprehensive system of knowledge. These volumes document the results of a joint effort at an in-depth analysis of a scientific revolution undertaken by a group of scholars over more than a decade. The aim was to reach a systematic understanding of both the knowledge base in classical physics that formed the point of departure for Einstein and his contemporaries and the nature of the process through which their research eventually overcame some of the conceptual foundations of classical as well as special-relativistic physics.

For this purpose, it was necessary to cover not only Einstein's individual pathway towards general relativity, but also other approaches to the problem of gravitation before and after the advent of special relativity. The aim was to reach an assessment of the "horizon of possibilities" of reacting to the crisis provoked by the conflict between the understanding of gravitation in classical physics and the challenge presented by the special theory of relativity. The horizon of possibilities is determined by the shared knowledge available to the historical actors. The reconstruction of this shared knowledge and its transformation is based on new approaches for describing the architecture of knowledge and for explaining its developmental dynamics, including the interaction between collective and individual processes.

We have thus attempted to provide a broader context to the reconstruction of Einstein's singular achievement. We surveyed the approaches to the problem of gravitation that existed in late classical physics, to examine the intellectual resources on which the different approaches relied, to determine the extent to which they were adequate to the task of responding to the crisis of classical physics, to explore alternative pathways that could have been but were not realized, and finally to evaluate the reasons why Einstein's general relativity eventually came to be accepted as the resolution of the crisis. The results of our reconstruction are documented in the form of detailed commentaries on the historical sources and in the form of new interpretations of the early history of general relativity.

The four volumes of this work comprise two sets. The first two volumes are dedicated to general relativity in the making, that is, to a detailed reconstruction of the research that led Einstein from special to general relativity in the years between 1907 and 1915. At the center of this reconstruction is the detailed "Commentary" on a key document written between 1912 and 1913, Einstein's so-called "Zurich Notebook." In the first volume this notebook is presented in its entirety for the first time. It is reproduced in facsimile accompanied by a new transcription. The second volume presents the comprehensive "Commentary" so that the reader may directly relate interpretation and historical source. The first two volumes furthermore comprise essays on the development leading up to the period documented in the notebook, assessments of the work documented by the notebook itself, and an analysis of the conclusive period of Einstein's search for the gravitational field equation. Taken together, the work assembled in these two volumes offers an encompassing view of Einstein's contributions to the genesis of general relativity.

The second set of two volumes is dedicated to theories of gravitation in the twilight of classical physics in a more general sense. In this part of the work, alternative approaches to the problem of gravitation around the time of Einstein's work are reviewed in terms of interpretative essays and English translations of key sources. The third volume deals with the tensions between the tradition of mechanics, the canonical place of the problem of gravitation, and the newly established tradition of field theory that raised expectations for a novel solution to this problem. These expectations were then strengthened by the advent of special relativity and led to an intense discussion about a relativistic theory of gravitation that forms the other nucleus of this volume. The fourth volume takes a closer look at possibilities for the establishment of a theory like general relativity along pathways that differed from Einstein's in that they employed more sophisticated mathematical means. The volume thus covers both a reassessment of David Hilbert's work and the suggestion of a fictive but historically plausible scenario for such an achievement.

The work presented in these volumes was originally pursued in the context of the *Arbeitsstelle Albert Einstein*, directed by Peter Damerow and myself, funded by the Senate of Berlin from 1991 to 1996, and hosted by the Max Planck Institute for Human Development and Education, at the Center for Development and Socialization headed by Wolfgang Edelstein. I am deeply grateful to the Berlin Senate, in particular to the former Senator of Science, Barbara Riedmüller, as well as to the former Senate Director, Jochen Stöhr, for the courageous and generous decision to support this unusual initiative, which aimed at exploring Einstein's scientific achievements in their intellectual, cultural, and political contexts. Under the auspices of Wolfgang Edelstein it has served in many ways as a pioneering venture for the foundation in 1994 of the Max Planck Institute for the History of Science. In addition to the collaborators of the *Arbeitsstelle*, Giuseppe Castagnetti, Werner Heinrich, and Tilman Sauer, members of its international scientific advisory board, Hubert Goenner, Michel Janssen, Karl von Meyenn, John D. Norton, Karin Reich, Erhard Scholz, and John Stachel, participated in one way or another in the research process – either by providing helpful comments or by engaging directly in joint projects.

In this way, the *Arbeitsstelle Albert Einstein* soon developed into a meeting point for an international group of scholars working on the history of general relativity and the locus of an unusual cooperation involving both senior experts in the field and young researchers, which continued later at the Max Planck Institute for the History of Science. These meetings involved the core group of authors of the first two volumes as well as the other members of the *Arbeitsstelle* and members of its board. In addition to the names already mentioned, Dieter Brill, Ulrich Majer, James Ritter, David Rowe, Matthias Schemmel, and Dirk Wintergrün also contributed at some point to our co-operation. The innumerable meetings and workshops were unique in their collaborative search for an interpretation of historical sources, with key ideas emerging from lively debate. In a lengthy process, the detailed protocols of these meetings were filtered, reworked, elaborated and reformulated to yield the "Commentary" which constitutes the most distinct outcome of the collective work. In this way, the core group worked together to analyze the sources and reconstruct the knowledge resources that are relevant for understanding the research documented in Einstein's notebooks, his publications, and correspondence. Following ground-breaking papers by John Stachel, John D. Norton, and a few other scholars, the continued investigation and reconstruction of Einstein's discovery process has led to many new insights, among them the identification of two distinct heuristic strategies in this discovery process, a physical and a mathematical strategy. This identification proved to be a breakthrough and an important interpretative tool for understanding Einstein's search for the gravitational field equation, even beyond the phase documented by the Zurich notebook. Our analysis has shed new light on the complex process of interaction between mathematical representation and the construction of physical meaning, a process of crucial importance also in other areas and periods of the history of science. Such epistemological insights were only possible because our joint work was not confined to a painstaking analysis of the historical sources in a traditional sense, but also comprised unusual approaches such as reconstructing the architecture of the shared knowledge at his disposal and actually retracing in detail Einstein's research process in a particular phase of his work. For the epistemological dimension of our discussions and for wider perspectives, the intense participation of Peter Damerow in our research endeavor turned out to be critical. He helped us wherever he could from falling into the traps of specialization and placed our work within the larger framework of a history of knowledge.

As work progressed, it quickly became clear that the clarification reached by deciphering Einstein's research notes from the period 1912–1913 would have serious consequences for our understanding the genesis of general relativity in its entirety. The Zurich Notebook shows that in 1912–1913 Einstein had already come within a hair's breadth of the final general theory of relativity. He failed, however, to recognize the physical meaning of his mathematical results, and turned to the alternative, physical strategy. Eventually he published, jointly with the mathematician Marcel Grossmann, the "erroneous" *Entwurf* ("outline") theory of 1913. Much of our work therefore focused on the question of how Einstein, in the period between 1913 and 1915, was able to overcome the obstacles which at first prevented him from realizing that the correct ansatz was the one obtained in his notebook and not the theory he published in 1913. The answer we found to this question led to the surprising insight that, contrary to what was commonly accepted, the long interval between the publication of the erroneous field equation and the return to the correct equation at the end of 1915 was not simply a period of stagnation. It was rather a period during which Einstein arrived at a number of insights that created the crucial preconditions that made the dramatic events of November 1915 possible. This result made it evident that the establishment and stabilization of the new physical concepts that emerged with general relativity first required an integration of further physical knowledge and a degree of elaboration of the mathematical formalism that went well beyond finding the correct field equation.

Another complementary line of research was, as mentioned above, dedicated to the study of other theories of gravitation before and after the advent of special relativity. In order to identify knowledge traditions that contributed to the emergence of general relativity, the scientific context of Einstein's search for a new theory of gravitation was systematically studied by analyzing a broad range of sources related to the work on alternative approaches, including also the work of less well-known authors.

I am particularly grateful to Michel Janssen who throughout the intricate research and production process leading to these volumes never hesitated to take up whatever challenges arose. He took the main responsibility for bringing the "Commentary," the core of the first two volumes, into the form it is presented here and also helped with critical acumen to sharpen the focus and improve the presentation of other contributions. Matthias Schemmel, who co-edited volumes three and four, and Lindy Divarci, assistant editor of all four volumes, played an essential role in coordinating this extended network of scholarly cooperation. Together with the associate editors, Christopher Smeenk and Christopher Martin, they carefully edited the various contributions, unified and improved the translations of original sources, checked and complemented the bibliographic references, and contributed in many other ways to providing a comprehensive resource for studying the early history of general relativity in context. They were assisted in their editorial work by Heinz Reddner, Stefan Hajduk, Yoonsuhn Chung, Miriam Gabriel, and Shaul Katzir.

The long-term cooperation required to produce the comprehensive analysis of the genesis of general relativity presented in this work was only possible due to the persistent institutional support provided by the Max Planck Society, first at the Max Planck Institute for Human Development and Education and, since its foundation in 1994, at the Max Planck Institute for the History of Science; additional support came from the Archive of the Max Planck Society and its director, Eckhart Henning. The close and ever reliable cooperation with other institutions, in particular with the Albert Einstein Archives at the Hebrew University of Jerusalem and its former Curator Ze'ev Rosenkranz and the Einstein Papers Project at the California Institute of Technology and its former director Robert Schulmann as well as its current director Diana Buchwald, were of great help in the completion of this ambitious project. The authors and editors are particularly grateful for the generous permission to reproduce or quote from Einstein's original documents. Other institutions and individuals offered their generous support as well, either in securing the documentary basis for our enterprise or helping in other ways, among them the Albert Einstein Institute (Max Planck Institute for Gravitational Physics), the Cohn Institute for the History and Philosophy of Science and Ideas at Tel Aviv University, the Institut für Zeitungsforschung, Dortmund, the Manuscript Department of the Staatsbibliothek zu Berlin, the Special Collections Department at the Library of the of the Swiss Federal Institute of Technology Zurich (ETH), the Niedersächsische Staats- und Universitätsbibliothek Göttingen (Manuscripts and Early Imprints Collection) and its director Helmut Rohlfing, the Mathematical Institute at the University of Göttingen, the Federal Archives in Koblenz, the National Science Foundation, the Istituto e Museo di

Storia della Scienza, Florence, and the Besso family (especially Laurent Besso, Lausanne). I also want to thank Jürgen Ehlers, Yehuda Elkana, Paolo Galluzzi, Peter Galison, Gerald Holton, Enrique Junowicz, Ron Overman, and Bernhard Schutz. A special thanks goes to the head of the library at the Max Planck Institute for the History of Science, Urs Schoepflin, who not only assured his team's unfailing support during the many years of work on these volumes, but who was also personally engaged in archival work, in sustaining the scholarly network, and in securing key documents on which these volumes are based.

The volumes appear in sequel to the International Year of Physics and the Einstein Year 2005, celebrating the centenary of Einstein's *annus mirabilis* and the overturn of the classical concepts of space and time. They propose an in-depth historical analysis of the consequences of this revolution for our understanding of gravity and, at the same time, of the structures of a scientific revolution that can be documented in an exceptionally comprehensive way. Yet the scope of the scientific revolution associated with Einstein's name goes well beyond the intellectual work on a new theory of gravity. Its full understanding also requires an analysis of other aspects such as Einstein's contributions to quantum theory or the role of cultural, technological, personal, and political contexts that could only be touched upon in these volumes. They are more amply treated in other publications, among them the three-volume survey *Albert Einstein – Engineer of the Universe* of Wiley-VCH associated with the exhibition of the same title, the *Einstein Companion* of Cambridge University Press, as well as a forthcoming book on the institutional contexts of the emergence of quantum theory to appear in this series, all of them emerging from the same context of collaboration that has made the present work possible. As research on Einstein's revolution of science is still in progress, parts of these volumes are, together with additional sources and interpretative documents and tools, accessible also via the Internet <http://einstein-virtuell.mpiwg-berlin.mpg.de/intro>.

MICHEL JANSSEN, JOHN D. NORTON, JÜRGEN RENN,
TILMAN SAUER, AND JOHN STACHEL

INTRODUCTION TO VOLUMES 1 AND 2: THE ZURICH NOTEBOOK AND THE GENESIS OF GENERAL RELATIVITY

When Albert Einstein died in April 1955, he left a small notebook among his many papers at the Institute for Advanced Study in Princeton. Its faintly gridded pages are covered with calculations. Some are tidy and unhurried. Others are hasty and incomplete. Some are annotated with a cryptic remark; others are unadorned. Some halt with a fragmented formula; others proceed mechanically to their conclusion. They come from another time and place, a silent trace of strenuous work from decades earlier and a continent away.[1]

Most of the calculations in this notebook date from the winter of 1912–1913. In August 1912 Einstein had left Prague, where he had taught for a year and a half, to become a full professor at his *alma mater*, the *Eidgenössische Technische Hochschule* (ETH) in Zurich. This is why the notebook is known among Einstein scholars as the Zurich Notebook. The bulk of it is devoted to a new theory of gravity, in which the ten components $g_{\mu\nu}$ of the metric tensor field encode the geometry of spacetime and double as the potentials of the gravitational field. Many of the end results of the investigations recorded in the notebook were published in the spring of 1913 in a paper Einstein co-authored with Marcel Grossmann, professor of mathematics at the ETH and one of his former classmates. The paper consists of two parts, a physical part written by Einstein, and a mathematical part written by Grossmann. The title modestly announces an "Outline [*Entwurf*] of a Generalized Theory of Relativity and a Theory of Gravitation" (Einstein and Grossmann 1913). Einstein continued to work on this *Entwurf* theory, as it is generally known in the historical literature, for the next couple of years, initially with Grossmann (see Einstein and Grossmann 1914b) and with Michele Besso, another friend from his college days. With Besso he investigated whether the new theory could account for the anomalous advance of the perihelion of Mercury, a well-known problem in Newtonian gravitational theory. They found that it

1 In 1982, the notebook, together with Einstein's other papers, was shipped from Princeton to Jerusalem, where it is now part of the Einstein Archives at Hebrew University (CPAE 1, Publisher's Foreword). Its call number is 3-006. A high-quality scan of the notebook is available electronically at the Einstein Archives Online (www.alberteinstein.info).

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 1
Einstein's Zurich Notebook: *Introduction and Source*.

could not.[2] These collaborations ceased in the spring of 1914, when Einstein moved to Berlin to become a salaried member of the *Preußische Akademie der Wissenschaften*. In the fall of 1914, he published a lengthy article intended as the authoritative, systematic exposition of the new theory. Its title, "The Formal Foundation of the General Theory of Relativity" (Einstein 1914), stands in marked contrast to the tentative title of the original Einstein-Grossmann paper. A year later, however, Einstein's confidence crumbled. In a series of four communications to the Prussian Academy in November 1915, he replaced the centerpiece of the *Entwurf* theory, a set of gravitational field equations of severely restricted covariance, by field equations of broad and ultimately general covariance, solving the problem of Mercury's perihelion in the process (Einstein 1915a, 1915b, 1915c, 1915d).[3] Einstein had thus arrived at the general theory of relativity, the crowning achievement of his career. He consolidated the theory over the next few years. In March 1916, he replaced the premature review article of 1914 by a detailed and self-contained exposition of the new theory (Einstein 1916a). He subsequently applied general relativity to new problems—such as gravitational waves (Einstein 1916b, 1918a) and cosmology (Einstein 1917)—and clarified its foundations (Einstein 1916c, 1917, 1918b, 1918c).[4]

As our joint commentary on the notebook makes clear, the material in the Zurich Notebook holds the key to understanding many aspects of these later developments. To facilitate reading the commentary in conjunction with the notebook itself, we placed the commentary at the beginning of volume two and a facsimile reproduction and a transcription of the notebook at the end of volume one.[5] A number of essays, which make up the balance of the volumes, fill in the background to the research documented in the notebook and address the ramifications of our analysis of the notebook for the reconstruction of the further development of the *Entwurf* theory and the transition to the theory of 1915. These essays are written in such a way that they can be read independently of the commentary and of each other.

The analysis of the notebook was quite a challenge. The notebook consists of working notes and was never intended to be read by others. Its one intended reader needed no narrative to explain the goals and presuppositions of the calculations, their successes and failures, the puzzlements and the triumphs. These would have been all too apparent to Einstein's eyes. They were not to ours. Our commentary reflects our best effort to understand Einstein's calculations and to supply some of the connective

2 See CPAE 4, Doc. 14 and the editorial note, "The Einstein-Besso Manuscript on the Motion of the Perihelion of Mercury," on pp. 344–359. For a popular account, see (Janssen 2003). For a history of the perihelion problem, see (Roseveare 1982).

3 The demise of the *Entwurf* theory and the subsequent developments are detailed in Einstein to Arnold Sommerfeld, 28 November 1915 (CPAE 8, Doc. 153). For analysis of (Einstein 1915c) on the perihelion problem, see (Earman and Janssen 1993).

4 For a concise history of Einstein's struggles with the conceptual basis of his theory and further references to the extensive literature on these topics, see (Janssen 2005).

5 Although the commentary only covers the notes on gravity, facsimiles and transcriptions of notes on other topics are also included. Insights from the analysis of the non-gravitational part of the Zurich Notebook were used in (Büttner et al. 2003).

tissue that he left out. We have reconstructed, as best we can, the content, goals and strategies of the calculations, usually on a line-by-line basis, sometimes even symbol by symbol.

Such efforts cannot hope to recover in full detail what transpired as Einstein made these entries in his notebook. Yet the relative completeness of many of the calculations and the apparent clarity of purpose repeatedly allowed us to discern simple and coherent content in pages that initially looked baffling and disjointed. In another context, Einstein often remarked on a problem not altogether different from the one we faced. Physical theories, he remarked, cannot be deduced from sensory experience. There remains considerable freedom of choice in the concepts and propositions one devises to account for experience. What makes stable theorizing possible is the restricted character of this freedom:

> The liberty of choice, however, is of a special kind; it is not in any way similar to the liberty of a writer of fiction. Rather, it is similar to that of a man engaged in solving a well-designed [cross]word puzzle. He may, it is true, propose any word as the solution; but, there is only *one* word which really solves the puzzle in all its parts (Einstein 1936, 294–295).[6]

Our experience with the Zurich Notebook has been similar. For any given page, one can propose many interpretations. However, when the calculation is relatively complete, and when it connects naturally with other pages, the majority of interpretations fail to solve the puzzle.[7] As readers of the commentary will find, our solution is incomplete in places. It is, however, much more complete than any of us dared hope at the outset. The portions that remain obscure are much smaller than those we now read with clarity. The notebook is open to a new readership.

We should make it clear that we did not have to start from scratch. Norton (1984) had already deciphered several key pages of the notebook before we joined forces. More generally, we drew on research done at the *Einstein Papers Project*[8] and on various studies of the history of general relativity. Many of these were presented at the *History of General Relativity* (HGR) conference series and published in several volumes of the series *Einstein Studies*.[9] Both series were inaugurated by the Nestor of our group, Stachel, who is also the founding editor of the Einstein edition. The anno-

---

6 The handwritten German original has: “Mit dieser Freiheit ist es aber nicht weit her; sie ist nicht ähnlich der Freiheit eines Novellen-Dichters sondern vielmehr der Freiheit eines Menschen, dem ein gut gestelltes Worträtsel aufgegeben ist. Er kann zwar jedes Wort als Lösung vorschlagen, aber es ist wohl nur eines welches das Rätsel in allen Teilen wirklich auflöst” (Einstein Archives, 122–858).

7 In the course of our collaboration, we hit upon what we have dubbed the “chicken scratch rule”: to inspire confidence, a proposed reconstruction should account for every last scratch on the page.

8 All five of us were involved in one way or another with the publication of the relevant Vols. 4 through 8 of *The Collected Papers of Albert Einstein* (CPAE), which appeared between 1993 and 2002.

9 (Howard and Stachel 1989) for HGR1 at Osgood Hill (1986) [Don Howard and Stachel are also the series editors for *Einstein Studies*]; (Eisenstaedt and Kox 1991) for HGR2 in Luminy (1988); (Earman, Janssen, and Norton 1993) for HGR3 in Johnstown (1991); (Goenner, Renn, Ritter, and Sauer 1999) for HGR4 in Berlin (1995); and (Kox and Eisenstaedt 2005) for HGR5 at Notre Dame (1998) and HGR6 in Amsterdam (2002). A volume for HGR7 in Tenerife (2005) is in preparation.

tation of the gravitational part of the Zurich Notebook in CPAE 4, which appeared in 1995, incorporates elements of (Norton 1984) as well as some of the early results of the work of our group.[10] Our joint commentary, however, is the first comprehensive analysis of the gravitational part of the Zurich Notebook. As such, it provides new insights on pages that were not deciphered before and important correctives to the interpretation of pages that were.

Our most important new results are highlighted in sec. 1 of the commentary. Here we want to draw attention to a few that are crucial to understanding the essays complementing the commentary in these volumes. Most of the material in the notebook documents Einstein's search for field equations for his new metric theory of gravity. Two strategies can be discerned in this search. We have labeled them the 'mathematical strategy' and the 'physical strategy'. Each of us would probably flesh out the distinction between the two somewhat differently, but the operative notion is fairly straightforward. Pursuing the mathematical strategy, Einstein scoured the mathematical literature (with the help of Grossmann) for expressions containing derivatives of the metric that could be used as building blocks for gravitational field equations. Pursuing the physical strategy, Einstein tried to find such building blocks drawing on the analogy between the gravitational field in his new theory and the electromagnetic field in the classical electrodynamics of Maxwell and Lorentz. Einstein vacillated between these two approaches in the notebook. The importance of this simple observation for our reconstruction of the research recorded in the notebook can hardly be overstated. The same is true for the reconstruction of the subsequent elaboration of the *Entwurf* theory and of the transition to general relativity in November 1915. The distinction between the two approaches plays a key role in two essays in these volumes, "Pathways out of Classical Physics: Einstein's Search for the Gravitational Field Equations" (this volume) and "Untying the Knot: How Einstein Found His Way Back to Field Equations Discarded in the Zurich Notebook" (volume two).

Another new result that came out of our analysis of the notebook and that needs to be mentioned here is more subtle. One of the requirements constraining the choice of gravitational field equations for Einstein was that they reduce to the field equation of Newtonian theory in the case of weak static fields. Einstein was searching for field equations of broad and, if possible, general covariance, i.e., for equations that have the same form in a broad range of spacetime coordinate systems. Newtonian theory—at least in the standard formulation, which Einstein was using—is formulated in such a way that its equations only retain their simple form under transformations from one inertial frame to another. To compare the equations of broad covariance of Einstein's theory to the equations of limited covariance of Newton's, we nowadays temporarily relinquish some of the covariance of the former. This is done by impos-

10 See CPAE 4, Doc. 10 and the editorial note, "Einstein's Research Notes on a Generalized Theory of Relativity," on pp. 192–199. Preliminary results of the work of our group can be found in (Castagnetti et al. 1993), (Janssen 1999), (Norton 2000), (Renn 2005a, 2005b), and (Renn and Sauer 1996, 1999, 2003).

ing extra conditions on the metric field. Such extra conditions are known as coordinate conditions. Once a coordinate condition is imposed, various terms in the Einsteinian equations vanish and the resulting truncated equations can readily be compared to the simple Newtonian equation. As this brief characterization of the role of coordinate conditions shows, they are not essential to the theory. They have the same status as gauge conditions in other field theories. Their purpose is simply to facilitate comparison with Newtonian theory. For the application of the theory to other problems it may in fact be convenient to impose a different coordinate condition. In all of this, the equations of broad covariance remain the fundamental equations of the theory, not the truncated ones obtained with the help of some coordinate condition.

What we found in the Zurich Notebook, however, is that Einstein used such conditions in a manner that deviates sharply from modern usage. He used them to truncate equations of broad covariance and looked upon the resulting truncated equations as the fundamental equations of the theory or candidates for them. The original equations of broad covariance were important to him only in that they made the covariance properties of the truncated equations more tractable. Starting from equations covariant under a well-defined, broad group of transformations, one can, at least in principle, find the covariance group of the truncated equations by determining the covariance properties of the condition used to do the truncating. This is what we see Einstein do over and over again in the notebook. We therefore introduced a special term for such conditions. To distinguish them from modern coordinate conditions we call them *coordinate restrictions*. During the period covered by the notebook, Einstein labored under the impression that coordinate restrictions were needed not just to recover Newtonian theory in the appropriate limit, but also to ensure that the theory be compatible with energy-momentum conservation. It was only in the course of developing the *Entwurf* theory in 1913–1914 that he came to realize that the covariance of the field equations automatically yields energy-momentum conservation, thereby anticipating an important application of one of the famous theorems of Emmy Noether (1918) relating symmetries and conservation laws.[11] Recognizing the role of coordinate restrictions is crucial not only for reconstructing the calculations in the Zurich Notebook, but also for understanding the subsequent elaboration of the *Entwurf* theory and the obstacles that had to be overcome before general relativity as we know it today could be formulated.

The older literature (e.g., Pais 1982, 222) routinely explained Einstein's initial rejection of generally-covariant gravitational field equations by supposing a lack of understanding of the need to use coordinate conditions to compare such equations to their Newtonian counterpart in the case of weak gravitational fields. Then Stachel (1989b) pointed out that Einstein actually compared his *Entwurf* theory's field equations with the field equation of Newtonian theory in the case of weak, *static* fields and that Einstein had presumed an excessively restrictive form for the metric representing

11 For details, see "Untying the Knot …" (volume two).

such fields.[12] Einstein expected that the metric tensor for weak static fields would have just one variable component, which would behave like the Newtonian field. This one component would govern the chronometry only. Since the remaining components were assumed to be constant, the spatial geometry would just be ordinary Euclidean geometry. This expectation contradicts the generally-covariant field equations Einstein finally adopted in 1915 and could by itself preclude their adoption. The suspicion that this was the real reason for Einstein's rejection of generally-covariant field equations was reinforced when Norton (1984, 117) drew attention to a page of the Zurich Notebook, on which Einstein rehearsed the now standard mathematical computations needed to reduce generally-covariant field equations based on the so-called Ricci tensor to the Newtonian equation by means of what is known as the harmonic coordinate condition. Presuming that Einstein used this condition as a modern coordinate condition, Norton conjectured that Einstein had rejected generally-covariant field equations based on the Ricci tensor because his expectations for the weak, static field were incompatible with the harmonic coordinate condition.

Einstein's difficulties with the static metric left unexplained, however, why he subsequently abandoned another, apparently serviceable set of gravitational field equations of near general covariance examined in the Zurich Notebook. Our analysis of the notebook makes clear that these gravitational field equations were abandoned largely because Einstein was not then using modern coordinate conditions but coordinate restrictions. Over the next two years, he worked energetically to develop a better understanding of the covariance properties of his *Entwurf* theory. The more sophisticated mathematical methods resulting from these efforts are central in his return to broader covariance in November 1915. That return involved his adoption of modern coordinate conditions. He concluded that the gravitational field equations of near general covariance rejected in the Zurich Notebook were acceptable after all and rushed them into print (Einstein 1915a). Freed of coordinate restrictions, Einstein now needed only weeks to discover his mischaracterization of static fields. He made some final modifications to the field equations published in (Einstein 1915a) and thus arrived at the familiar generally-covariant field equations of general relativity.

Not surprisingly, given this turn of events, Norton's essay, "What was Einstein's "Fateful Prejudice"?," in volume two analyzes the role of coordinate restrictions in Einstein's work. The essay carefully lays out the case for their presence in the notebook. It also reviews why Einstein's use of them is credible, even though they differ so much from the routine modern usage. Two considerations are key here: First, Einstein's use of coordinate restrictions is not unnatural in view of the historical development of his theory. His special theory of relativity was covariant just under Lorentz transformations. His goal was to expand that covariance to embrace transformations to accelerated frames of reference. General covariance goes well beyond this, includ-

12 In (Stachel forthcoming) it is shown that Einstein was seriously handicapped in the comparison of his theory to Newtonian theory by the mathematical tools available to him. In particular, he lacked the concept of an "affine connection" (see Stachel's "The Story of Newstein …" in vol. 4 of this series).

ing covariance under transformations not associated with a change of the state of motion, such as transformation from Cartesian to spherical spatial coordinates. So equations of broad covariance could plausibly be used by Einstein purely as an intermediate mathematical step, from which the final equations could be recovered by some harmless restriction of the covariance. Of course, what Einstein found again and again in the notebook was that he needed restrictions that seriously compromised his goal of extending the relativity principle. Second, as already indicated above, Einstein found that the coordinate restrictions he applied were closely associated with energy-momentum conservation in the weak-field limit of the theory. This naturally suggested to him that the full theory could only be made compatible with energy-momentum conservation at the expense of restricting its covariance.

More speculatively, Norton suggests that Einstein's use of coordinate restrictions may have been supported by a tacit reification of spacetime coordinates. In an addendum to the reprint of the *Entwurf* paper in the *Zeitschrift für Mathematik und Physik* in January 1914 (Einstein and Grossmann 1914a), Einstein advanced his notorious "hole argument" [*Lochbetrachtung*] against the physical admissibility of generally-covariant field equations. By his own later admission, he fell into this flawed argument because of just such a tacit reification. What if Einstein tacitly reified spacetime coordinates in just the same way during his calculations in the notebook? If he did, Norton argues, the reification would have forced him to restrict the covariance of the theory, even had he known full well how to use coordinate conditions in the modern sense. The conjecture is that Einstein did indeed reify coordinates in this way. It is essential to the conjecture that this reification remained beneath Einstein's conscious awareness, just as he later admitted it did with the hole argument. Had he been able to formulate it explicitly, he would presumably have recognized its inadmissibility right away.

Norton suggests that Einstein had the modern understanding of coordinate conditions all along, even though, for the reasons laid out above, he used coordinate restrictions in the notebook as well as in the further development of the *Entwurf* theory. In "Untying the Knot …" (volume two), however, Janssen and Renn suggest that Einstein only arrived at the modern understanding of coordinate conditions when he replaced the *Entwurf* field equations by equations of much broader covariance in (Einstein 1915a). This paper contains the first unequivocal example of Einstein applying a coordinate condition in the modern sense.

This disagreement is one of a number of disputes that have not been resolved as these volumes go to press. We make no apologies for this. It is a sign of the vitality of Einstein scholarship. We publish our book on the history of general relativity in the same spirit as Hermann Weyl published his famous book on the theory itself. As Weyl wrote in the preface to the first edition of *Raum-Zeit-Materie*:

> But it was definitely also not the intention of this book to turn the life, which manifests itself so forcefully today on the field of physical knowledge, with axiomatic thoroughness into a dead mummy at the point it happens to have reached this moment.[13]

The reader will note that only the introduction and the commentary are authored by all five of us. The rest of the volume consists of essays by one or two authors. In part, this reflects division of labor and differences in expertise. But it also reflects a lack of consensus about the interpretation of the material. Points of contention, besides the issue of coordinate conditions versus coordinate restrictions, are the relative importance of the mathematical and the physical strategy in the genesis of general relativity (as well as the interpretation of this distinction), the nature of Einstein's *modus operandi* as a creative scientist, and the extent to which the analysis of this particular episode provides insights into the practice of scientific theorizing, its conditions and mechanisms, in general. The first two of these disagreements call for some further comments.

The breakthrough to general relativity of November 1915 has been portrayed as a triumph of the mathematical strategy, prematurely abandoned in the Zurich Notebook in favor of the physical strategy that led Einstein to the problematic *Entwurf* theory (see, e.g., Norton 2000). This portrayal certainly fits nicely with the way in which Einstein presented his new results in November 1915. It is also how he came to remember his own achievement in later years, as has been documented in great detail in (Van Dongen 2002). In "Untying the Knot …" (volume two), Janssen and Renn nevertheless argue that Einstein made the transition from the *Entwurf* field equations to the field equations of November 1915 following the physical strategy. They reconstruct the developments of that eventful month taking Einstein at his word that the definition of the gravitational field in the *Entwurf* theory as the gradient of the metric was "a fateful prejudice" ("ein verhängnisvolles Vorurteil," Einstein 1915a, 782) and that replacing this gradient by the so-called Christoffel symbols was "the key to [the] solution" ("Den Schlüssel zu dieser Lösung," Einstein to Sommerfeld, 28 November 1915). With this substitution, the variational formalism for the *Entwurf* theory, developed in (Einstein and Grossmann 1914b) and (Einstein 1914) in close analogy with classical electrodynamics, leads almost automatically to the theory of November 1915. In an appendix to "Untying the Knot …," this insight is used to clarify the mathematical relation between the Einstein field equations and the *Entwurf* field equations.

Concerning Einstein's *modus operandi*, not all of us agree with Janssen's claim in "What Did Einstein Know and When Did He Know It? A Besso Memo Dated August 1913" (volume two) that more room needs to be made for the role of prejudice, wishful thinking, and opportunism in Einstein's work toward general relativity than was done in the accounts we started from (Norton 1984, Stachel 1989b).

Janssen's claim is based primarily on Einstein's handling of what we shall call the problem of rotation. An important test to which Einstein subjected candidate field equations in the notebook is whether they allow the Minkowski metric in uniformly rotating coordinates as a vacuum solution. The point of checking whether this rota-

13 "Es lag aber auch durchaus nicht in der Absicht dieses Buches, das auf dem Feld der physikalischen Erkenntnis heute so besonders kräftig sich rührende Leben an dem Punkt, den es im Augenblicke erreicht hat, mit axiomatischer Gründlichkeit in eine tote Mumie zu verwandeln" (Weyl 1918, vi).

tion metric, as we shall call it, is a vacuum solution was that Einstein wanted to make sure that his theory extended the relativity of uniform motion of special relativity to accelerated motion such as uniform rotation. When Einstein published the *Entwurf* field equations, he had satisfied himself that they met this important requirement. It turns out they do not. It took Einstein some time to discover this and even longer to accept it. As part of his collaboration with Besso on the Mercury anomaly shortly after the publication of the *Entwurf* paper, he made a half-hearted attempt to double-check that the rotation metric is indeed a vacuum solution of the *Entwurf* field equations. He was so convinced they were that he missed some factors of 2 and a minus sign in just the right places to bring about the outcome he expected.[14] This error and its eventual discovery in September 1915 are recounted in (Janssen 1999). The conclusion of this paper was that Einstein had simply been unlucky in not discovering his error sooner. However, a memo in Besso's hand bearing the date August 28, 1913, discovered in Switzerland in 1998 by Robert Schulmann, shows that Besso had clearly warned his friend in 1913 that the rotation metric is not a solution. Einstein initially accepted this verdict. He changed his mind again in early 1914.

The reconstruction of this episode in "What Did Einstein Know …" depends crucially on allowing a certain opportunistic streak in Einstein's general methods. Einstein's handling of the problem with the rotation metric is the most clear-cut case, but the Besso memo provides another example. The memo shows not only that Einstein had the key idea for the hole argument sometime in August 1913, but also that he had already taken important steps toward its eventual resolution in 1915. Janssen argues that it is partly because of opportunism that Einstein did not pursue this resolution in 1913. After all, such a resolution would be a most unwelcome reversal after Einstein had tried for months to make the lack of covariance of the *Entwurf* field equations more palatable. The exposure of such opportunistic moves by Einstein should not be seen as damning to his reputation as a scientist. What it shows rather is that creative science is messier and more complicated than many philosophers of science and science educators like to think.

The material discussed so far focuses on one phase in the genesis of general relativity, Einstein's search for field equations in the period 1912–1915. This phase is extremely well documented and has accordingly been discussed extensively in the recent literature on the history of general relativity. Yet, the field equations were only the capstone on the edifice of general relativity. Much of the foundation had been laid before Einstein started looking in earnest for field equations. The cornerstone of the theory—dubbed the equivalence principle in (Einstein 1912, 360, 366) and later singled out by Einstein as "the most fortunate thought of my life" ("der glücklichste Gedanke meines Lebens," CPAE 7, Doc. 31, [p. 21])—dates from 1907. Stachel's essay, "The First Two Acts" (this volume), examines all Einstein source material per-

14 This calculation can be found on [pp. 41–42] of the Einstein-Besso manuscript on the perihelion motion of Mercury (CPAE 4, Doc. 14). For further discussion of Einstein's struggles with rotation, see (Janssen 2005).

taining to the early stages of the genesis of general relativity.[15] "Act One" concerns his recognition of the close connection between acceleration and gravitation in Newtonian mechanics, based on the equality of inertial and gravitational mass, and his extension of that connection to all physical phenomena, which he called the equivalence principle. He took this principle, which he believed extended the relativity principle to accelerated frames of reference, as the key to any future relativistic theory of gravitation. In the crucial "Act Two," the quest for generalization of the relativity principle, which distinguished his efforts at a theory of gravitation from those of his special-relativistic competitors, had led him by late 1912 from a scalar to a tensor description of the gravitational field. In the theory for which he was looking, this tensor was to play a dual physical role, serving both as the metric in the line element $g_{\mu\nu}dx^{\mu}dx^{\nu}$ [16] describing the chronometry of time and the geometry of space (by then he had adopted Minkowski's four-dimensional viewpoint), and as the potentials for the inertio-gravitational field. It is at this point that the Zurich Notebook begins. In this reckoning, the search for the field equations, the focus of much of the rest of these volumes, comprises only the third act.

The genesis of general relativity, however, did not quite unfold in the form of a classic three-act drama between 1907 and 1915. The story begins before 1907 and continues well beyond 1915. In the essay that opens this volume, "Classical Physics in Disarray," Renn examines the state of knowledge about gravitation from which Einstein had to start. The work on gravity undertaken on this basis eventually culminated in the publication of the Einstein field equations (Einstein 1915d). This publication, the last of the four communications to the Prussian Academy of November 1915, marks the dramatic end of "Act Three". But the dust did not settle until 1918. Both the status of general covariance and the status of energy-momentum conservation in the theory remain unclear as "Act Three" draws to a close. Einstein himself tied together some of these loose ends in correspondence and further publications.[17] Several contributions by others helped clarify key aspects of the theory. The seminal paper of Emmy Noether (1918) on symmetries and conservation laws already mentioned above grew out of discussions among Göttingen mathematicians around David Hilbert and Felix Klein over energy-momentum conservation in general relativity.[18] The astronomer Karl Schwarzschild (1915)[19] replaced the approximate solution for

---

15 This is the only essay in these volumes that has already been published (Stachel 2002a, 261–292). This collection of Stachel's work on Einstein also contains a reprint of an important earlier paper on "Act Two" (Stachel 1989a)

16 The indices μ and ν run from 1 through 4 (corresponding to the spacetime coordinates $x^{\mu} = (x, y, z, t)$) and are summed over.

17 The status of general covariance is clarified in letters to Besso and Paul Ehrenfest in late 1915 and early 1916 (CPAE 8, Docs. 173, 178 and 180). For historical discussion, see, e.g., (Stachel 1989, 1993, 2002b), (Norton 1987), (Howard and Norton 1993), (Howard 1999), and (Janssen 2005). The status of energy-momentum conservation was clarified in (Einstein 1916c, 1918c). For discussion, see "Untying the Knot …" (volume two).

18 For further discussion, see (Rowe 1999), (Sauer 1999, 2005), and Renn and Stachel's "Hilbert's Foundation of Physics …" in vol. 4 of this series.

the metric field of a point mass that Einstein had used in his paper on Mercury's perihelion by an exact solution. The mathematicians Gerhard Hessenberg (1917), Tullio Levi-Civita (1917), and Hermann Weyl (1918) introduced the notion of parallel displacement, thereby clarifying the geometrical meaning of curvature representing gravity in the new theory.[20] This in turn led to the introduction of the notion of an affine connection, which provides a much more natural way to implement the equivalence principle than the metric tensor, with which Einstein had to make do in the development of the theory.[21]

The longest and most ambitious essay in these two volumes is "Pathways out of Classical Physics …" (this volume) by Renn and Sauer. This paper elaborates on "Classical Physics in Disarray" (this volume) and presents a comprehensive version of "Act Three" of the genesis of general relativity within a framework of historical epistemology that integrates historical analysis, epistemology, and cognitive science. This framework builds on efforts to integrate cognitive science and history of science by Peter Damerow (1996) and is based on adapting such concepts as "frames" and "mental models" introduced by Marvin Minsky (1975, 1987) and others to the needs of an historical analysis of knowledge. Damerow was closely involved with our group's efforts to decipher the gravitational part of the Zurich Notebook and to explore the ramifications of the results for the reconstruction of the genesis of general relativity. One of the advantages of the framework adopted by Renn and Sauer is that it allows them to trace continuities across sharp conceptual divides that separate various stations on Einstein's journey from classical physics to general relativity. The hope is that the application of the framework of historical epistemology to the genesis of general relativity will tell us more about conceptual innovation in general and about related issues in—to use a geological metaphor—the plate tectonics of knowledge.

REFERENCES

Büttner, Jochen, Jürgen Renn, and Matthias Schemmel. 2003. "Exploring the Limits of Classical Physics. Planck, Einstein and the Structure of a Scientific Revolution." *Studies in History and Philosophy of Modern Physics* 34B, 37–59.

Castagnetti, Giuseppe, Peter Damerow, Werner Heinrich, Jürgen Renn, and Tilman Sauer. 1994. *Wissenschaft zwischen Grundlagenkrise und Politik: Einstein in Berlin. Arbeitsbericht der Arbeitsstelle Albert Einstein*. Max-Planck-Institut für Bildungsforschung.

CPAE 4: Martin J. Klein, A. J. Kox, Jürgen Renn, and Robert Schulmann (eds.), *The Collected Papers of Albert Einstein*. Vol. 4. *The Swiss Years: Writings, 1912–1914*. Princeton: Princeton University Press, 1995.

CPAE 5: Martin J. Klein, A. J. Kox, and Robert Schulmann (eds.), *The Collected Papers of Albert Einstein*. Vol. 5. *The Swiss Years: Correspondence, 1902–1914*. Princeton: Princeton University Press, 1993.

---

19 See Matthias Schemmel's "An Astronomical Road to General Relativity …" (vol. 3 of this series) for the story of how Schwarzschild came to be interested in these matters.

20 The importance of this development was acknowledged in (Einstein 1918d), a review of (Weyl 1918); in Einstein's lectures on general relativity in Berlin in 1919 (CPAE 7, Doc. 19, [p. 10]); and in correspondence (see, e.g., Einstein to Hermann Weyl, 27 September 1918 [CPAE 8, Doc. 626]).

21 See Stachel's "The Story of Newstein …" in vol. 4 of this series.

CPAE 6: A. J. Kox, Martin J. Klein, and Robert Schulmann (eds.), *The Collected Papers of Albert Einstein*. Vol. 6. *The Berlin Years: Writings, 1914–1917*. Princeton: Princeton University Press, 1996.

CPAE 7: Michel Janssen, Robert Schulmann, József Illy, Christoph Lehner, and Diana Kormos Barkan (eds.), *The Collected Papers of Albert Einstein*. Vol. 7. *The Berlin Years: Writings, 1918–1921*. Princeton: Princeton University Press, 2002.

CPAE 8: Robert Schulmann, A. J. Kox, Michel Janssen, and József Illy (eds.), *The Collected Papers of Albert Einstein*. Vol. 8. *The Berlin Years: Correspondence, 1914–1918*. Princeton: Princeton University Press, 1998.

Damerow, Peter. 1996. *Abstraction and Representation. Essays on the Cultural Revolution of Thinking*. Dordrecht: Kluwer.

Earman, John, and Michel Janssen. 1993. "Einstein's Explanation of the Motion of Mercury's Perihelion." In (Earman et al. 1993, 129–172).

Earman, John, Michel Janssen, and John D. Norton (eds.). 1993. *The Attraction of Gravitation* (*Einstein Studies*, Vol. 5). Boston: Birkhäuser.

Einstein, Albert. 1912. "Lichtgeschwindigkeit und Statik des Gravitationsfeldes." *Annalen der Physik* 38: 355–369, (CPAE 4, Doc. 3).

———. 1914. "Die formale Grundlage der allgemeinen Relativitätstheorie." *Königlich Preußische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte*: 1030–1085, (CPAE 6, Doc. 9).

———. 1915a. "Zur allgemeinen Relativitätstheorie." *Königlich Preußische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte*: 778–786, (CPAE 6, Doc. 21).

———. 1915b. "Zur allgemeinen Relativitätstheorie. (Nachtrag)." *Königlich Preußische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte*: 799–801, (CPAE 6, Doc. 22).

———. 1915c. "Erklärung der Perihelbewegung des Merkur aus der allgemeinen Relativitätstheorie." *Königlich Preußische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte*: 831–839, (CPAE 6, Doc. 24).

———. 1915d. "Die Feldgleichungen der Gravitation." *Königlich Preußische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte*: 844–847, (CPAE 6, Doc. 25).

———. 1916a. "Die Grundlage der allgemeinen Relativitätstheorie." *Annalen der Physik* 49: 769–822, (CPAE 6, Doc. 30).

———. 1916b. "Näherungsweise Integration der Feldgleichungen der Gravitation." *Königlich Preußische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte*: 688–696, (CPAE 6, Doc. 32).

———. 1916c. "Hamiltonsches Prinzip und allgemeine Relativitätstheorie." *Königlich Preußische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte*: 1111–1116, (CPAE 6, Doc. 41).

———. 1917. "Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie." *Königlich Preußische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte*: 142–152, (CPAE 6, Doc. 43).

———. 1918a. "Über Gravitationswellen." *Königlich Preußische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte*: 154–167, (CPAE 7, Doc. 1).

———. 1918b. "Prinzipielles zur allgemeinen Relativitätstheorie." *Annalen der Physik* 55: 241–244, (CPAE 7, Doc. 4).

———. 1918c. "Der Energiesatz der allgemeinen Relativitätstheorie." *Königlich Preußische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte*: 448–459, (CPAE 7, Doc. 9).

———. 1918d. Review of: Hermann Weyl, *Raum–Zeit–Materie*. *Die Naturwissenschaften* 6: 373, (CPAE 7, Doc. 10).

———. 1936. "Physics and Reality." *The Journal of the Franklin Institute* 221 No. 3. Page reference to reprint in *Ideas and Opinions*, New York: Bonanza, 1954.

Einstein, Albert, and Marcel Grossmann. 1913. *Entwurf einer verallgemeinerten Relativitätstheorie und einer Theorie der Gravitation*. Leipzig: Teubner, (CPAE 4, Doc. 13).

———. 1914a. "Entwurf einer verallgemeinerten Relativitätstheorie und einer Theorie der Gravitation." *Zeitschrift für Mathematik und Physik* 62: 225–259. Reprint of *Einstein and Grossmann 1913* with additional "Comments" ("Bemerkungen," CPAE 4, Doc. 26).

———. 1914b. "Kovarianzeigenschaften der Feldgleichungen der auf die verallgemeinerte Relativitätstheorie gegründeten Gravitationstheorie." *Zeitschrift für Mathematik und Physik* 63: 215–225, (CPAE 6, Doc. 2).

Eisenstaedt, Jean, and A. J. Kox (eds.). 1992. *Studies in the History of General Relativity* (*Einstein Studies*, Vol. 3). Boston: Birkhäuser.

Goenner, Hubert, Jürgen Renn, Jim Ritter, and Tilman Sauer (eds.). 1999. *The Expanding Worlds of General Relativity* (*Einstein Studies*, Vol. 7). Boston: Birkhäuser.

Hessenberg, Gerhard. 1917. "Vektorielle Begründung der Differentialgeometrie." *Mathematische Annalen* 78: 187–217.

Howard, Don. 1999. "Point Coincidences and Pointer Coincidences: Einstein on the Invariant Content of Space-Time Theories." In (Goenner et al. 1999, 463–500).

Howard, Don, and John D. Norton. 1993. “Out of the Labyrinth? Einstein, Hertz, and the Göttingen Response to the Hole Argument.” In (Earman et al. 1993, 30–62).
Howard, Don, and John Stachel (eds.). 1989. *Einstein and the History of General Relativity* (*Einstein Studies*, Vol. 1). Boston: Birkhäuser.
Janssen, Michel. 1999. “Rotation as the Nemesis of Einstein’s Entwurf Theory.” In (Goenner et al., 127–157).
———. 2003. “A Glimpse Behind the Curtain of the Wizard/Un coup d’œil derrière le rideau du magicien.” In *The Einstein-Besso Manuscript: From Special Relativity to General Relativity/Le manuscrit Einstein-Besso: De la Relativité Restreinte à la Relativité Générale*. Paris: Scriptura and Aristophile.
———. 2005. “Of Pots and Holes: Einstein’s Bumpy Road to General Relativity.” *Annalen der Physik* 14: *Supplement* 58–85. Reprinted in J. Renn (ed.) *Einstein’s Annalen Papers. The Complete Collection 1901–1922*. Weinheim: Wiley-VCH, 2005.
Kox, A. J., and Jean Eisenstaedt (eds.). 2005. *The Universe of General Relativity* (*Einstein Studies*, Vol. 11). Boston: Birkhäuser.
Levi-Civita, Tullio. 1917. “Nozione di parallelismo in una varietà qualunque e conseguente specificazione geometrica della curvatura Riemanniana.” *Circolo Matematico di Palermo. Rendiconti* 42: 173–205.
Minsky, Marvin. 1975. “A Framework for Representing Knowledge.” In P. H. Winston (ed.), *The Psychology of Computer Vision*. New York: McGraw-Hill, 211–281.
———. 1987. *The Society of Mind*. London: Heinemann.
Noether, Emmy. 1918. “Invariante Variationsprobleme.” *Königliche Gesellschaft der Wissenschaften zu Göttingen. Mathematisch-physikalische Klasse. Nachrichten*: 235-257.
Norton, John D. 1984. “How Einstein Found his Field Equations, 1912–1915.” *Historical Studies in the Physical Sciences* 14: 253–316. Reprinted in (Howard and Stachel 1989, 101–159). Page references to this reprint.
———. 1987. “Einstein, the Hole Argument and the Reality of Space.” In John Forge (ed.), *Measurement, Realism, and Objectivity*. Dordrecht: Reidel, 153–188.
———. 2000. “‘Nature is the Realisation of the Simplest Conceivable Mathematical Ideas’: Einstein and the Canon of Mathematical Simplicity.” *Studies in History and Philosophy of Modern Physics* 31: 135–170.
Pais, Abraham. 1982. *‘Subtle is the Lord …’ The Science and the Life of Albert Einstein*. Oxford: Oxford University Press.
Renn, Jürgen. 2005a. “Standing on the Shoulders of a Dwarf: General Relativity—A Triumph of Einstein and Grossmann’s Erroneous *Entwurf* Theory.” In (Kox and Eisenstaedt 2005, 39–51).
———. 2005b. “Before the Riemann Tensor: The Emergence of Einstein’s Double Strategy.” In (Kox and Eisenstaedt 2005, 53–65).
Renn, Jürgen, and Tilman Sauer. 1996. “Einsteins Züricher Notizbuch. Die Entdeckung der Feldgleichungen der Gravitation im Jahre 1912.” *Physikalische Blätter* 52: 865–872.
———. 1999. “Heuristics and Mathematical Representation in Einstein’s Search for a Gravitational Field Equation.” In (Goenner et al., 127–157).
———. 2003. “Errors and Insights: Reconstructing the Genesis of General Relativity from Einstein’s Zurich Notebook.” In Frederic L. Holmes, Jürgen Renn, and Hans-Jörg Rheinberger (eds.), *Reworking the Bench: Research Notebooks in the History of Science*. Dordrecht: Kluwer, 253–268.
Roseveare, N. T. 1982. *Mercury’s Perihelion from Leverrier to Einstein*. Oxford: Clarendon Press.
Rowe, David. 1999. “The Göttingen Response to General Relativity and Emmy Noether’s Theorems.” In Jeremy Gray (ed.), *The Symbolic Universe: Geometry and Physics, 1890–1930*. Oxford: Oxford University Press, 189–234.
Sauer, Tilman. 1999. “The Relativity of Discovery: Hilbert’s First Note on the Foundations of Physics.” *Archive for History of Exact Sciences* 53: 529–575.
———. 2005. “Einstein Equations and Hilbert Action: What is Missing on Page 8 of the Proofs for Hilbert’s First Communication on the Foundations of Physics?” *Archive for History of Exact Sciences* 59: 577–590.
Schwarzschild, Karl. 1916. “Über das Gravitationsfeld eines Massenpunktes nach der Einsteinschen Theorie.” *Königlich Preußische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte* (1916): 189–196.
Stachel, John. 1989a. “The Rigidly Rotating Disk as the “Missing Link” in the History of General Relativity.” In (Howard and Stachel 1989, 48–61). Reprinted in (Stachel 2002a, 245–260).
———. 1989b. “Einstein’s Search for General Covariance, 1912–1915.” In (Howard and Stachel 1989, 62–100). Reprinted in (Stachel 2002a, 301–337).
———. 1993. “The Meaning of General Covariance: the Hole Story.” In John Earman, Allen I. Janis, Gerald J. Massey, and Nicholas Rescher (eds), *Philosophical Problems of the Internal and External World: Essays on the Philosophy of Adolf Grünbaum*. Konstanz: Universitätsverlag/Pittsburgh: University of Pittsburgh Press 129–160.
———. 2002a. *Einstein from ‘B’ to ‘Z’* (*Einstein Studies*, Vol. 9). Boston: Birkhäuser.

———. 2002b. "“The Relations between Things” versus “The Things between Relations”: The Deeper Meaning of the Hole Argument." In David B. Malament (ed.), *Reading Natural Philosophy. Essays in the History and Philosophy of Science and Mathematics*. Chicago and La Salle: Open Court, 231–266.

———. Forthcoming. *Einstein's Intuition and the Post-Newtonian Approximation*.

Van Dongen, Jeroen. 2002. *Einstein's Unification: General Relativity and the Quest for Mathematical Naturalness*. PhD. Thesis. University of Amsterdam.

Weyl, Hermann. 1918. *Raum–Zeit–Materie. Vorlesungen über allgemeine Relativitätstheorie*. Berlin: Springer.

JÜRGEN RENN

CLASSICAL PHYSICS IN DISARRAY

The Emergence of the Riddle of Gravitation

1. INTRODUCTION

1.1 A Creation ex nihilo?

The genesis of Einstein's special and general theory of relativity is an odd event in the history of science. From today's perspective, Einstein's theory represents the basis for modern astronomy, astrophysics, cosmology, and cosmogony. It comprises a broad range of observational and theoretical knowledge, covering, among others, phenomena related to planetary astronomy, to black holes, and to the expansion of the universe. Yet little of the knowledge that makes relativity theory a central asset of modern physics was available at the time when Einstein completed it by publishing his paper on general relativity in late 1915. Neither the bending of light in a gravitational field nor the expansion of the universe, let alone gravitational waves or black holes, were even suspected by contemporary astronomers. How then was it possible for Einstein without this knowledge to formulate a theory that has since withstood not one but several revolutions of astronomy and its instrumentation, including the development of radio-, X-ray, and space-borne astronomy?

A closer look at Einstein's investigative pathway does not resolve but rather complicates this puzzle, which may be called the "paradox of missing knowledge." Einstein first encountered the problem of formulating a relativistic theory of gravitation in 1907 when he was a clerk at the Swiss Patent Office in Bern. In the following eight years he pursued this problem with growing intensity, from 1911 to early 1914 as professor in Zurich and Prague, and from April 1914 as a member of the Prussian Academy and from 1917 as Director of the Kaiser Wilhelm Institute for Physics in Berlin.[1] As early as 1907 Einstein formulated several general conditions for the solution of his problem, among them the famous "principle of equivalence" and a generalization of the "principle of relativity." The first principle allowed him to arrive immediately at a number of surprising conclusions such as that of the bending of light in a gravitational field.[2] Five years later, however, he had to acknowledge that,

1 See (CPAE 5), Calender/Chronology, pp. 617–636.
2 See (Einstein 1907, 461).

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 1
Einstein's Zurich Notebook: *Introduction and Source*.

with the mathematical means at his disposal, the problem of formulating a relativistic theory of gravitation could not be solved. To his friend, Heinrich Zangger, Einstein thus wrote around June 1912:

> The further development of the theory of gravitation meets with great obstacles.[3]

Consequently, in the summer of 1912, he turned to his mathematician friend, Marcel Grossmann, who helped him to access more sophisticated mathematical tools, in particular, the absolute differential calculus of Elwin Christoffel, Gregorio Ricci-Curbastro, and Tullio Levi-Civita.[4] But after exploring these tools for about a year, an experience that is documented in the Zurich Notebook, Einstein's conviction grew that the problem he posed in 1907 was actually irresolvable. He thus limited himself to a partial solution, which he first published in 1913 together with Grossmann.[5] Soon after this publication Einstein even believed to have found a proof that his problem could not be solved as originally envisaged. In the following two years, he nevertheless continued to work on a relativistic theory of gravitation largely in isolation from and even against the resistance of the scientific community, which tended to regard his efforts as making little sense. Then, after three dramatic weeks towards the end of 1915 in which Einstein presented to the Prussian Academy week after week a new tentative solution,

---

3 "Die Weiterentwicklung der Theorie der Gravitation stösst auf grosse Hindern[i]sse." Albert Einstein to Heinrich Zangger, Prague, after 5 June 1912, (CPAE 5, Doc. 406). Unless otherwise noted, all translations are based on those in the companion volumes to the Einstein edition.

4 See (Kollross 1955, 278) according to which Einstein exclaimed: "Grossmann, you must help me or I'll go crazy." ("Grossmann, Du mußt mir helfen, sonst werd' ich verrückt!") See also the preface to the Czech translation (Einstein 1923) of his popular book on relativity where he wrote: "However, only after my return in 1912 to Zurich did I hit upon the decisive idea about the analogy between the mathematical problem connected with my theory and the theory of surfaces by Gauss—originally without knowledge of the research by Riemann, Ricci, and Levi-Civita. The latter research came to my attention only through my friend Grossmann in Zurich when I posed the problem to him only to find generally covariant tensors whose components depend only upon the derivatives of the coefficients of the quadratic fundamental invariant." ("Den entscheidenden Gedanken von der Analogie des mit der Theorie verbundenen mathematischen Problems mit der Gaußschen Flächentheorie hatte ich allerdings erst 1912 nach meiner Rückkehr nach Zürich, ohne zunächst Riemanns und Riccis, sowie Levi-Civitas Forschungen zu kennen. Auf diese würde ich erst durch meinen Freund Großmann in Zürich aufmerksam, als ich ihm das Problem stellte, allgemein kovariante Tensoren aufzusuchen, deren Komponenten nur von Ableitungen der Koeffizienten der quadratischen Fundamentalinvariante abhängen."), (CPAE 6, Doc. 42, 535, n. 4). For more extensive discussion, see "The First Two Acts" (in this volume).

5 See (Einstein and Grossmann 1913).

each revoking the previous one, he eventually succeeded in definitively solving his original problem.[6] How was it possible that Einstein could, without having an idea of the final solution and eight years before actually attaining it, already formulate the conditions it had to satisfy, and how could he persist in his search against the judgement of contemporary experts and in spite of his numerous failures? What was his explicit or implicit heuristics? And what accounts for its success?

The following is an attempt to prepare the answers to these questions by analyzing the roots of Einstein's achievements in the knowledge of classical and special relativistic physics, and by following the development of his theory until, in the summer of 1912, he recognized gravitation as the bending of space and time. As we shall see in this and the following contributions, both the peculiar emergence and the remarkable stability of Einstein's theory of gravitation with regard to the further development of physics and astronomy become plausible only if the genesis of general relativity is understood, not as a fortunate anticipation of future observational discoveries, but as a transformation of pre-existing knowledge. The next section provides a survey of this knowledge, in particular of classical physics as it became relevant to the emergence of a relativistic theory of gravitation. The concluding third section then attempts to explain how this knowledge became effective in laying the foundations for Einstein's successful heuristics.[7] As a prelude, let us look briefly at the long-term development of the knowledge on gravitation, which provided the presuppositions for Einstein's achievements.

1.2 A Short History of Gravitation

What can one possibly learn from a survey of the history of gravitation in order to understand the genesis of general relativity? Certainly, the long-range history of knowledge on gravitation turns out to be as peculiar as the emergence of general relativity. Above all it is characterized by the longevity of certain basic ideas on gravitational effects, as well as by the radical turnover of these ideas in the course of the historical development. For almost two millennia, the understanding of what we consider to be gravitational effects was dominated by Aristotelian natural philosophy. It divides such effects into two distinct classes; the motions of terrestrial and of celestial bodies. The downward motion of heavy terrestrial bodies is conceived as a "natural motion" towards the natural place of such bodies, the center of the earth. In contrast to the "violent motion" of terrestrial bodies, which is caused by a force, natural motion does not require any external moving cause but is the result of an intrinsic tendency of a heavy body acting in accordance with its nature. The characteristic motion of celestial bodies is, on the other hand, categorically separated from that of terrestrial bodies and conceived as an eternal circular motion.[8]

---

6 See (Einstein 1915a, 1915b, 1915c, 1915d).

7 For an attempt to provide comprehensive answers to the above questions based on the present analysis, see "Pathways out of Classical Physics …" (in this volume).

The Aristotelian account of gravitational effects was eventually supplanted by that of classical Newtonian physics, dominating their understanding for several centuries until the advent of relativity theory. Like Aristotle's natural philosophy, Newtonian physics also introduced a fundamental distinction between types of motion, in this case between uniform inertial motions and motions that are accelerated due to the action of a force. In Newtonian physics, gravitation is understood as a force acting like any other external force causing acceleration, be it human or natural, terrestrial or celestial. Accordingly, the accelerated motion of falling terrestrial bodies and the accelerated orbital motion of the planets are explained in the same way as being due to a universal gravitational force of attraction. This gravitational force acts without intermediary and without time delay through the intervening space between two bodies.

Since the mid-nineteenth century, yet another distinct idea was discussed and gained increasing support: gravitation as a space-filling field, transmitting the gravitational force not instantaneously but with a limited speed through an intervening medium called the "aether."[9] Finally, in general relativity, gravitation is conceived as a field that represents the curvature of the spacetime continuum itself and that is caused by the distribution of masses and energy in the universe. Since gravitation is not understood here as a force, motions of bodies within this field are no longer distinguished according to their inertial or gravitational character. They are rather all "natural" motions governed in the same way by the intrinsic geometry of spacetime.

What made these distinct conceptions convincing while they ruled the understanding of gravitation, and what eventually overturned them? Was the long-term domination of Aristotelian natural philosophy merely the consequence of its adoption as the official doctrine of the Catholic Church? And was the replacement of the Aristotelian concept of natural motion by the Newtonian explanation of the motion of fall in terms of a gravitational force simply the triumph of a newly introduced scientific method? If so, what accounts for the fact that the Newtonian conception itself was eventually superseded by that of general relativity, which returns to the interpretation of the motion of fall as a "natural" force-free motion, as conceived in Aristotelian natural philosophy? How can it be that, even after the establishment of the "scientific method," history of science was not simply dominated by gradual progress and that even the very foundations of the understanding of gravitation could still be overturned? As the following sketch will attempt to make plausible, the emergence and disappearance of such diverse core notions of gravitation—conceived as natural tendency, force, field, or as curvature of spacetime—becomes understandable on the basis of a history of knowledge that studies the extension as well as the architecture of the knowledge sustained by these core notions.

While the history of science has traditionally limited itself to the knowledge embodied in scientific theories and, more recently, also to that represented by experi-

8 For a survey, also of the following, see (Dijksterhuis 1986).

9 For a survey, see the introduction to vols. 3 and 4 of this series "Theories of Gravitation in the Twilight of Classical Physics" (in vol.3).

mental systems, the knowledge on gravitation here at issue predates considerably any systematic theoretical treatment in the framework of physics.[10] The most basic knowledge on heaviness, force, matter, and motion is based on experiences acquired by human activities almost universally in any culture. It includes, for instance, the perception of heavy bodies and their spontaneous tendency to fall downwards and the fact that their motion usually requires an effort. The outcome of these experiences is an "intuitive mechanical knowledge" embedded in a qualitative physics, which is built up in ontogenesis and guides human activities in their relation to the physical environment. A second kind of physical knowledge, which predates any systematic theoretical treatment, is the knowledge gained by the use of practical tools. In contrast to intuitive physical knowledge, this type of knowledge is closely linked to the production and use of technology by professional groups of practitioners and is hence subject to historical development. For a long historical period, these two forms of knowledge, intuitive and practitioners' knowledge, also formed the principal experiential foundation of scientific knowledge which is characteristically represented and historically transmitted in the form of written texts. Without realizing the extent to which the development of theoretical knowledge of gravitation depends on the reflection on these forms of "shared knowledge," its dynamics remain inexplicable.[11]

The historical persistence of Aristotelian natural philosophy, in particular, is related to the fact that it incorporates knowledge about the natural environment common to all human beings. The knowledge of intuitive physics is structured by mental models, that is, by specific forms of representation that allow inferences to be drawn from prior experiences about complex objects and processes, even if only incomplete information is available about them.[12] One example is the motion-implies-force model" which, when involved in the interpretation of a process of motion, yields the conclusion that the moved object is moved by a force exerted upon it by some mover. While this conclusion is incorrect from the perspective of classical physics, contradicting as it does Newton's principle of inertia, it represents elementary human experiences. In fact, when observing a moving object, for instance a vehicle moving on the street, one usually presumes that there is a mover at work that drives the object by its force, even when the mover itself and its force cannot be directly observed. Mental models possess slots that can be filled with empirically gained information but also with default assumptions, as in the case at hand, when the mover remains invisible.

10 The following outlines some basic notions of historical epistemology as it is pursued at the Max Planck Institute for the History of Science in Berlin; see the scientific reports of the institute (http://www.mpiwg-berlin.mpg.de/en/forschung/reports.html). For contributions to an account of the development of mechanical knowledge to which the following makes reference, see (Damerow, Renn and Rieger 2002; Büttner, Damerow and Renn 2001; Damerow et al.; Büttner et al. 2003; Renn 2001; Damerow and Renn 2001).

11 For the notion of shared knowledge, see (Büttner, Damerow and Renn 2001).

12 For the concept of mental model, see, e.g., (Minsky 1987; Gentner and Stevens 1983; and Davis, 1984). For a view on the potential of cognitive science and cognitive psychology for the history of science to which the present work is much indebted, see (Damerow 1996). This approach is extensively used in "Pathways out of Classical Physics …" (in this volume).

Such assumptions may simply result from prior experience, or may be "inherited" from higher-order mental models. In any case, information is assimilated to the slots of a mental model in the form of "frames," that is, cognitive structures with a well-defined meaning that categorize information as being suitable for the slots of the model or provide such information, if necessary, by their "default settings." In the case at hand, such frames are those which either identify a given object as a mover or as moved or which supply—if suitable empirical information is lacking—default examples of such objects from prior experience. A higher-order mental model may trigger the activation of mental models to interpret the natural environment; it is, in any case, context-dependent. While motions such as that of a vehicle tend to be interpreted in terms of the "motion-implies-force model," other motions, such as the spontaneous downward motion of a falling body due to gravitation, are interpreted by another mental model, the "heaviness-causes-fall model," implying that a heavy body will, without any intervention by a mover, fall downwards once the obstacles to such a motion are removed.

The embedding of a mental model in a theory such as Aristotelian natural philosophy, resulting from the reflection of intuitive physics, imposes further requirements on the explanatory potential of the model, in particular a drive towards universalization, consistency, and precision in the use of the model. In Aristotelian physics for instance, the motion-implies-force model became the foundation for the treatment of the class of "violent motions," whereas the heaviness-causes-fall model serves as the basis for the explanation of the "natural motion" of heavy bodies. Cases that do not clearly belong to one of these two categories, such as the motion of a projectile that is initially propelled by an external force and eventually drops to the ground due to its heaviness, represent a problem for this theory. Which is, in particular, the external force driving the motion of the projectile once it is separated from its original mover? And how does this force interact with the natural tendency downwards in the course of the projectile's motion? The awkward attempts at solving such problems, identifying, for instance, the surrounding medium as the missing external cause of the continued motion of a projectile, illustrate the new kind of challenges encountered by a mental model of intuitive physics when integrated into a theoretical framework.

Problems such as that of projectile motion were the subject of critical discussions and revisions of Aristotelian natural philosophy and eventually led, in the early modern era, to the ascendancy of a new mechanics, but only after such problems had become an important challenge for contemporary practitioners and theoreticians. In the early modern period, scientist-engineers accumulated, in the context of the technical ventures of the time,—from artillery via intercontinental navigation to monumental building projects—a rich base of experiences largely exceeding those of intuitive physics and also of the practical knowledge of antiquity. The bodies of specialized practical knowledge that they assembled were in fact the principal source of empirical knowledge for the new sciences of the Scientific Revolution. The preclassical mechanics and astronomy of Galileo, Kepler, and their contemporaries were constituted by attempts to assimilate the new knowledge resources of this period to the traditional knowledge structures inherited from antique and medieval science.

The early modern period was characterized, however, not only by new bodies but also by new images of knowledge.[13] In opposition to the prevailing Aristotelian worldview, early modern intellectuals searched for a framework in which celestial and terrestrial phenomena would be explained by the same causes. Accordingly, they cherished explanations with the capacity to integrate knowledge from these traditionally separated spheres. Galileo, for instance, attempted to explain the genesis of the planetary system on the basis of a mental model originating in his analysis of projectile motion.[14] This model makes it possible to imagine that projectile motion—or, in the cosmological case, planetary motion—proceeds as if it were generated by a preceding accelerated motion of fall along the vertical which is then deflected into the horizontal, thus producing the parabolic motion of a projectile or the circular motion of a planet around the sun, as the case may be. In his cosmogony, Galileo speculated that there was a fixed point in space from which all planets were originally released so that their deflection into a circular motion after the proper distances of fall—producing their correct distances from the sun—would generate the appropriate speeds of their revolving motions around the sun. While Galileo's model is still shaped by the traditional conceptual framework, interpreting for instance the motion of fall as a natural motion and a projectile motion as violent motion, it nevertheless integrates areas of knowledge that were separated from each other according to the classificatory distinctions of Aristotelian natural philosophy. It shows, in particular, how the "natural motion" of fall can be converted into the "violent motion" of projection and even the "celestial motion" of a planet.

The vision of a unified treatment of terrestrial and celestial physics embodied in attempts such as Galileo's was eventually realized in Newtonian physics on the basis of a new mental model, the "acceleration-implies-force model," which makes it possible to explain the motion of the planets and the motion of falling bodies on earth by the same universal force of gravitation. This model is still rooted in intuitive physics, as is suggested by the anthropomorphic concept of force involved in it and also by its character as a modification of the motion-implies-force model, from which it results by the specification that it is not motion but acceleration that is caused by a force. But clearly the acceleration-implies-force model is actually no longer part of intuitive physics or of practitioners' knowledge, as is made evident by its counter-intuitive consequences such as the implication that a force-free motion will never come to rest. The acceleration-implies-force model as the basic explanatory scheme of classical mechanics emerged rather from a reflection on the extensive experiential basis accumulated in preclassical mechanics and astronomy. The radical change in physical explanations, which it implies when compared to explanations based on the motion-implies-force model, was the result rather than the presupposition of a process of

13 See (Elkana 1981).
14 See (Büttner 2001).

knowledge integration. When considering the origin of the acceleration-implies-force model in the integration of knowledge achieved by preclassical science, the gravitational force of Newtonian physics, with its peculiar properties such as acting at a distance without any intermediary or causing all bodies to fall with the same acceleration, no longer appears to be just the strange but compelling idea of an individual genius. It rather emerges as being supported by the accumulated knowledge of preclassical and then classical science, comprising terrestrial experiences as well as astronomical observations, a circumstance that accounts for its stability for more than three centuries of classical physics.

The emergence of Newtonian physics out of the knowledge-integration of preclassical physics accounts, however, not only for its stability but also for the fact that it could eventually be revised again. In fact, Newtonian physics does not embody, as philosophers such as Kant believed, an *a priori* framework determining the fundamental categories for the further development of physics, but merely represents a temporary structure of knowledge organization subject to historical change. Although the knowledge from which it originated was rather comprehensive and grew steadily with the elaboration of classical mechanics and also with the advancement of technology, it was, after all, only a highly specialized body of knowledge whose specific theoretical structures were transmitted in a tradition of experts, carried on by institutions of higher learning and a sophisticated technical literature. It is therefore not surprising that it did not replace the intuitive physics by which we orient ourselves in our natural environment. In fact, as psychological investigations show, the intellectual means used to address physical problems outside such expert traditions strikingly resemble those at the roots of Aristotelian natural philosophy, although Aristotle's system has long ceased to dominate the academic world.[15]

The fact that Newtonian physics is removed from intuitive physics not only makes it difficult to understand for laymen but also represented, in the course of its history, a potential challenge even for professional science. Newton's concept of gravitational force as "action at a distance," for instance, was criticized by his contemporaries as well as by later scientists because it does not provide a mechanism of interaction compatible with what was to be expected from intuitive physics, that is, an interaction mediated by the contact of a material agent. Numerous attempts were thus undertaken to develop models of gravitational interaction capable of countering or circumventing this objection.[16] An example is provided by the suggestion of the eighteenth-century scientist Lesage, according to which gravitation is an "umbrella effect" by two material bodies apparently exerting a mutual force of attraction but actually only screening each other from an omnipresent "rain" of invisible particles impinging on them from all sides.[17] The net-result of this stream of invisible particles on a single,

---

15 See (McCloskey 1983; Bödecker 2004).

16 For a survey, see (Taylor 1876). For criticism of Newton's theory in the eighteenth and nineteenth centuries, see (Dundon 1972). See also the introduction to vols. 3 and 4 of this series "Theories of Gravitation in the Twilight of Classical Physics" (in vol.3).

17 See (Le Sage 1758). For discussion see (Prévost 1805; Langley 1898; Laudan 1972; Edward 2002).

isolated body would simply vanish, given that they pressure such a body from all sides. But if this stream on a given body is screened from one side by another nearby body, the impacts of the invisible particles no longer neutralize each other but actually drive the two bodies towards each other.

The tension between classical physics and expectations rooted in intuitive physics served, however, not only to provoke conciliatory attempts such as the one just described but also simply to keep alive the astonishment about the Newtonian doctrine and some of its peculiarities, in particular with regard to the puzzling features of inertial motion. The autonomous character of both natural and celestial motions, showing no noticeable intervention of an external cause, was thus never lost from sight, even long after the Aristotelian world picture had vanished. While Newtonian physics was extremely successful in accounting for the motions of the planets and other celestial bodies it could therefore never completely extinguish objections against its explanations questioning, for instance, the rather artificial split of the orbital motion of a planet into an inertial component due to an intrinsic tendency of motion and an accelerated component due to the invisible gravitational action of the sun. But for a long time such questions were mainly left to philosophers such as Hegel and played no significant role in the development of classical physics.[18]

The situation gradually changed, however, with the steady growth of scientific knowledge and, in particular, with the discovery and exploration of new phenomena outside the range of mechanics, such as those of optics and electromagnetism. Paradoxically, it was the rapid development of the disciplinary specialization in nineteenth-century science which showed that specialization was not just a means of isolating problems from the larger context of knowledge. Specialization also turned out to be a form of the production of shared knowledge that unavoidably creates intersections between different, highly structured bodies of knowledge, making it necessary to review their explicit and implicit foundations. It was precisely such intersections that brought the problems of gravitation and inertia eventually back to the fore. This first happened around the middle of the nineteenth-century—at the time without much further consequence—in the context of philosophical discussions of the structure and role of mechanics, which were triggered by the establishment of other subdisciplines of classical physics. Towards the end of the nineteenth century, a rich discussion had evolved around foundational questions concerning gravity and inertia, which, however, were still largely neglected by mainstream classical physics.[19] The pursuit of foundational questions was then taken up with much greater impact as a consequence of specific new problems that may be characterized as "borderline problems" between the major continents of classical physics. Among those driven by such borderline problems to revisit the foundations of physics and eventu-

---

18 See, e.g., (Hegel 1986a, 1986b). For a survey, see (Ihmig 1989, Petry 1993), and in particular (Borzeszkowski 1993, Damerow 1979, Damerow and Lefèvre 1980, Wahsner 1993).

19 For a survey, see vols. 3 and 4 of this series.

ally revolutionize the classical understanding of gravitation was Albert Einstein, initially also an outsider on the periphery of established physics.

What can one learn from such a survey of the history of knowledge on gravitation? It specifies the challenge of what it means to reconstruct the genesis of general relativity as a transformation of knowledge. As we have seen, radical breaks in the understanding of gravitation were not due to the invention of new paradigms or the creation of new ideas *ex nihilo*, but rather to the long-term processes of the accumulation and integration of knowledge resources preceding such breaks. It was this evolution of knowledge that also produced tensions within the existing cognitive structures as well as the potential for resolving them: in the course of a reflective reorganization of the accumulated shared knowledge, its architecture could be revised from bottom up. In the following section we will show, by sketching the major building blocks of classical physics and tracing their intersections at borderline problems, how the riddle of gravitation reemerged at the beginning of the 20th century due to the integration of the knowledge of classical physics, which gave rise to a disintegration of its foundational concepts.

2. GRAVITATION AMONG THE BORDERLINE PROBLEMS OF CLASSICAL PHYSICS

2.1 The Three-Partite Division of Classical Physics

The specific problems addressed by the special and general theories of relativity are rooted in the development of classical physics in the late nineteenth century.[20] In the following we will therefore look at classical physics at this time, surveying its extended and gradually evolving landscape with the double aim of introducing a world not necessarily familiar to modern readers and of locating the structural origins of the problems addressed by Einstein. In the late nineteenth century physics had evolved into three major branches, each treating a set of interconnected physical problems on the basis of characteristic knowledge structures. Such knowledge structures comprise chunks of knowledge embodying basic, often only qualitative rules, which allow the prediction of the essential properties and behavior of physical objects. These chunks are best described in terms of the above-mentioned "mental models." Such knowledge structures also comprise elaborate conceptual frameworks represented by specialized theories, representational tools such as mathematical formalisms, and practical knowledge concerning paradigmatic experimental systems. These knowledge structures were transmitted in the context of the disciplinary organization of nineteenth century physics by school and university training, textbooks, journals, as well as the participation in research projects.

20 For surveys, see (Jungnickel and McCormmach 1986a, 1986b; Harman 1982; D'Abro, 1951a, 1951b; Stichweh 1984).

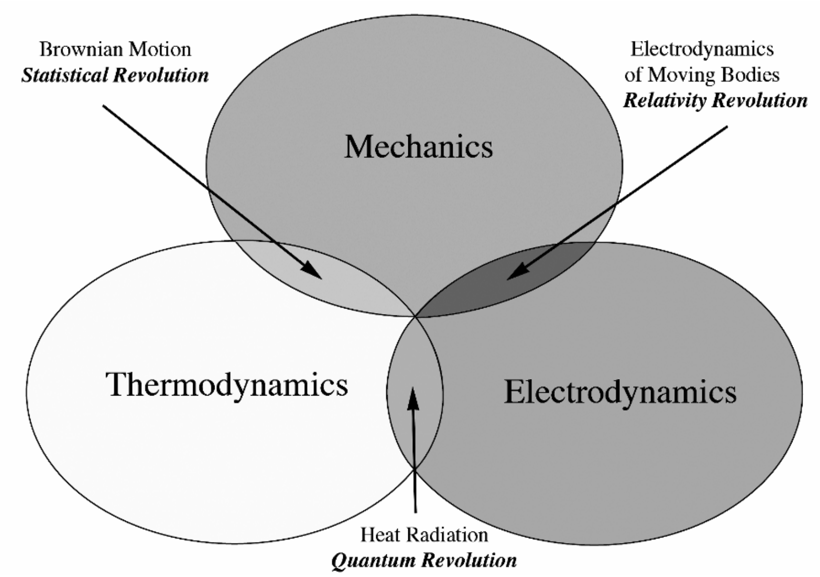


Figure 1: Revolutions at the Borderline Problems of Classical Physics

The most oldest structure of physics is mechanics, established as a major body of knowledge in the early modern period, following a tradition reaching back to antiquity, and brought into its classical form in the eighteenth and nineteenth centuries. Among its elementary mental models is, for example, a lever moveable around a fulcrum that enables a large weight to be lifted by a small force. Among its basic concepts are those of space, time, velocity, acceleration, force, momentum, energy, and mass. Among its outstanding objects of study are the motion of fall, planetary motion, motion along an inclined plane, collisions between elastic and inelastic bodies, and the motion of a pendulum. Its representational tools include, apart from ordinary and technical language, those of mathematics and, in particular, the language of ordinary differential equations.

By the end of the nineteenth century electrodynamics (including optics) and thermodynamics had been established by scientists such as James Clerk Maxwell, Heinrich Hertz, and Hendrik Antoon Lorentz, on the one hand, and Joseph Thomson and Rudolf Clausius, on the other, as specialized knowledge structures similar to that of classical mechanics. In fact, although they share some of their most fundamental concepts such as those of space and time with classical mechanics, they also comprise characteristic mental models, concepts and theories distinct from those of mechanics and transmitted within their own subdisciplinary traditions.

Electrodynamics deals not only with the electric and magnetic interaction of material bodies but also with electromagnetic radiation such as light. Among its elementary mental models is that of a charged body generating an electric field in its vicinity. Among its basic concepts are those of conductor, dielectric, electric charge, electric current, voltage, electric and magnetic field. It accounts for physical processes such as the attraction or repulsion between two charged particles, the generation of a magnetic field by a current, and the induction of voltage by a conductor moving in a magnetic field. Partial differential equations and vector analysis represent the prevalent mathematical formalisms used in electrodynamics.

Thermodynamics was formulated as an axiomatic theory in the mid-nineteenth century by Thomson and Clausius; its objects are all physical processes involving heat and its transformations. Among its elementary mental models is that of a heat-converting engine. Among its fundamental concepts are heat, temperature, energy, entropy, pressure, and volume. Physical processes studied by thermodynamics are the behavior of gases with changes of temperature, pressure, or volume, and the transformation of thermal into mechanical energy and vice versa. Partial differential equations are also an important mathematical tool for expressing the laws of thermodynamics.

The three-partite division of classical physics into mechanics, electrodynamics, and thermodynamics did not result in peaceful coexistence, let alone in a stable harmony of these branches and their theoretical foundations.[21] There were many open problems, regarding both the mutual relation and the internal structure of these sub-disciplinary bodies of knowledge. In fact, even within each field of research, a number of alternative theoretical formulations competed with each other. These alternatives distinguished themselves, for instance, by a different deductive organization of the knowledge, by modifying some aspects of the basic conceptual framework, by focussing on different paradigmatic objects, or by asserting different quantitative laws for some phenomena, which sometimes could be empirically tested.

2.2 Foundational Problems of Classical Mechanics

Even the conceptual foundations of classical mechanics were still being debated in the nineteenth century. Since this debate was particularly relevant to the later development of relativity theory, we will take a closer look at some of the foundational problems of classical mechanics as they were seen at the time. The establishment of physical theories other than mechanics demonstrated to students of mechanics in this period that its foundational concepts could not simply be considered as being intimately linked with the structure of human reasoning about nature, as was believed in the eighteenth century under the impression of the enormous success of Newtonian mechanics, by scientists and philosophers such as Jean le Rond d'Alembert and Immanuel Kant. Also the progress of research in mechanics itself, as exemplified by the contributions of Leonard Euler, Joseph Louis Lagrange, William Rowan Hamil-

21 Here the notion of classical physics is used as a historiographically descriptive term; for its contemporary use, see (Staley 2005).

ton, and others in the eighteenth and early nineteenth century, opened up new perspectives on its theoretical and conceptual foundations.

One elementary mental model of the Newtonian theory of motion is that of a discrete material particle moving in empty space (also called "mass point"); mechanical problems could, in general, be considered as being explicable as complex cases of this basic model. Another basic mental model is the "acceleration-implies-force model," introduced above. This model implies that the motions of material bodies can be divided into two classes, those which require a causal explanation by a "force," and those, called "inertial motions" or motions governed by the "principle of inertia," which do not require such an explanation. The latter class of inertial motions includes the state of rest and the rectilinear and uniform motions of particles, whereas all accelerated motions are to be considered as being caused by a force. Since the same motion may be accelerated or uniform depending on the point of reference, this distinction is only well defined with respect to a given laboratory that is located at a specific place at a specific time and that thus represents a "frame of reference." In Newtonian mechanics, the reference frame to be chosen is either the "absolute space" or any frame of reference in rectilinear uniform motion with respect to it, since such a motion makes no difference to the causal classification introduced above. For the frames of reference that are in rectilinear uniform motion with respect to absolute space the designation "inertial frames" was introduced in the late nineteenth century.[22] The conclusion that the laws of Newtonian mechanics should be equally valid in any one of these inertial frames is usually designated as the "Galilean principle of relativity."

If a body is accelerated with respect to an inertial frame, it must be subject to a force exerted upon it by one or more other bodies. The magnitude of its acceleration depends on the magnitude of the force, but also on the resistance with which it reacts to the force. This resistance is, in classical mechanics, an intrinsic, characteristic property of a body called its "inertial mass." In Newton's original conception it was understood as being proportional to the number of elementary particles composing the body.[23] Apart from this inertial mass, bodies also possess another kind of mass according to Newtonian mechanics, their "gravitational mass." In contrast to the inertial mass, the gravitational mass does not measure a resistance to the action of a force but rather represents itself the origin and source of a force, that of gravitational attraction, which, according to Newton's law of gravitation, is proportional to the product of two gravitational masses attracting each other and being inversely proportional to the square of their distance. Newton claimed that a single body may also have an inertial mass, while it possesses gravitational mass only in so far as it interacts with other bodies.[24] But in spite of the very different conceptual function of the two kinds of masses in Newtonian mechanics, they turn out to be exactly proportional to each other and can thus be considered identical (in the following we will therefore simply

22 See (Lange 1883, 1885a, 1885b, 1886).
23 See (Newton 1999, 23–24, 626).
24 See (Newton 1999). For discussion, see (Freudenthal 1986).

speak of an "equality" of inertial and gravitational mass).[25] For the motion of free fall caused by gravitational attraction, this remarkable coincidence has the peculiar consequence that, independent of their mass, all bodies fall with the same speed. In fact, the strength of the gravitational force exerted upon them, which is proportional to their gravitational mass, changes in exactly the same way as the magnitude of their resistance against this force, which is given by their inertial mass.

Even in Newton's time, the conceptual foundations of his theory of mechanics did not remain undisputed. In the nineteenth century, both the further progress in physics mentioned above and the development of a philosophy of science critical of its metaphysical aspects renewed the interest in foundational questions of mechanics. Among the problems discussed were the meaning of Newton's notion of absolute space and the difficulty of identifying the inertial reference frames in which solely the laws of classical mechanics are valid. Since absolute space itself is not directly accessible to experience, an inertial frame of reference can also only be indirectly identified by studying observable physical processes. If, for instance, in a given reference frame the bodies that are not subject to forces are those which are either at rest or in uniform and rectilinear motions, then the given frame of reference is indeed an inertial frame. But, according to the acceleration-implies-force model, the concept of force itself as a cause of motion is only indirectly defined by its connection to the concept of acceleration, which represents its effect. Therefore the presence or absence of forces can in general only be judged by observing and classifying the kind of motions a system of bodies is performing. But this operation in turn depends on the prior establishment of an inertial frame, so that one has apparently entered a vicious circle.

In practical applications of classical mechanics, however, the situation turns out to be less dramatic since here one mostly deals with forces produced by concrete material arrangements, such as a stretched elastic spring, under conditions in which other forces are either known or assumed to play no significant role. The role of inertial frames of reference for the actual functioning of mechanical knowledge may therefore be conceived as that of "laboratory models," i.e. mental models of physical systems, such as a laboratory with an experimental arrangement, for which the laws of classical mechanics turn out or are assumed to be valid. In practice, a laboratory model might therefore also be embodied by systems that are evidently accelerated as is in fact the case for all terrestrial laboratories. But considerations of this kind are hardly suited to respond to the foundational problems as raised by critics of mechanics, in particular in the second half of the nineteenth century. For instance, according to Newton, absolute space should represent an inertial frame of reference even when only a single body is present in the universe. But can it really make a physical difference, in the sense of the presence or absence of forces acting on the single body, whether or not that body is or is not accelerated in an otherwise empty universe?

25 This equality was discussed in the nineteenth century by Heinrich Hertz as a "miraculous riddle," ("wunderbares Räthsel"), see (Hertz 1999, 122).

Critics of mechanics such as the philosopher Ernst Mach argued that it makes no sense at all to apply the concept of motion or even of inertial mass to a single body.[26] He rather suggested that the entire theory of classical mechanics should be reformulated in terms of the relative motions of bodies with respect to each other and that also the concepts of inertial mass and inertial frame should be redefined along these lines. He argued, in particular, that the centrifugal forces curving the surface of water in a rotating bucket should not, as Newton suggested, be interpreted as being due to an effect of accelerated motion with respect to absolute space, but rather as being due to relative motion with respect to the other bodies in the universe ("Mach's bucket"). A few scientists close to Mach's ideas even demanded that the concept of an inertial frame should be given up entirely since a laboratory that is not subject to external forces and thus capable of providing a material embodiment of an inertial frame of reference cannot exist anywhere in the universe, in particular in view of the universal presence of gravitational interaction.[27] They argued that any frame of reference should be equally suitable for formulating the laws of mechanics. But their demand required a fundamental revision of Newtonian mechanics including the prediction of new effects for which they had little empirical evidence. Following Mach's interpretation of the Newtonian bucket experiment, they conceived, for instance, the effects observed in a rotating laboratory as not being due to the centrifugal forces arising because of its acceleration with respect to absolute space or to the inertial frames of classical mechanics. They rather interpreted such "inertial forces" as a manifestation of an hitherto not well understood interaction between physical masses in relative motion with respect to each other, those of the laboratory and those of the distant stars. But attempts to find direct, terrestrial empirical evidence of such an interaction failed.

Other critics of Newtonian mechanics neither worried about the concept of inertial frame nor demanded an entirely new mechanics but still found themselves dissatisfied with the Newtonian notion of force and the role of the interaction between mass points as a basic model of classical mechanics. The development of continuum mechanics and of analytical mechanics in the eighteenth and early nineteenth century in fact offered alternatives to these fundamental assets of mechanics in its original, Newtonian formulation.[28] Most mechanical arrangements do not resemble collections of point-like masses interacting with each other at a distance, as is the case for planetary astronomy. The interaction between the parts of a mechanical arrangement, such as a machine for instance, is realized by the geometric constraints that the arrangement imposes on the motion of its parts. A body moving along an inclined plane is constrained by the shape of the plane to follow its slope, or, to give another example, the motion of a bead along a curved wire is constrained by the shape of the wire. The deviation of this motion from rectilinearity can, in principle,

---

26 See (Mach 1921, chap. 2, sec. 6) or for an English version, (Mach 1960, chap. 2, sec. 6). For further discussion, see "The Third Way to General Relativity" (in vol. 3 of this series).

27 For extensive discussion, see "The Third Way to General Relativity" (in vol. 3 of this series).

28 For a survey see (Pulte 1989; Lützen 1993).

be explained by reconstructing the Newtonian forces exerted by the wire on the motion of the bead, a task that may become rather involved. But it appears much more natural and also more practical to account for the shape of its trajectory directly by the curvature of the wire, in particular as it turns out surprisingly that such an analysis may yield rather simple laws of motion which, under certain conditions, can be interpreted as a generalization of the principle of inertia.

Analytical mechanics as developed by Leonard Euler, Jean-Baptiste le Rond d'Alembert, Lagrange, and others achieves such an explanation by incorporating the geometric constraints of motion, such as the shape of the wire, into the coordinates used to describe the motion, for instance of the bead along the curved wire. In a suitable curvilinear coordinate system, the curved wire may even be described by what is, in these coordinates, the equivalent of a straight line, now being described in terms of a "variational principle" as the shortest or, more generally, an extremal path in such a coordinate system. The motion of the bead may then be considered as a "natural motion" in the sense of generalization of force-free inertial motion, at least when no other forces, apart from those exerted by the geometric constraints, are present. In short, the description of motion in terms of a Newtonian equation of motion on the basis of the acceleration-implies-force model can be replaced by a description on the basis of the "constrained-motion model" according to which the properties of a motion are explained, at least in part, by the geometry of the constraints imposed on the motion. The forces exerted by these constraints can be neglected as they are accounted for by the introduction of suitable coordinates adapted to the problem at hand. In this way, a considerable step was taken towards the elimination of the concept of force as a foundational concept of classical physics. In order to ban it completely from mechanics, however, one would have to conceive of all forces as being due to rigid constraints. An attempt in this direction was undertaken in the late nineteenth century by Heinrich Hertz who reformulated classical mechanics entirely in terms of such constraints, including invisible ones that were supposed to replace the forces known from ordinary mechanics.[29] However, because of the speculative character of the invisible constraints, this reformulation of classical mechanics did not find many followers.

2.3 Invisible Mechanisms and the Expansion of Mechanics

Hertz's rigid constraints may serve as an example for an "invisible mechanism." Such mechanisms are a characteristic tool for extending the range of applicability of a domain of knowledge beyond the set of problems to which it is directly applicable. Invisible mechanisms are usually conceived in terms of mental models from familiar territories of physical knowledge, in particular also from intuitive physics, such as the model of a rigid body or of a fluid. Under the assumption that what they actually represent is, for a good reason, only partly or indirectly accessible to sense experience,

29 See (Hertz, 1894, 1956).

as is the case for atoms, for instance, they may become the fundamental constituent of an all-encompassing theoretical framework such as atomism. This type of extension of a conceptual framework beyond its immediate range of application was a well-established practice within the tradition of mechanics since the early modern period. In fact, the emerging new science of mechanics had to compete with the all-encompassing Aristotelian word view embraced by the Church, a circumstance that proved to be a major challenge encouraging the elaboration of a mechanical world-view.[30] Furthermore, for a considerable period of time mechanics was without serious alternatives as a theoretical foundation of physics. In any case, the construction of invisible mechanisms involving mechanical models and knowledge remained a common practice throughout the reign of classical science.

The explanation of heat by the motion of atoms in the nineteenth century is a well-known example for such an extension of mechanics beyond its immediate range of application.[31] According to the kinetic theory of heat atoms were conceived according to the mental model of the discrete, more or less rigid bodies of our macroscopic experience, but were taken to be unobservable due to their smallness. Only their bulk effects such as heat or pressure were assumed to be directly accessible to sense experience. The kinetic theory of heat succeeded in representing the relation between the heat, temperature, pressure, and volume of a gas, for instance, by the relation between the motion of these atoms, their energy, momentum, number, etc., that is, by relations determined by mechanical and statistical laws. In this way, mechanics could be conceived as offering a theoretical foundation for thermodynamics as well. Since the early days of classical physics, invisible mechanisms were also considered in order to explain actions at a distance such as that of gravitation, electricity, magnetism, or the transmission of light, which, from the point of view of ordinary experiences with mechanical arrangements, appeared as particularly mysterious.

Invisible mechanisms for the explanation of the transmission of light, for instance, were conceived in terms of the motion of minute particles (in particular by Newton and his immediate followers) but also (first by Huygens and then by most physicists in the second half of the nineteenth century) in terms of the motion of an invisible continuous medium, the optical "aether."[32] The latter was understood according to the mental model of a medium such as water or air, as is familiar from intuitive mechanics. These media demonstrated how the transmission of a physical effect—like a water wave or sound—could be conceived without the transport of material, just by the propagation of a local perturbation from one point to the next. An invisible mechanism based on the model of a continuous medium was therefore, at least in principle, a plausible candidate for explaining the transmission of those physical effects which are not associated with any noticeable transport of material, such as the actions at a distance mentioned above. In the case of light, such an explanation

30 See, e.g., (Renn 2001; Montesinos and Solís 2001).
31 For a historical survey, see (Brush 1976a, 1976b).
32 For historical discussions, see (Cantor and Hodge 1981; Buchwald 1989).

received decisive additional support in the first half of the nineteenth century by the discovery of the wave properties of light, which follow naturally from its explanation as a perturbation of the optical aether.

2.4 The Problem of the Motion of the Optical Aether

By its very nature, an invisible mechanism has more properties than those which are directly involved in the physical processes to be explained by it. Even though, for instance, the optical aether itself was assumed not to be directly accessible to ordinary mechanical experience, it follows from its conceptualization as a mechanical medium that it must be in motion or at rest with respect to the macroscopically observable bodies immersed in it. Now it turns out that, for the optics of moving bodies, the different possible answers to the question of the state of motion of this invisible medium even have distinct observable consequences. In principle, it is possible that the aether locally shares the motion of a body (an assumption proposed by George Gabriel Stokes), or that it remains entirely unaffected by it (an assumption pursued by various scientists such as Thomas Young, Augustin Jean Fresnel, and Hendrick Antoon Lorentz), or that it is partially dragged along by the motion.[33]

A strong argument for an immobile aether was the phenomenon of stellar aberration known since the eighteenth century. If one observes the position of a star in the course of the year from different positions along the orbit of the earth, it is subject to certain regular variations, which if the star is at great distance cannot be due to its parallax, i.e., to the changing line of vision. These variations can, as James Bradley recognized at the beginning of the eighteenth century, rather be explained by the fact that the speed of light originating from the star is composed with the speed of the earth's motion in such a way that the starlight seems to originate from different directions in dependence on the direction of the earth's motion. However, on closer inspection, this simple explanation of aberration requires two specifications: first, if light is assumed to be carried by the aether, the composition of velocities explaining the aberration works in a straightforward manner only if it is further assumed that the aether is immobile; second, if propagation of light through a medium such as the glass of a telescope or water is taken into account, one has to deal with the fact that the speed of light in such media differs from that in the aether so that the effect of aberration should change if stars are observed through them. Observation shows, on the other hand, that aberration is completely independent of whether or not light passes through a transparent medium.

At the beginning of the nineteenth century, Fresnel explained this fact by assuming that transparent media, which move along with the earth through an otherwise immobile aether, carry it along with a certain fraction of their speed. But what exactly does this partial drag of the aether mean? Can it perhaps be observed directly or is it

33 For an extensive historical discussion of the problem of the optics and electrodynamics of moving bodies, see (Janssen 2004; Stachel 2005).

merely a hypothetical compensation in order to explain the absence of deviations from the normal aberration? This question was at the center of an experiment performed by Hippolyte Fizeau in 1851. He studied the propagation of light in flowing water in order to determine the variation of the speed of light in dependence on the speed of the moving medium. His experiment yielded a relation between the speed of light in water and the speed of the flowing water, which indicated a slight drag of the aether by the water, thus roughly confirming Fresnel's aether-drag hypothesis. Towards the end of the nineteenth century, the assumption of an immobile aether became generally accepted as the most plausible assumption about its state of motion, in particular after optics was included in Maxwell's electromagnetic theory, a development to which we shall return. On the basis of this assumption the problem then was to explain the apparent slight drag indicated by the peculiar way in which the speed of light in the flowing water was composed of the ordinary speed of light in the water at rest and the speed of the flowing water.

But a theory based on the assumption of an immobile aether faced much more serious problems. An immobile aether had to be conceived as providing itself a universal frame of reference for the motion of all other bodies, an embodiment of Newton's absolute space as it were. In short, it turned out that electromagnetism, as was the case for mechanics, also had unavoidable "cosmological" consequences. As the aether is a medium with concrete physical properties, the motion of a body with respect to this aether frame should somehow have noticeable effects. Now, if that is indeed the case, the aether model cannot be compatible with the principle of relativity of classical mechanics, according to which physical laws should be the same in two laboratories representing inertial frames and which are moving with rectilinear and uniform motion with respect to each other. Since the aether model entails that the speed of light with respect to the aether must always be the same, independently of the state of motion of the body emitting the light, such a violation of the relativity principle turns out to be an unavoidable consequence of this model. If, for instance, the speed of light is measured in two laboratories in relative motion with respect to each other, the two measurements cannot yield the same result, as at least one of the two laboratories must necessarily move with respect to the aether. An even simpler arrangement would be the measurement, in a single terrestrial laboratory, of the speed of a light ray once propagating in the direction of the earth's motion through the aether and then of the speed of the same ray when deflected by a mirror so that it propagates perpendicular to the direction of the earth's motion. An experiment along these lines was in fact performed by Michelson and Morley with the result that the speed of light is unaffected by the motion of the earth through the aether. Here then the explanation of optical effects by an aether model met with a very serious paradox.[34] On the one hand, the phenomenon of aberration and Fizeau's experiment required the aether to be essentially immobile, while, on the other hand, motion

34 For discussion, see (Janssen and Stachel 2004).

through the aether did not have the physical effects it should have had according to the aether model.

2.5 Foundational Problems of Classical Electrodynamics

In the course of the development of electrodynamics, light was recognized by Maxwell as a particular case of electromagnetic radiation, and optics henceforth became part of electrodynamics. As the electrodynamic theory founded by Maxwell was also based on an aether model, the electrodynamic aether now took over the role of the optical aether. The problems of the motion of the optical aether discussed above therefore became part of a knowledge area commonly referred to as the "electrodynamics of moving bodies." Although the aether represents an invisible mechanism based on mechanical concepts, it turned out to be extremely difficult, if not impossible, to construct a concrete mechanical model of the aether reproducing all physical properties required by the wealth of electromagnetic and optical experiences covered by Maxwell's theory. Even just the explanation of the properties of light required an aether that has the mechanical properties of an incompressible, jelly-like elastic medium and yet represents no obstacle to the motion of huge physical masses such as the earth or the sun.[35]

In fact, however, the aether could serve at least some of its essential explanatory functions, such as rendering comprehensible the transmission of electromagnetic waves, even without being exhaustively described in terms of the laws of classical mechanics. Certain aspects of the aether could instead be captured by the laws of electrodynamics themselves. Thus one could imagine that in the presence of an electric charge, the aether changes its state like an ordinary non-conducting, dielectric medium, which, according to the laws of electrodynamics, becomes "polarized." In this way, however, the aether model ceased to serve as a tool for linking electrodynamics to a mechanical foundation. It rather became itself part of an independent foundation of electrodynamics but nevertheless retained some of the properties due to its roots in intuitive physics, for instance, of having a state of motion—*this being the state of rest*.[36]

By the end of the nineteenth century the most developed theory of electrodynamics —by Hendrik Antoon Lorentz—was in fact based on an aether model whose dynamical behavior was entirely governed by electrodynamical laws, while it was assumed to be absolutely stationary.[37] Lorentz's aether was not even subject to Newtonian laws of motion. In fact, it exerts forces on bodies within it—without being affected itself by a counter-force as required by Newton's law of the equality of action and reaction. In order to describe the electromagnetic behavior of ordinary

35 For studies concerning the concept of aether in classical physics, see (Whittaker 1951–53; Hirosige, 1966; Schaffner, 1972; Buchwald 1985; Darrigol 1994; Warwick 1995; Janssen 2003).

36 The observation that the assumption of the aether being immobile amounts to the assignment of a mechanical property is due to Einstein, see (Einstein 1920).

37 For studies of Lorentz's aether theory, see (Hirosige 1969; Janssen 1995).

matter, which cannot escape the laws of mechanics, Lorentz's theory employed another invisible mechanism in addition to the aether. Matter was conceived to be made up of elementary particles carrying electric charges. While the particles themselves were modelled after the discrete bodies of ordinary mechanical experience and were hence also subject to the laws of mechanics, their charges were supposed to interact with the aether according to a mechanism similar to the interaction between charges and dielectric medium in macroscopic electrodynamics. The presence of charges generates a particular state of the surrounding aether, corresponding to an electromagnetic "field," which then propagates through the aether and thus can affect other charges. The introduction of the field concept had been the result of reflection on the experiments by Hans Christian Oerstedt, André-Marie Ampère, Michael Faraday, and others which showed that interactions involving the environment beyond the connecting line between two particles were characteristic of electromagnetic effects.[38] According to Lorentz's theory, charges interact with each other only via the aether, by generating fields and by sensing a "ponderomotive force" exerted by the field upon the elementary particles carrying the charges. By way of its complicated invisible mechanisms, which refer to two entirely distinct entities, the aether and the elementary particles, Lorentz's theory achieved an integration of the laws of electrodynamics with those of mechanics, albeit at the price of a fundamental dualism.

2.6 The World Pictures of Classical Physics and the Emergence of Borderline Problems

The enormous success of late nineteenth century electrodynamics in explaining physical properties of matter as well as of radiation stimulated some contemporary scientists to search for an electromagnetic foundation of all physics.[39] One could in fact imagine invisible mechanisms, now conceived in terms of electromagnetic processes, offering mental models to understand experiences hitherto explained in terms of mechanics. For instance, it is well known from electrodynamics that an electrically charged body, set in motion by some force, generates a magnetic field in its vicinity which results in an increased resistance against the moving force. Since this resistance adds to the inertial resistance, contemporary scientists wondered whether this process might perhaps serve as a model to understand the concept of inertial mass in purely electromagnetic terms as a kind of resistance against a moving force, thus underpinning a central concept of mechanics with an electromechanical foundation.

Other features of classical mechanics were even more suited to speculations about a possible electromagnetic origin. Since in the case of electricity, magnetism, and light, electrodynamics had been so successful in accounting for actions at a distance by an aether model, it was plausible to also attempt an explanation of the gravita-

---

38 For a survey, see (Darrigol 2000). See also (Steinle 2005).

39 For historical discussion of the electromagnetic foundations of physics, see (Jungnickel and McCormmach 1986a, 227–245).

tional action at a distance by means of the same intervening medium.[40] Newton's law of gravitational attraction between two masses even has the same form as Coulomb's law for the force between two electrical particles at rest. In both cases the force is inversely proportional to the square of the distance between the particles and proportional to the product of the masses in the first case, and the charges in the second. On the background of the electromagnetic aether model, in particular as developed by Lorentz, it was therefore natural to interpret the mass as representing a "gravitational charge" accounting for a perturbation of the surrounding aether and resulting in a "gravitational field," which is then transmitted through the aether and thus acts upon another gravitational charge.

Apart from the difficulty in developing a consistent model for an aether carrying such a gravitational field, either in addition to the electromagnetic fields or somehow produced by them, an aether model of gravitation also implied hitherto unknown features of the gravitational interaction. Consequences of an aether model for gravitation such as a finite speed for the transmission of gravitational interactions and the possible existence of gravitational waves were not only new with respect to classical mechanics but also found little support in the empirical knowledge on gravitation available at the turn of the century. Developing an aether model for gravitation was hence an unwieldy and speculative task comparable to constructing the entire theory of electrodynamics when only Coulomb's law of electrostatic interaction is known.[41]

Scientists in the second half of the nineteenth century not only attempted to construct an electromagnetic foundation of physics but also explored the possibility of extending the theoretical foundation of thermodynamics to the entire body of knowledge of classical physics.[42] The starting point for this line of research was the observation that in all branches of contemporary science the concept of energy played a central role and that different kinds of energy may be transformed into each other, mechanical energy into thermal energy, chemical energy into electrical energy, and so forth. The aim of this research program was therefore to understand all physical, if not all natural processes as transformations of energy. In this way, by overcoming the separation of the objects of the different branches of science, such as bodies in motion, electricity, or heat, all now conceived as different manifestations of energy, the separation of the branches themselves should also be superseded.[43]

In short, the different theoretical foundations of late nineteenth-century physics could be and were in fact conceived as being potentially alternate foundations for the entire body of knowledge of classical physics. Contemporary scientists thus discussed a mechanical "world picture," an electrodynamic world picture, and a thermodynamic world picture. Nevertheless, the dynamics of the intellectual development of physics in this period continued to be governed mainly by the many concrete

40 For further discussion, see the introduction to vols. 3 and 4 of this series "Theories of Gravitation in the Twilight of Classical Physics" (in vol. 3).

41 This observation is due to Einstein, see (Einstein 1913 1250).

42 For a historical discussion, see (Jungnickel 1986a, 213–227).

43 For historical discussion, see (Kuhn 1977, chap. 4, 66–104; Elkana 1974).

unsolved problems—both theoretical and empirical in character—on which the scientists were working. By the end of the century some of these problems also emerged as being related to more than one theoretical foundation. One such problem was the electrodynamics of moving bodies, which required the application of both the laws of electrodynamics and the laws of motion of mechanics. Another example of this class of borderline problem is heat radiation, which required the application of both the laws of radiation—covered by the laws of electrodynamics—and the laws of thermodynamics. Since these problems fall under the range of application of two different theoretical foundations, they represented not only a potential locus of conflict between fundamentally different conceptual frameworks, but also points of departure for their integration into more developed theoretical frameworks. Thus Planck's law of heat radiation was later seen as the first decisive contribution to quantum theory,[44] while the electrodynamics of moving bodies became, as we shall see, the core of the special theory of relativity.

At first sight, however, the problems of the optics and electrodynamics of moving bodies, as discussed previously, may merely appear to be due to the unfortunate choice of a mechanical aether as an invisible mechanism for electrodynamic explanations. But even Lorentz's renunciation of ascribing any mechanical properties to the aether did not help avoid the paradoxes of the propagation of light through the aether described above. He therefore had to modify his otherwise very successful theory of electrodynamics by additional assumptions about the effect of the aether on bodies moving through it in order to explain the lack of any noticeable effect due to the motion of a terrestrial laboratory through the stationary aether. If one considers, on the other hand, the absence of any noticeable "aether wind," not only from the perspective of the internal problems of electrodynamics, but from that of classical physics as a whole, it emerges in fact as a borderline problem between electrodynamics and mechanics. The absence of any aether wind due to the translatory motion of a laboratory through the aether is simply a confirmation of the principle of relativity of classical mechanics, according to which such a motion should indeed be unnoticeable, at least if acceleration is neglected. But this confirmation takes a form which itself is not compatible with classical mechanics, since the absence of an aether wind requires that the motion of the laboratory has no effect on the speed of light measured in the laboratory. This conclusion is in direct conflict with all ordinary experiences with the speeds of moving bodies, whose magnitude depends, of course, on the state of motion of the observer performing the measurement. The resolution of this paradox eventually required the abolition of the aether model rooted in intuitive physics, and with it, a complete revision of the concepts of velocity, space, and time.[45]

44 See (Büttner et al. 2003).

45 For discussion of the electrodynamics of moving bodies as a problem of classical physics, see also (Renn 2004, Rynasiewicz and Renn 2006).

2.7 The Crisis of Classical Electrodynamics

The existence of partially distinct conceptual frameworks within classical physics provided the substratum for a multiplicity of perspectives not only on its foundational questions but also on various specific research problems, and in particular on the borderline problems mentioned above. We have characterized the electrodynamics of moving bodies as a borderline problem, bringing to light a basic conflict between the notion of electromagnetic aether and the relativity principle of mechanics. But from within the conceptual framework of electrodynamics this conflict may have been perceived differently by contemporary scientists, namely as an unsolved puzzle of perhaps even minor significance. The undetectability of an aether wind could thus appear to them not so much a confirmation of the well-known relativity principle of classical mechanics but rather the net result of several causes conspiring to hide the presence of effects due to the motion of the earth through the aether.

Correspondingly, it was natural for scientists in the late nineteenth century to attack this problem by exploiting the conceptual and technical resources of electrodynamics. Lorentz introduced, in addition to assuming a "local" time characteristic of a physical system moving with a certain speed through the aether, the hypothesis of a length-contraction of bodies along the direction of their motion ("Lorentz contraction") in order to explain the "null-effect" of the aether wind on electrodynamic and optical measurements. The exclusive aim of this hypothesis was to introduce a mechanism compensating the expected but unobserved effects of an aether wind; this consideration also determined the quantitative specification of the length-contraction. Lorentz's qualitative justification for his hypothesis was the behavior of the molecular forces within bodies under motion through the aether, which he tentatively assumed to be analogous to that of the electric and magnetic forces.[46]

From the perspective of the mechanical framework of classical physics the same problem appears quite differently, as indicated above. The carefully and with much experimental effort established null effect, which was so difficult to explain within an electrodynamic theory based on an aether model, turned out to be simply another confirmation of the time-proven relativity principle of classical mechanics. Consequently, it was natural, within this framework, to discard the aether model as the source of the difficulties and to construct instead an electrodynamics based on the relativity principle. While for such an approach the null effect represented no challenge at all, it was confronted, on the other hand, with the task of explaining, in its terms, results that had constituted the triumph of an aether-based electrodynamics and optics since the early nineteenth century, such as the interference of light waves or the Fizeau experiment discussed above.

Since in the beginning of the twentieth century the wave theory of light encountered difficulties, also in other fields of enquiry but in particular regarding the interaction of light and matter, it seemed worthwhile to several scientists of this period, most

46 See (Lorentz 1892, 1895). For discussion see, e.g., (Janssen 1995, sec. 3.2).

notably to Albert Einstein and Walter Ritz, to tentatively reexamine and modify Newton's corpuscular theory of light as a framework within which to explain the new empirical knowledge.[47] Indeed, a corpuscular emission theory of light succeeded not only in accounting for the role of the color of light when setting free electrons of the irradiated matter. This effect was difficult to explain for a wave theory but was concisely described by Einstein's hypothesis of corpuscular quanta, whose energy is related to the color of light.[48] An emission theory was also a suitable basis for incorporating the relativity principle into the foundations of a theory of radiation and perhaps of electrodynamics in general.

The velocity of the light particles should not differ from the way velocities of other particles behave since its magnitude must, of course, depend on the state of motion of the reference frame in which it is measured. The affirmation of a "constancy of the speed of light" can, in this conceptual framework, only refer to the speed measured with respect to the source of light and not to any property independent from the frame of reference. There was, in particular, no need for any "absolute" frame of reference such as that provided by a stationary aether in order to give meaning to this constancy of the speed of light. But apart from the extraordinary difficulty in reconstructing electrodynamics and optics on the basis of an emission theory of light, this approach encountered additional obstacles in explaining other well-established aspects of scientific knowledge such as the properties of reflected light.[49] The available shared scientific knowledge thus decided against an emission theory of light although it evidently represented a promising approach to the special problem of the electrodynamics of moving bodies.

However, also an assessment of Lorentz's solution to this problem depended on the extent to which shared knowledge was taken into account for its evaluation. In fact, his proposal appears even more shaky when viewed not only from the internal perspective of classical electrodynamics but also in the light of new knowledge generated by another borderline problem: heat radiation. In the early twentieth century, the experimental and theoretical studies of heat radiation were recognized as being in conflict with an aether model such as the one assumed by Lorentz. According to the so-called "equipartition theorem" of the kinetic theory of heat, such an aether in thermal equilibrium with matter, should carry electromagnetic radiation of all possible frequencies, with each frequency being attributed an equal share of the total. In fact, the velocity of the light particles should not differ. But since there is, in principle, no limit to the frequencies at which the aether can vibrate, contrary to the case for ordinary matter, the aether with its infinite degrees of freedom will eventually absorb all the energy of a physical system so that a thermal equilibrium between aether and matter becomes impossible.[50]

---

47 For historical discussion, see the editorial note "Einstein and the Theory of Relativity" in (CPAE 2, 253–274). For the wider context of the reexamination of corpuscular theories of radiation, see also (Wheaton 1983).

48 See (Einstein 1905b).

49 For discussion, see (CPAE 4, Doc. 1, 35).

Hence neither the conceptual framework of electrodynamics nor that of mechanics offered a straightforward solution to the puzzles raised by the electrodynamics of moving bodies. Nevertheless, within the first framework—an aether-based electrodynamic field theory—it was at least possible to develop a somewhat roundabout solution by twisting and stretching this framework to its limits, which is what Lorentz did with his ever new additional auxiliary hypotheses. However, due to this protraction of the theory by auxiliary hypotheses, not only its anchoring in an aether model, but also its relation to the conceptual foundation of classical physics, had become loose. In particular the introduction of special variables describing lengths and time intervals in a moving system augmented the formalism of Lorentz's electrodynamics by elements for which a direct physical interpretation in terms of this foundation had become difficult. And yet the formulae developed by Lorentz to describe the electrodynamics of moving bodies were themselves compatible with all available knowledge, even with the principle of relativity from mechanics, which was otherwise at odds with the foundations of an aether-based electrodynamic theory.

2.8 Birth and Early Development of the Special Theory of Relativity

The elaboration of a consistent electrodynamics of moving bodies on the basis of an aether model was achieved mainly due to the work of Lorentz and Poincaré. As pointed out, the final results of their efforts constituted a new nucleus of knowledge whose linkages to the conceptual foundation of classical physics were problematic. Of course, precisely how problematic these linkages actually appeared to contemporary scientists depended on the range of shared knowledge they could or wished to take into account for their judgement, as we can see from the relevance to this judgement of such apparently unrelated phenomena as the thermal equilibrium of heat radiation. In other words it was the knowledge of classical science as a whole that provided a resource for reflection that could be used in different ways. Such reflections took on the form of different "perspectives" on this knowledge.

In the case at hand, it was precisely the decoupling from the roots of classical physics of the new body of knowledge constituted by the electrodynamics of moving bodies that created the possibility of a novel perspective on the conceptual foundations of physics, in addition to the perspectives rooted in mechanics, electrodynamics, and thermodynamics, respectively. In fact, as we have seen, Lorentz's auxiliary variables describing space and time intervals in a moving physical system could not fully derive their meaning from the concepts of space and time as they were basic to classical physics. Therefore, if the outcome of Lorentz's efforts was itself being considered as anchoring a conceptual framework, the new space and time variables could be conceived as gaining their meaning, not via a specification of these general classical concepts, which anyway was problematic, but primarily from their role within the electrodynamics of moving bodies. But this limited body of specialized knowledge

50 See the discussion in (Einstein 1905b).

was, on the other hand, in itself insufficient to equip the new space and time variables with the meaning of general space and time concepts, at least as long as their implicitly defined meaning within this body of knowledge could not be related to the more universal meaning of these concepts in physical thinking.

Lorentz's theory thus offered a natural starting point for a process of reflection that is repeatedly observable at decisive moments in the history of science. This process implies that marginal elements of a complex architecture of knowledge torn asunder by inner tensions become the starting points for a reconstruction as a result of which a new structure emerges that is, however, essentially composed of the building blocks already available. This process may be designated as a "Copernicus Process." Indeed conceptual turnovers unfold in a similar way to the revolution of Copernicus, who created a new world system by placing an initially marginal celestial body—the sun—into the center, while, instead of starting with a *tabula rasa,* retaining the previously elaborated complex machinery of planetary astronomy. Similarly, the central role of the aether and peripheral role of the new variables for space and time in Lorentz's theory is essentially reversed in special relativity; here the concept of aether no longer plays a role while Lorentz's auxiliary quantities become the new foundational concepts of space and time. Most mathematical details, on the other hand, in particular the so-called Lorentz transformations between reference frames in uniform motion with respect to each other, remain unaffected by this displacement of the conceptual center.

The special theory of relativity, published in 1905 by Einstein,[51] shows that it is indeed possible to embed the new space and time variables of the electrodynamics of moving bodies within such general physical concepts of space and time, with the consequence of also restructuring these general concepts. Einstein carefully defined space and time measurements in a given frame of reference with the help of an elementary procedure conceived in terms of measuring rods and clocks. This procedure involves only a minimum of well-defined physical assumptions, such as the existence of rigid measuring rods, the possibility of synchronization procedures between clocks, etc. In fact, it is the peculiar feature of this procedure that, on the one hand, it is compatible with and makes precise some essentials of our ordinary understanding of space and time measurements but that, on the other hand, little else follows from it without the addition of further assumptions. In particular, it entails nothing about the relation between space and time measurements performed in frames of reference in relative motion to each other. Einstein thus succeeded in demonstrating that the concepts of space and time in classical physics involve at least two layers of knowledge, one that involves elementary mental models of rods and clocks, and a second, clearly distinguished one that introduces the additional theoretical assumptions required for the universal time and space concepts, equal for all physical systems whether in motion or not, which underlie classical physics. Disentangling these two levels he could now replace the layer of additional physical requirements leading to the classi-

51 See (Einstein 1905c).

cal concepts of space and time, such as the assumption of an infinite signal speed or the assumption that the simultaneity of events is independent of the state of motion, by the layer representing the physical requirements embodied by the electrodynamics of moving bodies.

Einstein thus introduced the notion of a "relativity of simultaneity" and identified the ordinary addition of velocities as not pertaining to the elementary level of the concepts of space and time. It was just this ordinary and seemingly self-evident addition rule, however, which led to the incompatibility between the constancy of the speed of light, supported not only by the Fizeau experiment but essentially by the entire body of knowledge incorporated in Lorentz's electrodynamics, and the principle of relativity from classical mechanics. The assumption that velocities along one and the same direction can simply be added or subtracted like numbers in fact immediately leads to the conclusion that, if the velocity of light has a certain constant value in one reference system, it cannot have that same value in another system in relative motion to the first one, but must be augmented or diminished in dependence on the direction of motion. But on account of the epistemological insights associated with the introduction of the special theory of relativity, one could abandon this addition rule as part of the classical understanding of space and time and yet retain the more elementary level of mental models of space and time measurements, on which one could now build differently structured concepts. It turned out to be possible to introduce the constancy of the speed of light, along with the requirements imposed by the relativity principle, as a basic assumption of the new conceptual framework from which then a new addition rule for velocities could be inferred. This addition rule makes it possible to avoid the contradiction between the constancy of the speed of light and the relativity principle, which among other things requires that velocities of bodies depend on the state of motion of the reference frame from which they are observed. Indeed, for the velocities of ordinary material bodies the new addition rule comes close to the classical one, while the speed of light always retains its constancy; the new addition rule furthermore implies that the speed of light plays the role of a limiting velocity for physical interactions. In this way, the special theory of relativity had achieved a synthesis between the foundation of classical electrodynamics, from which the principle of the constancy of the speed of light was taken, and the foundation of classical mechanics, from which the principle of relativity was inherited.

The new concepts of space and time of the special theory of relativity of 1905 are built from combining Einstein's elementary procedure for measuring space and time by rods and clocks, which determine space and time only in one reference frame, with transformation equations relating the measurements obtained in one frame to those in another frame. The specifics of this new, higher-level structure of the concepts of space and time are entirely determined by the knowledge extracted from classical mechanics and classical electrodynamics. From classical electrodynamics it follows, in particular, that the speed of light should be constant at least in one frame of reference. From classical mechanics it follows that the admissible frames should be inertial frames of reference and that the basic physical laws, and hence also the constancy of the speed of light, should be the same in all such frames. From these requirements the so-called “Lorentz transformations” can be derived, which relate the space and time variables in one inertial frame of reference to those in another one.

While similar transformations were initially derived by Lorentz as an artifice for solving internal problems of the electrodynamics of moving bodies, their refinement and reinterpretation by Einstein led to new concepts of space and time whose potential range of applicability was now the entire domain of physics. These new concepts were already so closely interwoven with electrodynamics that it turned out to be rather unproblematic to implement them in this area of knowledge, with little need for major restructurations. The foundational equations of electrodynamics, due to Maxwell and Lorentz, did not have to be modified in order to be compatible with the new prescriptions for transforming physical variables from one reference frame into another. However, this was not the case for Newton’s dynamics which did require such modifications. This difference is not surprising since classical electrodynamics was already conceived as a field theory in which physical actions travel with speeds limited by the speed of light, as required by the special theory of relativity. Newton’s dynamics, on the other hand, allows for instantaneous actions at a distance, such as that of gravitation, and conflicts with the requirement of special relativity that no physical interaction can propagate with a speed greater than that of light. Nevertheless, in the course of a “mopping up” operation in the years following Einstein’s formulation of special relativity, it turned out to be possible to adapt greater parts of classical physics to the new concepts of space and time. The assimilation of ever larger portions of physical knowledge to these new concepts provided them in turn with an ever growing stability, reflected also by the quick acceptance of the special theory of relativity among contemporary physicists.

This restructuration of physical knowledge by new concepts was greatly eased when in 1908 Hermann Minkowski developed a mathematical formalism for representing the totality of physical events in space and time, together with the relations between these events as implied by the special theory of relativity.[52] This mathematical formalism consists of a four-dimensional manifold in which coordinate systems, representing the physical frames of reference, can be employed to characterize each physical event by four numbers, its three space and one time coordinate. If the time coordinate is introduced in such a way that it is always multiplied by the constant speed of light, it also acquires the physical dimension of a spatial variable so that Minkowski’s four-dimensional world becomes even more similar to the three-dimensional Euclidean space familiar from classical physics, being, however, extended by one more dimension.

The mathematical analogy becomes almost complete if in this space the notion of a “metric” is introduced, allowing a determination of the distance between two events in space and time. In ordinary Euclidean space a metric has the immediate physical significance of giving the spatial distance between two points. It can be expressed in

52 See (Minkowski1908).

terms of a “line element,” determining the distance ds between two infinitesimally close points in terms of their infinitesimal coordinate differences dx, dy, dz. In the case of the usual orthogonal Cartesian coordinates one thus obtains according to the Pythagorean Theorem:

$$ds^2 = dx^2 + dy^2 + dz^2.$$

In the case of more general coordinates or of a non-Euclidean three-dimensional space, the line element is given by a more general expression:

$$ds^2 = \sum_{i,j=1}^{3} g_{ij} dx_i dx_j$$

where g_{ij} are the components of the so-called “metric tensor,” which in general may even be functions of the coordinates. In the case of the usual coordinates of the four-dimensional Minkowski’s geometry, the line element is given by:

$$ds^2 = dx^2 + dy^2 + dz^2 - c^2 dt^2,$$

corresponding to the components of a metric tensor with constant components which, represented in matrix form, are given by:

$$g_{ij} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -c^2 \end{bmatrix},$$

where c^2 is the square of the speed of light. In Minkowski’s four-dimensional world, line element and metric do not possess an immediate interpretation in terms of the concept of length; the line element can, for instance, represent a spatial interval in one case and a time interval in another. But line element and metric do represent in a concise way the causal structure of physics according to special relativity. In particular, an event that can causally affect another one must have a distance that is negative or zero as measured according to the metric of Minkowski’s world since two events can only be causally connected by interactions that propagate with a speed less than or equal to that of light.

In addition to the characterization of the causal structure of physics by the metric, the introduction of a distance in the four-dimensional world of spacetime events also offers the possibility of giving a succinct characterization of the Lorentz transformations. In ordinary three-dimensional Euclidean space the notion of a distance enters into the definition of rotational transformations between two coordinate systems. In fact, they can be defined as those linear and homogeneous transformations which leave the distances between arbitrary points invariant. It now turns out that, with the help of the four-dimensional metric that defines distances in Minkowski’s world, Lorentz transformations can simply be defined in terms of four-dimensional rotations and hence acquire geometrical significance.

The mathematical representation of events in space and time by "Minkowski space" or "Minkowski spacetime," as this four-dimensional manifold is also called, and the representation of Lorentz transformations by rotations in this space has far-going implications for the mathematical representation of other physical quantities and relations as well. One immediate consequence is the characterization of the motion of a point particle by a four-dimensional "world-line" in Minkowski space, representing the sequence of events in space and time corresponding to this motion. The usual kinematic description of the motion of a point-particle is thus replaced by a description in terms of the geometric properties of such a line. In a sense, the dynamics of physical changes in a three-dimensional world now becomes a "statics" in the four-dimensional representation. Minkowski thus formulated the requirement that in the mathematical description of physical processes, the time coordinate should not play an exceptional role, but rather be treated on an equal footing with the other three coordinates.

The motion of a particle corresponding to a specific world-line may of course, still be given in terms of relations between its coordinates referring to a given coordinate system, for instance by prescribing the spatial position of the particle with respect to this system as a function of the local time in this system. But the geometric properties of the particle's world-line in the four-dimensional Minkowski space remain independent of any specific coordinate system. All that changes, if another coordinate system is introduced in the four-dimensional world, is the relative position of the curve with respect to this new system, and hence the description of the curve in terms of space and time coordinates. One such coordinate representation results from another one by a Lorentz transformation, provided that both coordinate systems correspond to inertial frames of reference. The use of Minkowski's four-dimensional "spacetime continuum" as a framework for physical theories naturally leads to the requirement that physical magnitudes should be represented by geometric objects in this framework. This requirement automatically ensures that the representation of such a physical magnitude in terms of space and time coordinates referring to a given reference frame is related to the coordinate representation with respect to another frame by a Lorentz transformation. In this case the physical magnitude may be characterized as being "Lorentz invariant."

Minkowski's formalism thus provided not only an elegant representation of the mathematical properties of the Lorentz transformations, but also a heuristic framework guiding the reformulation of physical knowledge in terms of the new concepts of space and time. It offered, on the one hand, a mathematical criterion for immediately deciding whether or not a given physical quantity is Lorentz invariant, and thus conforming to the principles of special relativity; on the other hand, it allowed the creation of a multitude of mathematical objects which were candidates for representing physical magnitudes or laws within special relativistic physics. On the background of Minkowski's formalism it was possible, for instance, to recognize immediately that Newton's law of gravitation is not Lorentz invariant—Newton's expression for the gravitational force does not correspond to a geometric object in Minkowski space—and thus needs to be adapted to the principles of special relativity. The same formalism immediately suggested, on the other hand, a variety of Lorentz invariant force laws that could all be conceived as candidates for a special relativistic modification of Newton's law of gravitation.[53]

Special relativity, in particular after its reformulation by Minkowski, was largely perceived as founding a new kinematics, in the sense of a new spatio-temporal scaffolding for the description of physical processes, which replaces the "absolute" space and time concepts of classical physics. Although the new scaffolding was initially constructed on the basis of knowledge acquired mainly in a specific subdomain of electrodynamics, it tended to be perceived as having or pretending to have the same quasi *a priori* status for physical science as classical kinematics. While in the future physical research would uncover new forces and hence new dynamical relations, it was nevertheless expected that these would fit into the kinematic framework established once and for all by special relativity; in this sense special relativity was conceived as structuring new physical experience in terms of space and time rather than as being itself potentially affected by it. But even in the first years after its formulation the theory already led to far-going conceptual changes in classical physics, beyond the domain of kinematics. Naturally, the dynamical laws of classical physics, such as the relation between force and acceleration or the principle of energy conservation, had to be adapted to the new spatio-temporal framework, as we have already mentioned. Some of these adaptations, however, profoundly changed the understanding of concepts crucial for dynamics such as those of mass and energy. The impact of the special theory of relativity on the conceptual structure of classical physics was therefore not limited to a revision of its spatio-temporal scaffolding.

The special theory of relativity led, in particular, to the consequence that energy possesses inertial mass and that, more generally, energy and mass are but two aspects of a single conserved quantity whose relation is described by the famous equation $E = m \cdot c^2$, where E stands for the energy, m for the mass, and where c^2 is the square of the speed of light. In principle this "mass-energy relation" allows for the possibility that energy may be transformed into matter and matter into energy, in a well-defined ratio, although the equation itself does not specify the processes by which such transformations can be achieved. It follows that, whenever a body loses or gains energy, it also loses or gains inertial mass; furthermore, if energy—in whatever form—is subject to motion it behaves like a body with an inertial mass of the amount $m = E/c^2$. Einstein arrived at these conclusions in 1905 by theoretically studying transformation processes in which a body loses, for instance, energy in the form of radiation.[54] Considering such a process from two different frames of reference in relative motion with

53 For discussion, see the introduction to vols. 3 and 4 of this series "Theories of Gravitation in the Twilight of Classical Physics" (in vol.3).

54 See (Einstein 1905a).

respect to each other, he could use the Lorentz transformations to determine the effect of the energy loss on the motion of the body—and hence on its inertial mass.

The mass-energy relation is at the root of the striking implications of special relativity for dynamics. It requires, for instance, that stresses be systematically taken into account when considering the effect of forces on the motion of an extended body, even when in classical physics the stresses have no effect on this motion, e.g. when they are produced by a pair of equal and opposite forces acting along the same line. But according to the mass-energy relation, stresses embody energy and thus may change the inertial properties of the moving body. The elaboration of such consequences of special relativity had the effect of changing the fundamental role that the concept of inertial mass played in classical physics. Even when dealing with an extended body, in classical physics it was always possible to describe the effect of forces on its overall motion—leaving aside deformations—by a single quantity, its inertial mass. It was therefore possible, for many purposes, to treat an extended body according to the mental model of a single mass point having the same inertial mass as the extended body. This was another reason why the elementary dynamics of mass points had a more fundamental status in classical mechanics than the treatment of extended bodies, in addition to the fact that theories such as hydrodynamics and elasticity theory could be built on the conceptual basis of particle dynamics. In special relativity, on the other hand, there is in general no single quantity such as mass characterizing the inertial behavior of an extended physical system. The work of Max von Laue around 1911 made it clear that for this purpose no less than ten functions are required, which together form the components of a geometric object in Minkowski space called the "stress-energy tensor."[55]

2.9 Gravitation as a Stumbling Block of Special Relativity

How could gravitation be made to fit into the framework of the relativity theory of 1905? A modification of Newton's law of gravitational attraction was clearly necessary since it implies an instantaneous action at a distance, while the spatio-temporal framework of special relativity requires that no physical action can propagate with a speed faster than that of light. It quickly turned out that it was not at all difficult to adapt Newton's law to this spatio-temporal framework; as the work of Poincaré, Minkowski and others between 1905 and 1910 showed, there were even several possibilities for performing the necessary adjustments.[56] Furthermore, these adjustments could be brought without much difficulty into agreement with the well-known astronomical results confirming Newton's original theory. But while from a purely kinematical view point no essential difficulties arose, these special relativistic gravitational force laws were quite unsatisfactory from a broader perspective. In par-

55 See (Laue 1911a, 1911b).

56 For historical discussion, see the introduction to vols. 3 and 4 of this series "Theories of Gravitation in the Twilight of Classical Physics" (in vol.3).

ticular, since they took the form of an "action at a distance," even if with a temporal delay, they did not fulfill the ambition to create a field theory of gravitation, in analogy to the theory of the electromagnetic field.

Attempts to construct a relativistic field theory of gravitation could essentially follow the model represented by Lorentz's theory of electromagnetism. In contrast to an interaction between particles that can be described in terms of a mutual force, the "Lorentz model" takes into account that an interaction mediated by a field involves not only the interacting material particles but also their vicinity, in fact, their global environment. As discussed above, the Lorentz model describes in terms of a global field how this environment is affected by matter, considered the "source" of the field, and how this field in turn determines the motion of matter, now conceived as a "probe" exposed to the field. A mathematical representation of physical processes interpreted according to this model therefore necessarily comprises two parts: i) a field equation describing how a localized source, such as a particle, creates the global field or, alternatively, an analogous equation determining the "potential" from which the field can then be derived; in classical physics the corresponding second-order partial differential equation for the gravitational potential was known as the "Poisson equation," and ii) an equation of motion describing how the global field determines the motion of a localized probe; in Newtonian mechanics the corresponding equation is that between acceleration, force, and inertial mass $a = (F/m)$, representing the acceleration-implies-force model.[57]

In order to apply Lorentz's model of a field theory to the case of a relativistic theory of gravitation, one had to identify an appropriate mathematical representation of the gravitational field as well as of its source, and one had to find a generalization of the Poisson equation that was to be Lorentz-invariant. The familiar quantities from classical physics, the gravitational field derivable from a scalar gravitational potential and the gravitational mass, represented the default solutions to the first two problems. In addition, it was not difficult to write down a generalization of the Poisson equation compatible with special relativity. Nevertheless, the most obvious implementations of Lorentz's model of a field theory along these lines turned out to lead to serious difficulties, menacing the very framework of special relativity.

The early attempts at a special relativistic treatment of gravitation were, in particular, confronted with the dynamical consequences of special relativity following from a revised understanding of the concept of mass.[58] As we have seen, in classical mechanics the inertial mass of a body is quantitatively equal to its gravitational mass, although the two masses are conceptually distinct. This quantitative identity leads to the consequence that in a gravitational field all bodies fall with the same acceleration, whatever their constitution ("Galileo's principle"). The transformations that the concept of inertial mass underwent due to the advent of special relativity could not

57 For more extensive discussion, see "Pathways out of Classical Physics …" (in this volume).

58 See "Einstein, Nordström, and the Early Demise of Scalar, Lorentz Covariant Theories of Gravitation" (in vol. 3 of this series).

remain without effect for the understanding of this relation within the new relativistic mechanics. From the mass-energy relation it follows that the loss or gain of energy of a body changes its inertial mass—but does this change also affect its gravitational mass? If so, one would expect that Galileo's principle remains valid also in the relativistic context. It turned out, however, that it was not easy to preserve this principle in a special relativistic theory of the gravitational field since it seemed to follow from a straightforward implementation of the classical gravitational field equation within a special relativistic context that bodies fall with different accelerations in a gravitational field if their initial velocities and hence their inertial masses are different.

This problematic consequence may be glanced from a simple thought experiment, which might have been at the roots of Einstein's 1907 decision to abandon the attempt to find a special relativistic field theory of gravitation. Consider a stone falling vertically and a projectile shot horizontally at the same time the stone is dropped. According to classical physics, both hit the ground at the same time. If this situation is now considered from a moving reference frame, which moves with the same speed as the projectile along the horizontal, the roles of the stone and projectile are reversed. In the moving reference frame, the projectile falls vertically to the ground, while the stone now follows a projectile motion with a horizontal component in the opposite direction. According to classical physics, also within the moving reference frame, the two bodies should hit the ground at the same time. According to special relativity, however, the two events, which happen simultaneously in one frame of reference, cannot happen at the same time—because of the relativity of simultaneity, i.e., its dependence on the state of motion of the reference system—simultaneously in the other reference system moved with respect to the former. Galileo's principle, according to which all bodies fall with the same acceleration, can hence not be valid in both frames of references; it hence cannot be valid at all.[59]

This difficulty amounted not only to a clash between the new theory and a time-honored insight of classical mechanics, but pointed also to an internal conflict between the kinematical and the dynamical dimensions of special relativity that was generated by the attempt to incorporate gravitation into its framework. The dynamical consequences of special relativity, as embodied in the revision of the concept of mass, made it in fact impossible to simply transfer the elements of the classical gravitational field equation into a relativistic kinematic framework. In the classical Poisson equation, the

59 This reconstruction fits with Einstein's remark: "These investigations, however, led to a result which raised my strong suspicions. According to classical mechanics, the vertical acceleration of a body in the vertical gravitational field is independent of the horizontal component of its velocity. Hence in such a gravitational field the vertical acceleration of a mechanical system or of its center of gravity works out independently of its internal kinetic energy. But in the theory I advanced, the acceleration of a falling body was not independent of its horizontal velocity or the internal energy of the system. This did not fit with the old experimental fact that all bodies have the same acceleration in a gravitational field. This law, which may also be formulated as the law of the equality of inertial and gravitational mass, was now brought home to me in all its significance." In (Einstein 1954, 286–287).

mass density of matter acts as the source of the gravitational potential. But what should take its place in the relativistic counterpart of this equation? From the mass-energy relation it follows that any form of energy—even the energy of the gravitational field itself—should act as the source of a gravitational field, at least if the close relation between gravitational and inertial mass is to be upheld also in a relativistic theory. But in which form should the energy of an extended physical system act as the source of a gravitational field? It follows from the mass-energy relation, as we have seen above, that the inertial behavior of such a system in special relativity can no longer be easily characterized by a single function such as the classical inertial mass; it was therefore to be expected that similar difficulties arise also with regard to the gravitational effects of an extended physical system. The relativistic action-at-a-distance laws proposed by Poincaré and Minkowski, referring only to the interaction of point masses were of course, far from coping with such demands on a relativistic theory of gravitation and also for this reason had little impact on the further development of gravitational theories.[60] But also the most straightforward four-dimensional generalizations of the Poisson equation suggested only a single function, such as the mass as the source of the gravitational potential, and therefore encountered similar difficulties caused by the transition from the concept of mass to the concept of mass-energy.

A further, more "kinematical" difficulty characteristic of special relativistic field theories of gravitation is related to a simple mathematical property of Minkowski's four-dimensional formalism.[61] From the fact that, in Minkowski space, the scalar product of the "four-velocity" of a moving body with itself equals the square of the velocity of light, it follows that this four-velocity has always to be perpendicular to its "four-acceleration," since the derivative of the constant velocity of light vanishes. However, this condition imposes unacceptably restrictive conditions on the choice of a gravitational potential in the most obvious four-dimensional theory. In particular, the gravitational potential must be constant along the world-line of a particle, a very restrictive condition indeed. In summary, the framework of special relativity revealed itself as being too narrow for assimilating the knowledge on gravitation that was accumulated in the context of classical physics, at least as long as this framework was not somehow further elaborated or even stretched.

2.10 Exploring the Limits of Special Relativity

The various and partially conflicting demands on a relativistic theory of gravitation form the background of the different approaches to the problem of such a theory pursued by scientists like Albert Einstein, Max Abraham, Gunnar Nordström, and others

---

60 For more extensive discussion, see the introduction to vols. 3 and 4 of this series "Theories of Gravitation in the Twilight of Classical Physics" (in vol.3).

61 For further discussion of this problem, see "Einstein, Nordström, and the Early Demise of Scalar, Lorentz Covariant Theories of Gravitation" (in vol. 3 of this series).

in the period between 1912 and 1915.[62] Theories in which the gravitational potential is represented by a single function were advanced in this period by Abraham and Nordström, while a theory in which this potential was represented by a more complex mathematical object was pursued by Einstein. In fact, these theories embody the different possibilities offered by the shared mathematical and conceptual resources available at the time to respond to the challenges of a special relativistic field theory of gravitation.

Using earlier hints provided by Einstein's studies of special static gravitational fields, Abraham proposed in 1911 the first four-dimensional field theory of gravitation, obtained by a simple modification of the most straightforward special relativistic field theory.[63] In contrast to this "standard" theory, Abraham's theory assumes the speed of light to be variable. It thus succeeds in avoiding the difficulty related to Minkowski's formalism just mentioned, that is, the orthogonality between four-velocity and four-acceleration implied by the constancy of the speed of light, but, as it turned out, at the price of violating a fundamental principle of special relativity. It was, in fact, soon realized that Abraham's theory is no longer Lorentz-invariant but rather makes the geometry of Minkowski space depend on the variable speed of light which in turn is a function of the gravitational potential. This radical step met with criticism because it led to an internal contradiction. But even apart from this difficulty, Abraham's theory did not offer a response to the "dynamical" problems related to the modification of the concept of mass discussed above.

These problems quickly became the central issue in the development of Nordström's apparently more conservative alternative approach, which was intended to remain within the framework of special relativity.[64] But whether more conservative or not, this approach had the same starting point as Abraham's, the problem of bringing the most obvious four-dimensional generalization of a gravitational field equation into harmony with the most elementary requirements of a theory of gravitation. His solution for avoiding the restrictive conditions on the gravitational potential resulting from the "standard" special relativistic gravitational field theory was to let the mass of a body depend on the gravitational potential, instead of assuming a dependence of the speed of light on the gravitational potential. In this way, it seemed that he had only to introduce a minor unorthodoxy into special relativity, rather than giving up its entire conceptual framework, as was effectively the case with Abraham.

The theory that Nordström first proposed in 1912 did not, however, succeed in avoiding other objections against a special relativistic approach. Like Abraham's theory, Nordström's initial approach was based on considering only the mass as the source of the gravitational field, neglecting other forms of energy. In order to avoid

62 This is discussed more extensively in the introduction to vols. 3 and 4 of this series "Theories of Gravitation in the Twilight of Classical Physics". See also "Einstein, Nordström, and the Early Demise of Scalar, Lorentz Covariant Theories of Gravitation"; "The Summit Almost Scaled …" (all in vol. 3 of this series).

63 See (Abraham 1912).

64 See (Nordström 1912).

this difficulty, a further elaboration of Nordström's theory therefore had to take into account the recent advances in understanding the role of energy for the inertial behavior of bodies. We have seen above that the mass-energy relation implies that in special relativity the inertial behavior of extended physical systems can, in general, no longer be described by the single function "inertial mass" but that this description rather requires a complex, 10-component object, the "stress-energy tensor." Given the demand that the inertial and the gravitational behavior of such an extended system should be governed by the same physical quantity embodying its energetic properties, this stress-energy tensor was also a natural candidate for the source of the gravitational field, thus effectively replacing the classical notion of gravitational mass. Nordström's theory, however, was guided by the default settings for the Lorentz model inherited from Newtonian theory in assuming that the gravitational potential is represented by just one single function which had to correspond to another single function representing the source of this potential. In the classical case, this single function characterizing the source of the gravitational field was given by the gravitational mass. Now, in the light of the new insights into special relativistic dynamics and of the constraints imposed by Nordström's theory, it became necessary to extract such a single function from the complex stress-energy tensor for a physical system acting as the source of a gravitational field.

The attempts, in particular by Nordström and Einstein, to address this problem necessitated further revisions of Nordström's original theory. Eventually Nordström was forced to assume that the gravitational potential influences not only the mass of a body, but also length and time measurements performed under the influence of a gravitational field. However, this further revision implied that Nordström's theory of gravitation had essentially undercut its own conceptual roots in the spatio-temporal framework of special relativity. Indeed, as it turned out, this framework was no longer directly accessible through measurements as these are affected by the presence of a gravitational field.

These and other attempts between 1911 and 1915 to confront the challenge that the inclusion of gravity posed for special relativity led to a situation similar to that of classical aether-based electrodynamics at the turn of the century, when it was confronted with the challenge of incorporating the electrodynamics of moving bodies. In both cases it was possible to cope with these challenges by continually adjusting and modifying the original framework which, however, created tensions between the elaborated theories and their conceptual roots. But further development under the spell of these challenges not only produced internal inconsistencies and a proliferation of possible modifications, it also generated the preconditions for new perspectives as a consequence of the enhanced opportunities to reflect upon the accumulated knowledge. In the case of the electrodynamics of moving bodies, these preconditions were constituted, among other aspects, by the new space and time variables in Lorentz's theory.

In the case of the relativistic field theory of gravitation we have also encountered the emergence of such elements of a new perspective, even though we have reviewed only a small segment of its development. Indeed, a sober look at Abraham's and Nordström's efforts could not only have suggested that a satisfactory field theory of gravitation might transcend the limits of special relativity, but could have also revealed hints as to how such a theory might be constructed: by letting the geometry of Minkowski space depend on the gravitational potential (Abraham); by representing the gravitational potential not by a single function but by a 10-component object on a pair with the stress-energy tensor and having this tensor as its source (Laue and Nordström); and by including an effect of the gravitational potential on the measurement of space and time (Nordström).

In summary, our survey of the evolution of classical physics has shown that the accumulation of knowledge in its highly organized disciplinary structures produced borderline problems which also challenged the classical understanding of gravitation. The fact that the problem of a relativistic theory of gravitation was deeply rooted in a structural crisis of classical physics makes it evident that the genesis of general relativity was not just a one-man affair and a lucky individual discovery but the result of a profound transformation of the extended system of knowledge of classical physics that may well have taken place, albeit under a different form and possibly with a different outcome, even if Einstein had never lived. This survey makes it thus clear in which sense general relativity was not created *ex nihilo*. What remains open, however, is the second question posed in the introduction concerning the success of the heuristics that guided Einstein in its creation. While this survey makes it plausible that this success must have been due to the fact that Einstein's heuristics exploited the knowledge of classical and special relativistic physics, it remains to be clarified how it did so concretely. This is the question at the center of the next section.

3. THE ROOTS OF EINSTEIN'S HEURISTICS IN THE KNOWLEDGE OF CLASSICAL AND SPECIAL RELATIVISTIC PHYSICS

3.1 The Drama of General Relativity

How is Einstein's work on general relativity related to the knowledge resources reviewed in the last section? At first glance it seems that the processes described above merely set the stage for the truly decisive events associated with his struggle for a relativistic theory of gravitation. In fact, this struggle resembles a heroic drama which, according to Einstein's own retrospective account, has three major acts:[65] The drama began in 1907 when he conceived the idea of the so-called "equivalence principle," which was at the origin of his unusual and lonely path towards a theory of gravitation, and with which he was to realize a generalization of the principle of relativity. The drama reached a first culmination in 1912, when Einstein introduced a 10-component metric tensor as a representation of the gravitational potential, substituting its representation in classical physics by a single scalar function. At that point, the

65 See "The First Two Acts" (in this volume).

full extent of the deviation of the new theory of gravitation from classical physics became visible. Finally, in 1915, the drama reached its climax when Einstein, after a desperate search and numerous failed attempts, formulated the gravitational field equation of general relativity as a differential equation for the metric tensor.

However, describing the three steps associated with the years 1907, 1912, and 1915 as turning points of a drama or of an individual biography obviously does not explain why each of them was so successful in advancing the solution of the problem to create a relativistic theory of gravitation, in particular as they do not even seem to lie along a straight path towards such a solution. Einstein's contemporaries strongly objected, for instance, against the introduction of the metric tensor which they considered to be an unnecessary complication with respect to the simple scalar gravitational potential of the Newtonian theory.[66] Even a psychological explanation of the origin of the three breakthroughs in Einstein's thinking can hardly account for their effectiveness in creating general relativity as a durable solution to the problem of a relativistic theory of gravitation, emerging at the dusk of classical physics.

In the following, we shall therefore concentrate on the way in which Einstein's three crucial steps functioned in the transformation of the knowledge of classical science. For this purpose, we will describe what happened in terms of mental models effecting an integration of physical and mathematical knowledge. The emphasis is not on the question of how exactly these models entered Einstein's individual thinking. What matters here is that they originated either in the shared scientific knowledge of the time, such as the model of a field equation taken over from Lorentz's electrodynamics, or resulted from a reflection on conflicts within this shared knowledge of the kind described in the previous section, such as the conflict between Galileo's principle and special relativity.

3.2 Einstein's Realization of the Conflict between Galileo's Principle and Special Relativity

In September 1907 Einstein agreed to the offer by Johannes Stark to write a review article on the theory of relativity which he was expected to complete within two months.[67] At the beginning of November, when he had completed the first half of the paper, he was yet uncertain whether gravitation should be included—at least he did not mention this topic in a letter to the editor Stark which he wrote at that time.[68] When the paper was submitted on the fourth of December, it did comprise a short section on gravitation.[69] The obvious problem confronting Einstein was the reconciliation of Newton's law of gravitation with the requirement of a finite speed of the

66 See vols. 3 and 4 of this series.
67 See Einstein to Johannes Stark, 25 September 1907 (CPAE 5, Doc. 58, 74–75).
68 See Einstein to Johannes Stark, 1 November 1907 (CPAE 5, Doc. 63, 77–78); see also the discussion in (Fölsing 1993, 266).
69 See (Einstein 1907, sec. V, 454–462).

propagation of physical effects following from special relativity. In a later recollection, Einstein wrote:

> Like most physicists, at this period, I endeavoured to find a "field-law," since, of course, the introduction of action at a distance was no longer feasible in any plausible form once the idea of simultaneity had been abolished.
>
> The simplest way was, of course, to keep the Laplace scalar potential of gravity and to extend the Poisson equation by adding, in such a way as to comply with the special theory of relativity, a term differentiated with respect to the time. (Einstein 1933, 6)

Whatever precise form Einstein had given to his attempt at a special relativistic gravitation theory, he soon encountered the above-mentioned intrinsic conflict between the kinematical and dynamical implications of special relativity for the problem of gravitation, which shifted the central conflict from that between Newton's law and special relativity to that between special relativity and Galileo's principle. The conflict suddenly transformed this long-familiar principle—and the equality of inertial and gravitational mass on which it is based in classical physics—into a touchstone for a new theory of gravitation. It was as if the elaboration of a relativistic theory of gravitation had directed a spotlight onto this previously rather inconspicuous asset of classical physics. This is strikingly confirmed by Einstein's own recollection:

> This principle, which can also be stated as the law of the equivalence of inertial and gravitational mass, impressed me as being of fundamental importance. I wondered how this law could exist, and believed that it held the key to the real understanding of inertia and gravitation. I never seriously doubted its exact validity, even though I did not know about the beautiful experiments of Eötvös, which, if I remember alright, were not known to me until a later date.
>
> I gave up, therefore, the attempt, which I have sketched above, to treat the problem of gravitation within the framework of the special theory of relativity; it was clearly inadequate, since it failed to take into account just the most fundamental property of gravitation. (Einstein 1933, 7)

The ground was thus prepared for reflecting upon this property of gravitation and rethinking the relevant knowledge of classical physics taking into account that the framework of special relativity was evidently too narrow for capturing the problem of gravitation.

3.3 The Synthesis of Knowledge by the Elevator Model

Even within classical mechanics Galileo's principle appears to be an odd coincidence, due to the equality of gravitational and inertial mass rather than a fundamental property anchored in its conceptual structure. However, when Einstein reconsidered it in the context of his attempt at a relativistic theory of gravitation, it somehow seemed to hold the key to a deeper understanding of gravitation and inertia. But what exactly was the question to which Galileo's principle suggested an answer? Einstein eventually noticed that it did fit with Mach's peculiar view on mechanics, questioning the privileged role of inertial frames of reference and the notion of acceleration with respect to absolute space. Mach had claimed that the curvature of the surface of water in Newton's bucket experiment might also occur if the bucket were at rest while the cosmic masses rotated around it. Mach thus established an equivalence between a situation in which a physical effect—the curvature of the water surface—is caused by acceleration, and a situation in which the same effect can be interpreted as being caused by forces exerted on the water by the cosmic masses. If Galileo's principle is now reconsidered from the point of view of such an equivalence relation, it becomes plausible to reduce Mach's argument to a comparison of two much simpler situations; one in which just the usual linear acceleration of free fall is considered instead of a rotational motion, and one in which just the usual gravitational force of the Earth is contemplated instead of a speculative force of moving cosmic masses.

The problem to which Galileo's principle provided a response was, in other words, the generalization of the principle of relativity to accelerated motions which, for Einstein, was associated with Mach's view. He later recalled:

> After the special theory of relativity had shown the equivalence for formulating the laws of nature of all so-called inertial systems (1905) the question whether a more general equivalence of coordinate systems existed was an obvious one. In other words, if one can only attach a relative meaning to the concept of velocity, should one nevertheless maintain the concept of acceleration as an absolute one? From the purely kinematic point of view the relativity of any and every sort of motion was indubitable; from the physical point of view, however, the inertial system seemed to have a special importance which made the use of other moving systems of coordinates appear artificial.
>
> I was, of course, familiar with Mach's idea that inertia might not represent a resistance to acceleration as such, so much as a resistance to acceleration relative to the mass of all the other bodies in the world. This idea fascinated me; but it did not provide a basis for a new theory. I made the first step towards the solution of this problem when I endeavoured to include the law of gravity in the framework of the special theory of relativity. (Einstein 1933, 5–6)

It was evidently the highlighting of Galileo's principle by the problems of a special relativistic theory of gravitation that had linked this principle, in Einstein's mind, with Mach's interpretation of mechanics, and in particular with the idea to generalize the principle of relativity also to accelerated frames of references. According to standard classical mechanics, inertial and accelerated frames of reference have a fundamentally different status. But when, as Galileo's principle implies, all bodies fall in the same way in a uniform and homogeneous gravitational field, an observer falling with them could imagine living, at least temporarily, in an inertial frame of reference although he is himself falling with growing speed. In fact, he would neither feel his own weight nor observe any forces acting on the bodies falling with him. It is as if the gravitational field had been neutralized by his own accelerated motion. His accelerated frame of reference would therefore be indistinguishable from an inertial frame. This indistinguishability between an accelerated and an inertial frame could, in turn, be made into a criterion for the validity of Galileo's principle. If, in particular, the new theory of gravitation were to incorporate this principle, it would have to be also a generalized theory of relativity allowing for accelerated frames of reference. This way of reinterpreting Galileo's principle as the equivalence of a freely falling reference frame with an inertial one, and as the starting point for resolving the problems raised by Mach, must have struck Einstein as a flash-like insight, as is made evident by his later recollections. In his famous Kyoto lecture, for instance, Einstein reminisced:

> I was sitting in my chair in the patent office at Bern when all of a sudden a thought occurred to me: 'If a person falls freely he will not feel his own weight.' I was startled. This simple thought made a deep impression on me. It impelled me towards a theory of gravitation. (Pais 1982, 179)

Other recollections not only confirm the sudden nature of this insight but also hint at the intellectual background that had prepared it and, in particular, at the role of Mach's critique as suggesting the problem to which Galileo's principle, freshly discerned in its significance, provided the striking and unexpected resolution:

> Now it came to me: the fact of the equality of inertial and gravitational mass, i.e., the fact of the independence of the gravitational acceleration from the nature of the falling substance may be expressed as follows: In a gravitational field (of small spatial extension) things behave as they do in a space free of gravitation, if one introduces into it, in place of an "inertial system," a frame of reference accelerated relative to the former. ... The concept of "acceleration relative to space" then loses all meaning and with it the principle of inertia along with the paradox of Mach. (Einstein 1979, 61, 63)

Einstein's rethinking of Galileo's principle was guided not only by his familiarity with Mach's foundational critique of mechanics but also by other aspects of his earlier intellectual experience, such as his recognition, in the context of his work on the electrodynamics of moving bodies, that electric and magnetic fields only have a relative existence depending on the state of motion. In one of his recollections he stressed the analogy of this crucial insight from the context of special relativity with the relative existence of the gravitational field as revealed by accelerated motion:

> At this point, there occurred to me the happiest thought of my life. Just as is the case with the electric field produced by electromagnetic induction, the gravitational field has similarly only a relative existence. *For if one considers an observer in free fall, e.g., from the roof of a house, there exists for him during this fall no gravitational field—at least not in his immediate vicinity.* Indeed, if the observer drops some bodies, then these remain relative to him in a state of rest or in uniform motion, independent of their particular chemical or physical nature (in this consideration the air resistance is, of course, neglected). The observer therefore has the right to interpret his state as "at rest."[70]

---

70 "Da kam mir der glücklichste Gedanke meines Lebens in folgender Form: Das Gravitationsfeld hat an einem betrachteten in ähnlicher Weise nur eine relative Existenz wie das durch magnetelektrische Induktion erzeugte elektrische Feld. *Denn für einen vom Dache eines Hauses frei herabfallenden Beobachter existiert während seines Falles*—wenigstens in seiner unmittelbaren Umgebung—*kein Gravitationsfeld*. Lässt der Beobachter nämlich irgend welche Körper los, so bleiben sie relativ zu ihm im Zustand der Ruhe bezw. gleichförmigen Bewegung, unabhängig von ihrer besonderen chemischen und physikalischen Natur. Der Beobachter ist also berechtigt, seinen Zustand als "Ruhe" zu deuten." (CPAE 7, Doc. 31, 265). English translation in (Holton 1971). See also (Miller 1992, 325; Pais 1982, 178).

The way in which Einstein exploited the knowledge resources available to him obviously depended on his specific perspective on these resources. But in whatever prior experience this perspective was actually grounded, there can be no doubt that his first crucial step towards a relativistic theory of gravitation—the introduction of what he later called his equivalence principle—has its origin in a thorny process of knowledge integration that started with the attempt to assimilate the classical knowledge on gravitation to the framework of special relativity. How exactly did the equivalence principle structure this process?

The equivalence principle states that all physical events in a uniform and homogeneous gravitational field are equivalent to those happening in a uniformly accelerated frame of reference. It is often described, also by Einstein himself, in terms of a "thought experiment" with a closed laboratory, which is also known as the "elevator experiment." However, if it is merely conceived as a thought experiment, then it must remain a riddle why Einstein held on to it for so long and why he made it the central element of his heuristics, even against the resistance of almost all his colleagues. On the other hand, if one interprets it as an operation with the mental model of a laboratory, then it becomes clear that it could have dramatic effects on the architecture of the knowledge of classical physics.

Einstein's laboratory may be considered in two states that, in classical physics, are described by different concepts. In the first case the laboratory is uniformly accelerated by an arbitrary external force with respect to an inertial system. It represents the mental model of a system with inertial forces. Among its slots are the state of motion of the system, the inertial forces in its interior, and the motion of bodies caused by them. In the second case the laboratory is at rest in an inertial system and is exposed to a homogeneous gravitational field. It now represents the mental model of a system with gravitational force. Among its slots are the state of motion of the system, the gravitational force acting on the bodies within the system, and the motions caused by it.

Einstein noticed that the observer in the interior of the laboratory cannot distinguish between these two cases. A laboratory of which it is only empirically known that bodies in its interior fall to the bottom with uniform acceleration satisfies the conditions for the slots of both mental models, which are connected with each other in this way. In fact, the slots for the motion of the bodies within the laboratory are, in both cases, filled with the same empirically given data, while the slots for which no empirical information is available, such as those for the states of motion, can be filled by default assumptions delivered by the interpretation of the laboratory either as being in accelerated motion or as being exposed to a gravitational field. The integration of the two mental models questions or even dissolves essential distinctions of classical and special relativistic physics, in particular that between accelerated systems and inertial systems, and that between inertial forces and gravitational forces.

Even years before the formulation of only the first approach to a generally relativistic theory of gravitation, the integrated model—in the following called the "elevator model"—became the central asset of Einstein's heuristics. In fact, this integrated model allowed him to relate two separate knowledge areas of physics to each other. In particular, it enabled the investigation of special cases of the gravitational field by means of the study of accelerated motion. Accelerated motion, in turn, could be analyzed by methods of special relativity, assuming that the physical effects at a certain point in the accelerated laboratory are the same as those in a laboratory moving uniformly with the same instantaneous speed, that is, in a "comoving inertial frame of reference." On the basis of this successful, if only partial integration of the knowledge of classical mechanics and special relativity, Einstein was able to draw conclusions concerning the empirical consequences of a theory of gravitation still to be formulated, among them the prediction of the deflection of light and of the redshift in a gravitational field,[71] as well as the explanation of the anomalous perihelion motion of Mercury.[72]

The bending of light in a gravitational field, for instance, could simply be inferred from the observation that, in an accelerated laboratory, light rays must be curved as a consequence of the superposition of the motion of the laboratory and of the light. The conclusion that this is also the case for a gravitational field was in accordance with the assumption that energy has not only inertial but also gravitational mass, so that the energy of light should be subject to gravitational attraction. Such straightforward considerations are at the roots of the most striking observational predictions of general relativity, made long before its completion and confirmed only years afterwards. Among them is Einstein's prediction that, during a solar eclipse, one should be able to observe a variation of the apparent positions of stars whose light passes close to the sun and is therefore deflected by its gravitational attraction, a prediction first published in 1911 following up on Einstein's 1907 considerations,[73] and confirmed during a solar eclipse in 1919.[74] Among these striking predictions is also that of the existence of gravitational lenses: distant massive objects which create images and intensify the light of even more distant objects aligned with them. Einstein first determined the properties of such gravitational lenses in 1912, three years before completing general relativity.[75] He later abandoned the idea because he doubted that such objects could ever be observed; an observation that, in fact, was only achieved much later. These examples illustrate how the knowledge resources of classical physics made it possible to identify effects predicted by general relativity even before its formulation. In fact, the deflection of light, revealed as a qualitative effect by the elevator model as early as 1907, could, without much difficulty, be combined with mental models from ray optics involving simple constellations such as that of a solar eclipse or a gravitational lens.

71 See (Einstein 1907, sec. V).
72 See Albert Einstein to Conrad Habicht, 24 December 1907, (CPAE 5E, 82).
73 See (Einstein 1911).
74 Apart from a factor of 2; for a historical discussion, see (Earman and Glymour 1980).
75 For a discussion of the early history of gravitational lensing, see (Renn, Sauer and Stachel 1997, Renn and Sauer 2003).

3.4 Conceptual Challenges Implied by the Elevator and Bucket Models

Einstein's elevator model was not only a heuristic device capable of predicting remarkable observational implications of the new theory of gravitation, it also highlighted its incompatibility with special relativity and allowed Einstein to anticipate some of the conceptual changes associated with such a theory. The bending of light in a gravitational field, for instance, suggested that the speed of light is no longer constant, in contradiction with one of the fundamental principles of special relativity. This qualitative conclusion could be underpinned, as Einstein did in his review paper of 1907,[76] by an analysis of time synchronization in an accelerated frame of reference that made it evident that the concept of time in a gravitational field had to be further differentiated with respect to that of special relativity. In fact, time, in the sense of a synchronized network of clocks, could be defined with respect to any of the comoving inertial frames anchored at particular points of an accelerated frame of reference. But since these different "local inertial frames" are in relative motion to each other, two such networks do not coincide since, according to special relativity, simultaneity means something different in each of them. It is hence no longer possible to simply refer to "the time" of an accelerated frame of reference (or of a frame with a gravitational field) since one can no longer expect that two standard clocks, initially synchronized at the same location in a gravitational field, would still remain synchronized when one is transported to a region with a different gravitational potential; in this sense the "homogeneity" of time is lost in a theory of gravitation incorporating Einstein's principle of equivalence.

An extension of the considerations that had led to the elevator model allowed Einstein to arrive at similar insights concerning the necessity of revising the concept of space in the new theory of gravitation. Such an extension of the original model, to include more general cases of accelerated motion and, in particular, the case of rotation, was suggested both by Mach's analysis of inertial effects in classical mechanics and by the Lorentz model. The combination of knowledge resources from mechanics and field theory was, in fact, characteristic of Einstein's work on a new theory of gravitation.[77] The Lorentz model, rooted in the tradition of field theory, made it possible to conceive of gravitation and inertia as two aspects of one gravito-inertial field, in analogy to the unification by Maxwell's field equations of the electric and magnetic field to one electromagnetic field. In a letter to Paul Ehrenfest from spring 1912 Einstein wrote with reference to static gravitational fields:

> My case corresponds to the electrostatic field in the theory of electricity, whereas the more general static case would further include the analogue of the static magnetic field. I am not yet that far.[78]

---

76 See (Einstein 1907, sec. V).

77 For further discussion, see "The Third Way to General Relativity" (in vol. 3 of this series).

78 "Mein Fall entspricht in der Elektrizitätstheorie dem elektrostatischen Felde, wogegen der allgemeine[r]e statische Fall noch das Analogon des statischen Magnetfeldes mit einschliessen würde. So weit bin ich noch nicht." Einstein to Paul Ehrenfest, before 20 June 1912, in (CPAE 5, Doc. 409, 486).

Mach's analysis of classical mechanics, on the other hand, suggested identifying the inertial forces in accelerated frames of reference as being due to an interaction between bodies in motion. His interpretation of the mental model of a rotating bucket, in particular, made it plausible to search for a theory that admits rotating frames of reference in the sense of a generalization of the principle of relativity and to conceive the inertial effects occurring in them as a specific case of this hitherto unknown interaction. In this vein, Einstein wrote, also in spring 1912, to his friend Michele Besso:

> You see that I am still far from being able to conceive rotation as rest![79]

While attempts to formulate a purely Machian mechanics on the basis of introducing such inertial interactions had been hopeless, the combination of this idea with the perspective of field theory could now provide the latter with precise information, otherwise missing, concerning the properties of *dynamical* gravitational fields. In other words, an integration of the Machian bucket model with the Lorentz model made it possible to identify the inertial forces in a rotating frame as the analogon to a magnetostatic field, thus filling the slots of the Lorentz model with specific knowledge about the special case of a stationary gravitational field.

When Einstein began to search for a gravitational field equation in 1912, the bucket model became, next to the elevator model, the second crucial mental model of his heuristics, exploiting the possibilities of analyzing properties of a generalized gravitational field by considering the inertial effects of accelerated motion known from classical mechanics. The bucket model pointed to the effects of Coriolis-like forces on light rays in a stationary gravitational field, which turned out to be such that light rays cannot go back along the same path from one point to the other.[80] From this insight it follows that the force exerted by such a stationary gravitational field is velocity-dependent, in this case changing with the direction of the velocity of light. However, even within the framework of special relativity, the inclusion of rotating reference frames was connected to a conceptual difficulty that had its origin in the length contraction of rapidly moving bodies and the problem of defining a rigid body in special relativity. Einstein and Max Born had already encountered this difficulty in 1909.[81] It had also been found independently by Ehrenfest.[82] Ehrenfest argued that the circumference of a cylinder, which is slowly set into motion around its axis should, according to special relativity, show a contraction with respect to the state of rest, while its radius is not affected by such a contraction since it lies orthogonally to the motion. For the rotating cylinder, the ratio between circumference and radius

---

79 "Du siehst, dass ich noch weit davon entfernt bin, die Drehung als Ruhe auffassen zu können!" Einstein to Michele Besso, 26 March 1912, in (CPAE 5, Doc. 377, 435).

80 See Paul Ehrenfest to Albert Einstein, St. Petersburg, before 3 April 1912, in (CPAE 5E, 439–445). For a more detailed historical analysis, see "The First Two Acts" (in this volume).

81 See (Born 1910).

82 See (Ehrenfest 1909).

should therefore, from the viewpoint of an observer at rest, deviate from the number π that, in Euclidean geometry, determines this ratio.

This difficulty was related to what became known as the "Ehrenfest paradox" which gave rise to a controversial discussion about the definition and role of rigid bodies in relativity theory. Einstein, however, took this paradox as a hint relevant for a generalization of relativity theory on the basis of his equivalence principle. Immediately after the discovery of the problem he wrote to Arnold Sommerfeld:

> The treatment of the uniformly rotating rigid body seems to me to be of great importance because of an extension of the relativity principle to uniformly rotating systems that is based on a line of reasoning analogous to that which I tried to purse for uniformly accelerated translation in the last § of my paper that was published in the *Zeitschr. f. Radioaktivit.*[83]

When Einstein interpreted the Ehrenfest paradox in terms of the bucket model, it took on a new significance, implying conceptual consequences for the new theory of gravitation which sharpened its conflict with special relativity. In one of his papers published in 1912 (Einstein 1912a), Einstein argued that the ratio between the circumference and radius of a circle or a disk in a rotating laboratory is no longer given by π because, as he later explained in detail,[84] the Lorentz contraction affects a ruler posed along the circumference and moving in the momentary direction of the rotating motion, while it does not affect a ruler posed along the radius of the disk. As a consequence, one needs more such contracted rulers to measure the circumference of the rotating disk which thus exceeds the length of 2π times the radius expected from ordinary geometry. In other words, the spatial properties of a stationary gravitational field no longer satisfy Euclidean geometry.

For Einstein the equivalence principle, and with it the use of mental models such as the elevator and bucket models, was eventually subsumed under a more general heuristic principle, his generalized principle of relativity. According to this principle, the new theory of gravitation should not make a distinction between inertial and gravitational effects, which are rather to be described as effects of an integrated gravito-inertial field. It should, in particular, admit reference frames in arbitrary states of motion and describe the inertial effects occurring in them as the effects of such a generalized gravito-inertial field. As will become clear in the following, this principle played a crucial role in shaping the mathematics employed by Einstein to formulate the gravitational field equation, imposing, in particular, its covariance with respect to general classes of coordinate transformations. However, it follows from the preceding discussion that Einstein's generalized principle of relativity not only had implications

83 "Die Behandlung des gleichförmig rotierenden starren Körpers scheint mir von grosser Wichtigkeit wegen einer Ausdehnung des Relativitätsprinzips auf gleichförmig rotierende Systeme nach analogen Gedankengängen, wie ich sie im letzten § meiner in der Zeitschr. f. Radioaktivit. publizierten Abhandlung für gleichförmig beschleunigte Translation durchzuführen versucht habe." Einstein to Arnold Sommerfeld, 29 September 1909, in (CPAE 5, 210–211).

84 For an extensive historical discussion, see (Stachel 1989). For a modern assessment, see (Vishveshwara 2003).

for the mathematical instruments used in the creation of general relativity, but that it also incorporated knowledge resources of classical physics, in particular, the knowledge on inertial effects in accelerated frames of reference which were traditionally explained as being due to the existence of an absolute space.

3.5 The Synthesis of Knowledge by the Curved-Spacetime Model

Classical physics offered a number of tools for the further exploration of the properties of a generalized gravitational field. Exploiting these tools, Einstein was eventually led to a major breakthrough in his understanding of gravitation. As a result of this breakthrough, gravitation was now conceived as being due to the curvature of space and time. How did this breakthrough come about? In order to describe how gravitation affects motion Einstein first made use of the classical acceleration-implies-force model, conceiving gravitation as a force. In a second step, he then realized that the way in which a gravitational field influences the motion of a body exposed to it could also be accounted for by the constrained-motion model, also familiar from classical physics, wherein the equation of motion is formulated in terms of a variational principle. Finally, in a third step, he realized that an equation of motion thus formulated also perfectly matches the slots of another mental model, this time rooted in the shared knowledge of contemporary mathematics, the "curved-surface model." As a result, these two models were joined to become an integrated model, the "curved-spacetime model," which opened the way for a complete and coherent mathematical description of motion in a relativistic gravitational field by virtue of the mathematical structures associated with the curved-surface model. It follows from this revised description of the effect of gravitation that the gravitational field itself could now be conceived, instead of being a force in the sense of Newtonian physics, as expressing the geometric properties of a generalized spacetime continuum. In fact, this new world generalizes Minkowski's spacetime in the same sense as a curved surface represents a generalization of a plane surface, its geometric properties being governed not by a constant but by a variable metric. This metric, in turn, emerged as the appropriate mathematical representation of the gravitational potential. Because of its origin in the integration of two mental models, the metric tensor embodies both the knowledge of the generalized gravito-inertial field revealed by the equivalence principle, and the knowledge of the causal structure of spacetime following from special relativity and incorporated in Minkowski's formalism.

A closer look at the emergence of this breakthrough reveals it as a rather straightforward implication of the knowledge so far accumulated. Until 1911 Einstein had committed himself mainly to exploring, by means of the equivalence principle, the effects and conceptual changes characterizing a new theory of gravitation, evidently without seriously attempting its construction. Only in early 1912 was he challenged by the publications of Max Abraham to elaborate such a theory, at least for the special case of a static gravitational field.[85] He did so by applying the Lorentz model to the results of his use of the equivalence principle. In particular, he attempted to describe the inertial effects in a uniformly accelerated frame in terms of a field equation for a static gravitational field, and the motion of particles by an equation of motion in accordance with the acceleration-implies-force model. For this purpose he explored how a force-free motion in an inertial frame would look if considered from the perspective of an accelerated frame of reference from which it could then be interpreted, on the basis of the equivalence principle, as a motion subject to a gravitational force.

Having derived transformation equations between the inertial and the accelerated frame of reference compatible with the available knowledge, in particular about the propagation of light, Einstein arrived at transformation equations for space, time, and a variable speed of light. Since in Einstein's theory of static gravitation the speed of light is variable, his results were obviously no longer compatible with special relativity. As discussed in the previous section, Max Abraham had followed an entirely different strategy. He believed to have resolved the problem of gravitation within the framework of Minkowski's four-dimensional representation in a simple and surprisingly elegant way, but Einstein soon recognized that Abraham's use of the Minkowski formalism was not compatible with the assumption of a variable speed of light. In reaction to Einstein's criticism, Abraham became the first, in February 1912, to propose that the four-dimensional line element, defining the infinitesimal distance between points in Minkowski space in terms of a constant metric, has to be replaced by a variable line element whose functional dependence on the coordinates is determined by a gravitational potential associated with the variable speed of light c.

While the appearance of such a generalized line element in Abraham's work is striking, pointing as it does at what was to become a central mathematical object of general relativity, neither Abraham's nor Einstein's contemporary research offered a context for realizing this key role. On the contrary, one of the essential properties of the line element in Minkowski space—its invariance under Lorentz transformations—was lost by the introduction of a variable speed of light, as Einstein did not fail to notice. As a consequence, the formalism introduced by Abraham clearly did not represent a coherent generalization of Minkowski's spacetime. On the other hand, Einstein's theory of the static gravitational field, for the time being, did not offer even a hint at a general framework that would allow him to go beyond the special case of static fields. In addition, it even suffered from a problem of compatibility with the essential physical requirement of energy conservation. In fact, this problem had not only necessitated an adjustment of its field equation but also a restriction of Einstein's central heuristic device, the equivalence principle, to infinitesimally small fields. In short, the attempts at a relativistic theory of gravitation pursued by Einstein and Abraham had met with a dilemma in which neither mathematical formalism nor physical heuristics seemed to provide an indication as how to proceed further.

By the end of May 1912, Einstein encountered at least a partial resolution of this dilemma in the course of revisiting the equation of motion of his theory. This turning

85 For more extensive discussion and references to Abraham's papers, see "The Summit Almost Scaled …" (in vol. 3 of this series).

point—referred to above as the second step in Einstein's realization of gravitation as being due to the curvature of spacetime—is documented in a postscript to his second 1912 paper on the static theory, in which he proposed the modification of its field equation just mentioned.[86] Einstein had also reconsidered the equation of motion of his theory, originally shaped, as mentioned, according to the acceleration-implies-force model. He now successfully attempted to rephrase it, without changing its physical content, according to the constrained-motion model.

The elaborate mathematical formalisms associated with this model in analytical mechanics, such as the variational calculus of Euler, Lagrange, and Hamilton, had proven to be standard tools of classical physics for succinctly representing the structure of complex theories, in particular when these theories could not be interpreted in terms of simple mechanical interactions, as was the case, for example, for optics or electrodynamics. The essential ingredient of such reformulations in terms of a variational calculus is usually the integral of a physical quantity, depending on the trajectory of a particle or of a light ray, whose extremal value is to be determined by a variation under certain constraints. In this way, it is possible to single out among several possible trajectories along a curved surface the one that actually represents the "natural motion" under the given constraints. A direct interpretation of abstract variational formalisms in terms the physical meaning of the constrained-motion model is, under certain conditions, still possible but not essential for the successful development and implementation of these formalisms.

In his reformulation of the equation of motion in a static gravitational field Einstein again followed as closely as possible the precedent of special relativity, just as he had done in his original formulation guided by the acceleration-implies-force model. In this case the precedent was given by Planck's formulation of the equation of motion of a particle in the framework of a variational calculus adapted to Minkowski space.[87] The physical quantity, whose variation Planck had considered, is the integral over the square root of the negative line element along the trajectory of the particle. The requirement that this integral be an extremal has a direct geometric interpretation: the world line of a force-free particle in Minkowski space is given by a geodesic, that is, by the straightest possible line connecting the initial and the end point of the particle's motion. It now turned out that the corresponding expression for the motion in a static gravitational field takes on exactly the same form as that given by Planck for a gravitation-free Minkowski space, the only difference being that the speed of light is now assumed to be variable:

$$\delta\left(\int\sqrt{c^2dt^2 - dx^2 - dy^2 - dz^2}\right) = 0,$$

where δ stands for the variation of the subsequent integral.

86 See (Einstein 1912b, 458).
87 See (Planck 1906a, 1907).

Einstein was not only struck by the simple form of this equation and its complete analogy to the special relativistic case, but also by the possibility it immediately offered for a generalization beyond the special case of static gravitational fields:

> Here too—as was proved by Planck for the usual [i.e., special] theory of relativity —it is seen that the equations of analytical mechanics possess a significance that extends far beyond Newtonian mechanics. Hamilton's equation as finally written down lets us anticipate [*ahnen*] the structure of the equations of motion of a material particle in a dynamical gravitational field.[88]

Nevertheless, according to later recollections,[89] it must have taken Einstein a while before he took the third and final step in his understanding of gravitation as the curvature of spacetime. For some time, he was indeed struck by the dilemma of what coordinates could actually mean in physics given that coordinate differences could no longer be interpreted as being the immediate result of measurements with ideal rods and clocks, as was discussed above in connection with the loss of the "homogeneity" of time in a gravitational field. Only after his return from Prague to Zurich at the end of July 1912 did Einstein realize that a generalization of the equation of motion along the lines indicated by the above quotation, yielding an expression of the form:

$$\delta\left(\int\sqrt{\sum_{i,j=1}^{4} g_{ij}dx_i dx_j}\right) = 0,$$

which now involves 10 variables g_{ij} characterizing the gravitational potential, could be assimilated to the curved-surface model as conceived in Gaussian surface theory and its successive elaborations by Bernhard Riemann, Elwin Christoffel, Tullio Levi-Civita, and others.[90] If considered from the perspective of this model, the components of the gravitational potential g_{ij} can be conceived as those of a metric tensor determining the geometry of a generalized curved surface. In this way, Einstein also found an answer to his question concerning the physical meaning of coordinates. In fact, the differentials of the general curvilinear coordinates dx_i, just serving to number spacetime points, are related to the measurable magnitude ds, representing the invariant line element, with the help of the components of the metric tensor g_{ij}, by means of the following expression:

---

88 "Auch hier zeigt sich—wie dies für die gewöhnliche Relativitätstheorie von Planck dargetan wurde—daß den Gleichungen der analytischen Mechanik eine über die Newtonsche Mechanik weit hinausreichende Bedeutung zukommt. Die zuletzt hingeschriebene Hamiltonsche Gleichung läßt ahnen, wie die Bewegungsgleichungen des materiellen Punktes im dynamischen Gravitationsfelde gebaut sind." (Einstein 1912b, 458)

89 See the preface to the Czech translation of (Einstein 1923) of his popular book on relativity. See also (Einstein 1981, 137).

90 For a historical survey of the mathematical aspects, see (Reich 1994). For the role of Einstein's Machian heuristics in this step, see "The Third Way to General Relativity" (in vol. 3 of this series).

$$ds^2 = \sum_{i,j=1}^{4} g_{ij} dx_i dx_j .$$

This expression corresponds to the term under the integral in Einstein's variational formulation of the equation of motion, as given above, and lends an immediate geometric interpretation to this formulation, which was probably at the roots of Einstein's embracing of the curved-surface model. Indeed, expressed in general curvilinear coordinates, the equation simply represents the condition for a geodesic line on such a surface. The fact that the number of indices is 4 instead of 2, as in ordinary surface theory, could hardly infringe on the plausibility of this interpretation.

Once the curved-surface model was taken into consideration and generalized to a four-dimensional context, the analogy between the generalization of the concept of a straight line to a curved geometry and that of the concept of an inertial motion to a situation in which accelerated frames of reference are admitted, must have been immediate, in particular since force-free motion and geodesics are also closely related in classical physics. A motion along a curved surface, which is not subject to any external forces, proceeds along a geodesic line—the most natural generalization of a straight line for such surfaces. This description can immediately be transferred to the case of a force-free motion that is observed from an arbitrary accelerated frame of reference in a four-dimensional framework. Using the curvilinear coordinates, with the help of which one can describe such an accelerated reference frame, this motion can now simply be represented by a geodesic line in the four-dimensional spacetime. In view of the equivalence of inertial and gravitational forces assumed by Einstein, it was plausible to extend this insight also to arbitrary gravitational fields, including those which cannot be generated by accelerated reference frames.[91]

The assimilation of the equation of motion to the generalized curved-surface model had a number of profound consequences for Einstein's theory. First of all, the curved-surface model consolidated what had merely been a speculative generalization of his equation of motion for the static case to general, dynamical gravitational fields. In particular, it became possible to immediately conclude that force-free motions in general gravitational fields are represented by geodesic world lines, and also that the gravitational potential is, in general, represented by the 10 independent (from a total of 16) components g_{ij} of a metric tensor. Second, the interpretation of Einstein's equation of motion in terms of a mental model embedded in elaborated mathematical theories opened up a plethora of technical resources available to the construction of a relativistic theory of gravitation, in particular for the search of a gravitational field equation. In mid-1912, even though Einstein himself was far from being familiar with these resources, he eventually appropriated them with the help of his mathematician friend Marcel Grossmann. Third, this interpretation had far-reach-

91 By August Einstein had found an expression for the general equation of motion in a gravitational field, see Albert Einstein to Ludwig Hopf, Zurich, 16 August 1912, (CPAE 5E, 501–502).

ing implications for the conceptual understanding of a relativistic theory of gravitation. In fact, Einstein's earlier observations concerning the implications of the equivalence principle for the concepts of space and time could now be systematically related to the newly available concepts of surface theory and its elaborations, such as the concepts of coordinates, line element, metric tensor, invariants, and curvature. As we have emphasized, these and other conceptual implications follow from the fact that the mental model that had guided Einstein's formulation of the equation of motion, the constrained-motion model, could be combined with the curved-surface model to yield the integrated curved-spacetime model. Hence motion under the influence of a gravitational field could now be conceived as a natural motion in the sense of generalized inertial motion, governed by the curvature of spacetime. In this way, it was possible to combine the synthesis of gravitation and inertia prepared by the equivalence principle with the insight into the causal structure of physical interactions represented by the metric structure of Minkowski's spacetime.

3.6 The Lorentz Model as a Problem for the Integration of Physical and Mathematical Knowledge

Einstein's pathway from his first attempt at a theory of gravitation in 1907 to his insight into the general laws of motion under the influence of a relativistic gravitational field in mid-1912 had been an unusual one, involving steps that had led him far from the ordinary course of the adaptation of the knowledge of classical physics to the framework of special relativity. This unusual pathway was marked, in particular, by the heuristics of the equivalence principle that had initiated it and by its preliminary culmination in the no-less singular recognition of the four-dimensional metric tensor as the appropriate representation of the gravitational potential. From this point onwards, the further course was, apparently, well laid out. Einstein had found what was necessary to formulate an equation of motion in a gravitational field. What was lacking was a gravitational field equation determining how this field was created by its sources, matter and energy. In fact, the twin constellation of field equation and equation of motion rooted in the tradition of classical physics and ideally embodied in Lorentz's theory of electromagnetism had directed Einstein's search for a relativistic theory of gravitation at least since he considered the unification of gravitation and inertia in analogy to the unification of the electric and magnetic fields, and certainly since Abraham had published his proposal for a field theory of gravitation, obviously shaped by the same model.

In view of the guidance provided by the Lorentz model, the completion of a relativistic theory of gravitation by an appropriate field equation might have appeared to be a rather straightforward task once the equation of motion had been found. However, the task of finding a field equation turned out to be the most challenging one among all Einstein had ever tackled in his struggle for a relativistic theory of gravitation. First of all, he was confronted with the daunting mathematical problem that the representation of the gravitational potential by the metric tensor requires a field equation not for a single function but for a 10-component object. Second, an acceptable gravitational field equation replacing the Poisson equation of classical physics had to be compatible not only with the various insights Einstein had gathered during his earlier work on the problem, such as the necessity of introducing new concepts of space and time, but also with the immense resources of knowledge on gravitation and its relation to other physical interactions accumulated by classical and special relativistic physics. Einstein had already encountered the vicissitudes of these treasures while elaborating his theory of the static gravitational field. There he had stumbled, as mentioned above, on the unpleasant discovery that what appeared to be a most natural and plausible candidate for a static field equation actually turned out to be incompatible with the inescapable requirement of the conservation of energy and momentum. This episode thus confronted him in effect with the insight that the identification of an acceptable gravitational field equation represented an even more challenging task of knowledge integration than the promising hints and intermediate results obtained due to the heuristics of the equivalence principle could have led him to believe.

In particular, Einstein could not avoid taking into account that the action of the gravitational field under ordinary circumstances was well known and satisfactorily described by Newton's law of attraction. The relativistic field equation of gravitation therefore had to yield the same results as this law under appropriate circumstances. We will refer to this requirement as the "correspondence principle," representing an important building block of the heuristics that guided Einstein's search for a gravitational field equation. The new field equation had obviously to be compatible also with the well-established knowledge on energy and momentum conservation. This requirement, which must have played an outstanding role for Einstein given the just mentioned precedent offered by the theory of the static field, is referred to in the following as the "conservation principle," another crucial building block of his heuristics. But in addition to these well-established knowledge resources, Einstein's earlier research under the guidance of the equivalence principle had revealed a number of insights that were to be covered by the new field equation, in particular, the insight into the unified nature of gravitation and inertia. As pointed out above, this insight was suggested not only by the elevator and the bucket models, but also more generally by the possibility of conceiving inertial effects in accelerated frames of reference as effects of a generalized gravitational field and, accordingly, accelerated frames as being on equal footing with inertial frames of reference. Following Einstein's own terminology, this heuristic building block is therefore called here the "generalized principle of relativity." He expected that it could be implemented by imposing the mathematical requirement of an invariance, or rather "covariance" of the gravitational field equation under as general as possible coordinate transformations.

The breakthrough of mid-1912, marked by the recognition of the key role of the metric tensor and the formulation of a general equation of motion, was the result of a perfect match between a mental model of physical knowledge and a mental model of mathematical knowledge, mediated by a mathematical representation originally motivated by physical considerations. However, when Einstein began to search for a general gravitational field equation shortly afterwards, no such match was available. In particular, no mathematical representation lent itself to the expression of all the requirements listed above, let alone one that could be assimilated to a well-established mathematical framework. It was hence not even clear whether or not these requirements were at all compatible with each other since this question could only be settled on the basis of a mathematical representation embodying them. The integration of mathematical and physical knowledge associated with the formulation of a general gravitational field equation was therefore evidently not just a matter of finding an appropriate mental model, cutting across the borderlines of classical physics as was the case for the mental models related to the equivalence principle, or of combining two matching mental models as in the case of the curved spacetime model.

As it turned out, the search for a relativistic field equation of gravitation required instead a much more complex process of research, involving the systematic examination of candidate solutions and even the elaboration of a complete theory which was then discarded again, a process in the course of which some of the fundamental knowledge structures governing its heuristics and fueling its progress eventually had to be revised. It is this revolution of fundamental knowledge structures that makes Einstein's search for the gravitational field equation in the years between 1912 and 1915 an outstanding challenge for any attempt to understand the transition from classical to modern physics from the viewpoint of an historical epistemology. While other contributions to these volumes detail the structures of this revolution, we have focused here on the roots of Einstein's heuristics in the knowledge of classical and special relativistic physics. The principle aim of this contribution was to highlight some of the conditions under which general relativity could have emerged from a transformation of the knowledge of classical and special relativistic physics, and also to help understand the first steps of Einstein's pathway out of the disarray of classical physics.

REFERENCES

Abraham, Max. 1912. "Zur Theorie der Gravitation." *Physikalische Zeitschrift* 13: 1–4. (English translation in vol. 3 of this series.)

Ashtekar, Abhay et al. (eds.). 2003. *Revisiting the Foundations of Relativistic Physics: Festschrift in Honor of John Stachel*. Dordrecht: Kluwer.

Bödecker, Katja. 2004. *Die Entwicklung intuitiven physikalischen Denkens im Kulturvergleich*. Dissertation, Technische Universität, Berlin.

Born, Max. 1910. "Über die Definition des starren Körpers in der Kinematik des Relativitätsprinzips." *Physikalische Zeitschrift* 6: 233–234.

Borzeszkowski, Horst H. 1993. "Hegel's Interpretation of Classical Mechanics." In (Petry 1993).

Brush, Stephen G. 1976. *The Kind of Motion We Call Heat: A History of the Kinetic Theory of Gases in the 19th Century. 1: Physics and the Atomists*. (*Studies in Statistical Mechanics* 6.) Amsterdam: North Holland.

Buchwald, Jed Z. 1985. *From Maxwell to Microphysics*: *Aspects of Electromagnetic Theory in the Last Quarter of the Nineteenth Century*. Chicago: University of Chicago Press.

———. 1989. *The Rise of the Wave Theory of Light: Optical Theory and Experiment in the Early Nineteenth Century*. Chicago: University of Chicago Press.

Büttner, Jochen. 2001. "Galileo's Cosmogony." In (Montesinos and Solís 2001, 391–402).

Büttner, Jochen, Peter Damerow, and Jürgen Renn. 2001. "Traces of an Invisible Giant: Shared Knowledge in Galileo's Unpublished Treatises." In (Montesinos and Solís 2001, 183–201).

Büttner, Jochen, Peter Damerow, Jürgen Renn, and Matthias Schemmel. 2003. “The Challenging Images of Artillery - Practical Knowledge at the Roots of the Scientific Revolution.” In W. Lefèvre, J. Renn, and U. Schoepflin (eds.), *The Power of Images in Early Modern Science*. Basel: Birkhäuser, 3–27.

Cantor, Geoffrey N., and Michael J. S. Hodge (eds.). 1981. *Conceptions of Ether: Studies in the History of Ether Theories, 1740–1900*. Cambridge: Cambridge University Press.

CPAE 2. 1989. John Stachel, David C. Cassidy, Jürgen Renn, and Robert Schulmann (eds.), *The Collected Papers of Albert Einstein*. Vol. 2. *The Swiss Years: Writings, 1900–1909*. Princeton: Princeton University Press.

CPAE 3. 1993. Martin J. Klein, A. J. Kox, Jürgen Renn, and Robert Schulmann (eds.), *The Collected Papers of Albert Einstein*. Vol. 3. *The Swiss Years: Writings, 1909–1911*. Princeton: Princeton University Press.

CPAE 4. 1995. Martin J. Klein, A. J. Kox, Jürgen Renn, and Robert Schulmann (eds.), *The Collected Papers of Albert Einstein*. Vol. 4. *The Swiss Years: Writings, 1912–1914*. Princeton: Princeton University Press.

CPAE 5. 1993. Martin J. Klein, A. J. Kox, and Robert Schulmann (eds.), *The Collected Papers of Albert Einstein*. Vol. 5. *The Swiss Years: Correspondence, 1902–1914*. Princeton: Princeton University Press.

CPAE 5E. 1995. *The Collected Papers of Albert Einstein*. Vol. 5. *The Swiss Years: Correspondence, 1902–1914*. English edition translated by Anna Beck. Princeton: Princeton University Press.

CPAE 6. 1996. A. J. Kox, Martin J. Klein, and Robert Schulmann (eds.), *The Collected Papers of Albert Einstein*. Vol. 6. *The Berlin Years: Writings, 1914–1917*. Princeton: Princeton University Press.

CPAE 7. 2002. Michel Janssen, Robert Schulmann, József Illy, Christoph Lehner, and Diana Kormos Buchwald (eds.), *The Collected Papers of Albert Einstein*. Vol. 7. *The Berlin Years: Writings, 1918–1921*. Princeton: Princeton University Press.

D’Abro, A. de.1951a. *The Rise of the New Physics, Vol. 1: Its mathematical and physical theories*. New York: Dover.

———. 1951b. *The Rise of the New Physics, Vol. 2: The rise of the new physics*. New York: Dover.

Damerow, Peter. 1979. “Handlung und Erkenntnis in der genetischen Erkenntnistheorie Piagets und in der Hegelschen “Logik”.” In W. R. Beyer (ed.), *Hegel-Jahrbuch 1977/1978*. Köln: Pahl-Rugenstein, 136–160.

———. 1996. *Abstraction and Representation: Essays on the Cultural Revolution of Thinking*. Dordrecht: Kluwer.

Damerow, Peter, Gideon Freudenthal, Peter McLaughlin, and Jürgen Renn. 2004. *Exploring the Limits of Preclassical Mechanics*, (2nd ed.). New York: Springer-Verlag.

Damerow, Peter, and Wolfgang Lefèvre. 1980. “Die wissenschaftshistorische Problemlage für Hegels “Logik”.” In W. R. Beyer (ed.), *Hegel-Jahrbuch 1979*. Köln: Pahl-Rugenstein, 349–368.

Damerow, Peter, Jürgen Renn and Simone Rieger. 2002. “Mechanical Knowledge and Pompeian Balances.” In G. Castagnetti and J. Renn (eds.), *Homo Faber: Studies on Nature, Technology, and Science at the Time of Pompeii*. Rome: L’Erma di Bretschneider, 93–108.

Damerow, Peter and Jürgen Renn. 2001. “Scientific Revolution: History and Sociology of.” In N. J. Smelser and P. B. Baltes (eds.), *International Encyclopedia of the Social and Behavioral Sciences*. Elsevier Science Ltd., 13749–13752.

Darrigol, Olivier. 1994. “The Electron Theories of Larmor and Lorentz: a comparative study.” *Historical Studies in the Physical and Biological Sciences* 24: 265–336.

———. 2000. *Electrodynamics from Ampère to Einstein*. Oxford: Oxford University Press.

Davis, Robert. 1984. *Learning Mathematics: The Cognitive Approach to Mathematics Education*. New Jersey: Ablex Publishing Corporation.

Dijksterhuis, Eduard J. 1986. *The Mechanization of the World Picture: Pythagoras to Newton*. Princeton: Princeton University Press.

Dundon, Stanislaus J. 1972. *Philosophical Resistance to Newtonianism on the Continent 1700–1760*. PhD Dissertation., St. John’s University, New York.

Earman, John and Clark Glymour. 1980. “Relativity and Eclipses: The British Expeditions of 1919 and Their Predecessors.” *Historical Studies in the Physical Sciences* 11: 49–85.

Edward, Matthew R. (ed.). 2002. *Pushing Gravity: New Perspectives on Le Sage’s Theory of Gravitation*. Montreal: Apeiron.

Ehrenfest, Paul. 1909. “Gleichförmige Rotation starrer Körper und Relativitätstheorie.” *Physikalische Zeitschrift* 10: 918.

Einstein, Albert. 1905a. “Ist die Trägheit eines Körpers von seinem Energieinhalt abhängig?” *Annalen der Physik* 18: 639–641, (CPAE 2, Doc. 24).

———. 1905b. “Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt.” *Annalen der Physik* 17: 132–148, (CPAE 2, Doc. 14).

———. 1905c. “Zur Elektrodynamik bewegter Körper.” *Annalen der Physik* 17: 891–921, (CPAE 2, Doc. 23).

———. 1907. “Über das Relativitätsprinzip und die aus demselben gezogenen Folgerungen.” *Jahrbuch für Radioaktivität und Elektronik* 4: 411–462, (CPAE 2, Doc. 47).
———. 1911. “Über den Einfluß der Schwerkraft auf die Ausbreitung des Lichtes.” *Annalen der Physik* 35: 898–908, (CPAE 3, Doc. 23).
———. 1912a. “Lichtgeschwindigkeit und Statik des Gravitationsfeldes.” *Annalen der Physik* 38: 355–369, (CPAE 4, Doc. 3).
———. 1912b. “Zur Theorie des statischen Gravitationsfeldes.” *Annalen der Physik* 38: 443–458, (CPAE 4, Doc. 4).
———. 1913. “Zum gegenwärtigen Stande des Gravitationsproblems.” *Physikalische Zeitschrift* 14: 1249–1262, (CPAE 4, Doc. 17). (English translation in vol. 3 of this series.)
———. 1915a. “Die Feldgleichungen der Gravitation.” *Königlich Preußische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte* (1915) XLVIII–XLIX: 844–847, (CPAE 6, Doc. 25).
———. 1915b. “Erklärung der Perihelbewegung des Merkur aus der allgemeinen Relativitätstheorie.” *Sitzungsberichte der Preußischen Akademie der Wissenschaften: 2. Halbband* XLVII: 831–839, (CPAE 6, Doc. 24).
———. 1915c. “Zur allgemeinen Relativitätstheorie.” *Königlich Preußische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte*: 778–786, (CPAE 6, Doc. 21).
———. 1915d. “Zur allgemeinen Relativitätstheorie. (Nachtrag).” *Königlich Preußische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte*: 799–801, (CPAE 6, Doc. 22).
———. 1919. “Spielen Gravitationsfelder im Aufbau der materiellen Elementarteilchen eine wesentliche Rolle?” *Preussische Akademie der Wissenschaften, Sitzungsberichte* 349–356, (CPAE 7, Doc. 17).
———. 1920. *Äther und Relativitätstheorie. Rede. Gehalten am 5. Mai 1920 an der Reichs-Universität zu Leiden*. Berlin: Julius Springer. (English translation in vol. 3 of this series.)
———. 1923. *Theorie relativity speciální i obecná: Lehce srozumitelný výklad*. Prague: F. Borový.
———. 1933. *The Origins of the General Theory of Relativity. Being the first Lecture of the Georg A. Gibson Foundation in the University of Glasgow delivered on June 20th, 1933*. Glasgow: Jackson, Wylie and Co.
———. 1954. *Ideas and opinions by Albert Einstein*. New York: Crown. (Edited by Carl Seelig, new translations and revisions by Sonja Bargmann.)
———. 1979. *Autobiographical Notes: A Centennial Edition*. La Salle/Chicago: Open Court.
———. 1981. *Mein Weltbild*. Frankfurt am Main/Berlin/Wien: Ullstein (first published 1934).
Einstein, Albert, and Marcel Grossmann. 1913. *Entwurf einer verallgemeinerten Relativitätstheorie und einer Theorie der Gravitation*. Leipzig: Teubner, (CPAE 4, Doc. 13).
Elkana, Yehuda (ed.). 1974. *The Discovery of the Conservation of Energy.* London: Hutchinson.
Elkana, Yehuda. 1981. “A Programmatic Attempt at an Anthropology of Knowledge.” In E. Mendelsohn and Y. Elkana (eds.), *Sciences and Cultures. Sociology of the Sciences*, vol. 5. Dordrecht: Reidel, 1–76.
Freudenthal, Gideon. 1986. *Atom and Individual in the Age of Newton: On the Genesis of the Mechanistic World View.* (Translated by Peter McLaughlin.) Dordrecht: Reidel.
Fölsing, Albrecht. 1993. *Albert Einstein: eine Biographie*. Frankfurt am Main: Suhrkamp.
Gentner, Dedre and Albert L. Stevens. 1983. *Mental Models*. Hillsdale, NJ: Erlbaum.
Harman, Peter M. 1982. *Energy, Force, and Matter: The Conceptual Development of Nineteenth-Century Physics*. Cambridge: Cambridge University Press.
Hertz, Heinrich. 1894. *Die Prinzipien der Mechanik in neuem Zusammenhange dargestellt*. Leipzig: Johann Ambrosius Barth.
———. 1956. *The Principles of Mechanics: Presented in a New Form*. New York: Dover Publications.
———. 1999. *Die Constitution der Materie: eine Vorlesung über die Grundlagen der Physik aus dem Jahre 1884*. (Edited by Albrecht Fölsing.) Berlin/Heidelberg: Springer.
Hirosige, Tetu. 1966. “Electrodynamics Before the Theory of Relativity, 1890–1905.” *Japanese Studies in the History of Scienc*e 5: 1–49.
———. 1969. “Origins of Lorentz’s Theory of Electrons and the Concept of the Electromagnetic Field.” *Historical Studies in the Physical Sciences* 1: 151–209.
Hegel, Georg Wilhelm Friedrich. 1986a. *Dissertatio philosophica de orbitis planetarum/Philosophische Erörterung über die Plantenbahnen*. Weinheim: Acta Humaniora, VCH. (Translated, introduced, and commented by Wolfgang Neuser).
———. 1986b. *Phänomenologie des Geistes*. Vol. 3 of *Werke* (20 vols.). Frankfurt am Main: Suhrkamp. (Translated 1977 by A. V. Miller in *Hegel’s Phenomenology of Spirit*. Oxford: Oxford University Press.)
Holton, Gerald. 1971. “On Trying to Understand Scientific Genius.” *The American Scholar* 41: 95–110.
Ihmig, Karl-Norbert. 1989. *Hegels Deutung der Gravitation*. Frankfurt am Main: Athneäum.

Janssen, Michel. 1995. *A Comparison between Lorentz's Ether Theory and Special Relativity in the Light of the Experiments of Trouton and Noble*. Dissertation, University of Pittsburgh.

———. 2003. "The Trouton Experiment." In (Ashtekar et al. 2003, 27–54).

Janssen, Michel and John Stachel. 2004. *The Optics and Electrodynamics of Moving Bodies*. Preprint 265, Max Planck Institute for the History of Science. Forthcoming in John Stachel's *Going Critical* (Dordrecht: Springer).

Jungnickel, Christa and Russel McCormmach. 1986a. *The Now Mighty Theoretical Physics. Intellectual Mastery of Nature*, vol. 2. Chicago and London: The University of Chicago Press.

Jungnickel, Christa and Russel McCormmach. 1986b. *The Torch of Mathematics 1800–1870. Intellectual Mastery of Nature*, vol. 1. Chicago and London: The University of Chicago Press.

Kollross, Louis. 1955. "Erinnerungen-Souvenirs." In C. Seelig (ed.), *Helle Zeiten - Dunkle Zeiten: In Memoriam Albert Einstein*. Zürich/Stuttgart/Wien: Europa Verlag, 17–31.

Kuhn, Thomas S. 1977. *The Essential Tension: Selected Studies in Scientific Tradition and Change*. Chicago: University of Chicago Press.

Lange, Ludwig. 1883. *Die geschichtliche Entwicklung des Bewegungsbegriffs*. Leipzig: Teubner.

———. 1885a. "Über das Beharrungsgesetz." *Berichte der Königlichen-Sächsischen Gesellschaft der Wissenschaft zu Leipzig, mathematisch-physikalisch Klasse* 36: 335–351.

———. 1885b. "Über die wissenschaftliche Fassung des Beharrungsgesetzes." *Wundts philosophische Studien* 2: 276–297.

———. 1886. *Der Bewegungsbegriff der Himmelskunde von Copernicus bis zu Newton (1543–1687)*. Leipzig: Wilhelm Engelmann.

Langley, Samuel P. 1898. "The Lesage Theory of Gravitation, translated by C. G. Abbot." In *Smithsonian Report*, 139–160.

Laudan, Laurens. 1972. "Georg-Louis LeSage: A case study in the interaction of physics and philosophy." *Akten des II. Internationalen Leibniz-Kongreß*, 2: 241–252.

Laue, Max von. 1911a. *Das Relativitätsprinzip*. Braunschweig: Vieweg.

———. 1911b. "Zur Dynamik der Relativitätstheorie." *Annalen der Physik* 35: 524–542.

Le Sage, George-Louis. 1758. *Essai de chimie méchanique*. Geneva: Rouen.

Lorentz, Hendrik A. 1892. "La Théorie électromagnétique de Maxwell et son application aux corps mouvants." *Archives Néerlandaises des Sciences Exactes et Naturelles* 25: 363–552.

———. 1895. *Versuch einer Theorie der electrischen [elektrischen] und optischen Erscheinungen in bewegten Körpern*. Leiden: Brill.

Lützen, Jesper. 1993. *Interactions between Mechanics and Differential Geometry in the 19th Century*. Preprint Series, vol. 25. København: Københavns Universitet. Matematisk Institut.

Mach, Ernst. 1921. *Die Prinzipien der physikalischen Optik: historisch und erkenntnispsychologisch entwickelt*. Leipzig: J. A. Barth.

———. 1960. *The Science of Mechanics*. La Salle, Illinois: Open Court.

McCloskey, Michael. 1983. "Naive Theories of Motion." In (Gentner and Stevens 1983).

Miller, Arthur I. 1992. "Albert Einstein's 1907 Jahrbuch Paper: The First Step from SRT to GRT." In J. Eisenstaedt and A. J. Kox (eds.), *Studies in the History of General Relativity.* (*Einstein Studies*, Vol. 3.) Boston/Basel/Berlin: Birkhäuser, 319–335.

Minkowski, Hermann. 1908. "Raum und Zeit." *Physikalische Zeitschrift* (10) 3: 104–111.

Minsky, Marvin. 1987. *The Society of Mind*. London: Heinemann.

Montesinos, José and Carlos Solís (eds.). 2001. *Largo Campo di Filosofare*. Orotava: Fundación Canaria Orotava de Historia de la Ciencia.

Newton, Isaac. 1999. *Die mathematischen Prinzipien der Physik*, edited and translated by Volkmar Schüller. Berlin: de Gruyter.

Pais, Abraham. 1982. *Subtle is the Lord. The Science and Life of Albert Einstein*. Oxford/New York/Toronto/Melbourne: Oxford University Press.

Petry, Michael J. 1993. *Hegel and Newtonianism*. (*International Archives of the History of Ideas,* vol. 136.) Dordrecht: Kluwer Academic Publishers.

Planck, Max. 1906. "Das Prinzip der Relativität und die Grundgleichungen der Mechanik." *Deutsche Physikalische Gesellschaft. Verhandlungen* 8: 136–141.

———. 1907. "Zur Dynamik bewegter Systeme." *Königlich Preußische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte,* 542–570.

Prévost, Pierre. 1805. *Notice de la vie et des écrits de George-Louis Le Sage de Genève*. Geneva.

Pulte, Helmut. 1989. *Das Prinzip der kleinsten Wirkung und die Kraftkonzeptionen der rationalen Mechanik: eine Untersuchung zur Grundlegungsproblematik bei Leonhard Euler, Pierre Louis Moreau de Maupertuis und Joseph Louis Lagrange*. Stuttgart: Steiner.

Reich, Karin. 1994. *Die Entwicklung des Tensorkalküls: vom absoluten Differentialkalkül zur Relativitätstheorie*. (*Science Networks: Historical Studies*, Vol. 11.) Basel/Boston/Berlin: Birkhäuser.

Renn, Jürgen (ed.). 2001. *Galileo in Context*. Cambridge: Cambridge University Press.
Renn, Jürgen. 2004. "Die klassische Physik vom Kopf auf die Füße gestellt: Wie Einstein die Spezielle Relativitätstheorie fand." *Physik Journal* März: 49–55.
Renn, Jürgen, and Tilmann Sauer. 2003. "Eclipses of the Stars - Mandl, Einstein, and the Early History of Gravitational Lensing." In (Ashtekar et al. 2003, 69–92).
Renn, Jürgen, Tilman Sauer and John Stachel. 1997. "The Origin of Gravitational Lensing. A Postscript to Einstein's 1936 Science Paper." *Science* (275) 5297: 184–186.
Rynasiewicz, Robert, and Jürgen Renn. 2006. "The Turning Point for Einstein's *Annus Mirabilis*." *Studies in History and Philosophy of Modern Physics* 37: 5–35.
Schaffner, Kenneth F. 1972. *Nineteenth-Century Aether Theories*. Oxford: Pergamon Press.
Stachel, John. 1989. "The Rigidly Rotating Disk as the "Missing Link" in the History of General Relativity." In D. Howard and J. Stachel (eds.), *Einstein and the History of General Relativity.* (*Einstein Studies*, Vol. 1.) Boston, Basel, and Berlin: Birkhäuser, 48–62.
———. 2005. "Fresnel's (Dragging) Coefficient as a Challenge to 19th Century Optics of Moving Bodies." In J. Eisenstaedt and J. J. Kox (eds.), *The Universe of General Relativity.* (Einstein Studies, Vol. 11.) New York: Birkhäuser.
Staley, Richard. 2005. "On the Co-Creation of Classical and Modern Physics." *Isis* 96:530–558.
Steinle, Freidrich. 2005. *Explorative Experimente: Ampère, Faraday und die Ursprünge der Elektrodynamik*. Stuttgart: Steiner.
Stichweh, Rudolf. 1984. *Zur Entstehung des modernen Systems wissenschaftlicher Disziplinen: Physik in Deutschland 1740–1890*. Frankfurt am Maine: Suhrkamp.
Taylor, William B. 1876. "Kinetic Theories of Gravitation." In *Annual Report of the Board of Regents of the Smithsonian Institute*. 205–282.
Vishveshwara, C. V. 2003. "Rigidly Rotating Disk Revisited." In (Ashtekar et al. 2003, 305–316).
Warwick, Andrew. 1995. "The Sturdy Protestants of Science: Larmor, Trouton, and the Earth's Motion Through the Ether." In J. Buchwald (ed.), *Scientific Practice. Theories and Stories of Doing Physics*. Chicago: University of Chicago Press.
Wahsner, Renate. 1993. "Hegel's Interpretation of Classical Mechanics." In (Petry 1993).
Wheaton, Bruce R. 1983. *The Tiger and the Shark: Empirical Roots of Wave-Particle Dualism*. Cambridge: Cambridge University Press.
Whittaker, Edmund Taylor. 1951–53. *A History of the Theories of Aether and Electricity.* London, New York: T. Nelson.

JOHN STACHEL

THE FIRST TWO ACTS

PROLOGUE: THE DEVELOPMENT OF GENERAL RELATIVITY, A DRAMA IN THREE ACTS

In 1920, Einstein wrote a short list of "my most important scientific ideas".[1] The final three items on the list are:

1907 Basic idea for the general theory of relativity

1912 Recognition of the non-Euclidean nature of the metric and its physical determination by gravitation

1915 Field equations of gravitation. Explanation of the perihelion motion of Mercury.

Einstein's words provide the warrant for comparing the development of general relativity to a three-act drama:

Act I (1907) The formulation of the "basic idea," to which he soon referred as the equivalence principle.

Act II (1912) The mathematical representation of the gravitational field by a symmetric second rank tensor field, which enters into the line element of a four-dimensional spacetime; hence this tensor is usually referred to as the (pseudo-)metric of spacetime.[2]

Act III (1915) The formulation of the now-standard Einstein field equations for the metric field, and use of its spherically-symmetric solution to explain the anomalous precession of the perihelion of Mercury.

1 Einstein Archives, Hebrew University of Jerusalem, Control Index No. 11 196. (Hereafter, only the number of such items will be cited.) It appears from Einstein to Robert Lawson, 22 April 1920 (1–010), that it was written for the biographical note in (Einstein 1920a). This is the English translation of the fifth German edition of (Einstein 1920b). I thank Dr. Josef Illy of the Einstein Papers for locating Einstein's letter to Lawson. In (CPAE 8), it is incorrectly calendared under 1917 (see pp. 1005–1006).

2 Properly speaking, the term "metric" should be restricted to line elements with positive-definite signature; those with an indefinite signature are more properly termed "pseudo-metrics." But Einstein, and following him most physicists, referred to the four-dimensional tensor field with Minkowski signature as the metric tensor, and I shall follow that usage.

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 1
Einstein's Zurich Notebook: *Introduction and Source*.

Act III certainly does not represent the end of general relativity, but a certain point of closure in its development, signaled by Einstein in his 1916 review paper (Einstein 1916a):

> "According to the general theory of relativity, gravitation plays an exceptional role as opposed to the other forces, in particular the electromagnetic..." (p. 779);

consequently, he concluded his exposition with a "Theory of the Gravitational Field" (pp. 801–822).

Up to this point, the story had been essentially an account of Einstein's struggles.[3] Now that the final form of the gravitational field equations had been achieved and one of its predictions validated, the theory became the property of the physics and astronomy communities. Its further development and interpretation became a subject of discussion among many participants, among whom Einstein's voice did not always carry the day.[4]

This book tells the story from the opening curtain of Act I in 1907 until the curtain goes down on Act III at then end of 1915. The great bulk of it is devoted to Act III, embracing the events between 1912 and 1915. In particular, it centers on the understanding of the Zurich Notebook, which opens Act III and has made a signal contribution to our understanding of subsequent events. As in most plays, the final act contains the *dénouement*; but the crucial events that lead up to it take place in the first two acts. It was his formulation of the equivalence principle in Act I, and constant adherence to it as the guiding thread in his search for a relativistic theory of gravitation that set Einstein apart from other physicists who were working on the problem of fitting gravitation within the framework of the (special) theory of relativity. And as usual, the high point of the drama comes in Act II, with Einstein's remarkable decision to represent gravitation, not by a scalar field, but by the ten components of a tensor field that also describes the chronogeometry of a non-flat four-dimensional spacetime.

It is at this point, with an expression for the line element in terms of the metric tensor field, that the Zurich Notebook opens. Clearly, it cannot be fully understood or evaluated without prior knowledge of what happened during the first two acts. Unfortunately, no equivalent of the Zurich Notebook has been found for this period from 1907–1912. So this chapter attempts to present what can be learned—or surmised—about what happened on the basis of Einstein's published papers and correspondence, as well as his later reminiscences.[5] In keeping with the documentary character of this book, rather than attempting to summarize them, I shall often cite Einstein's words *in extenso*.

---

3 David Hilbert played a significant role in the final moments of Act III, although not the one that is often attributed to him. See "Hilbert's Foundation of Physics: From a Theory of Everything to a Constituent of General Relativity" (in vol. 4 of this series).

4 For Einstein's side of this discussion during its first years, see (CPAE 7).

5 For another account, see (Pais 1982, Part IV, 177–296). For some critical comments, see (Stachel 1982); reprinted in (Stachel 2002, 551–554).

ACT I: THE EQUIVALENCE PRINCIPLE: "THE MOST FORTUNATE THOUGHT OF MY LIFE"

In 1920,[6] Einstein recalled how he first arrived at the ideas behind the equivalence principle:

> While I was occupied (in 1907) with a comprehensive survey of the special theory[7] for the "Yearbook for Radioactivity and Electronics," I also had to attempt to modify Newton's theory of gravitation in such a way that its laws fitted into the theory. Attempts along these lines showed the feasibility of this enterprise, but did not satisfy me, because they had to be based on physical hypotheses that were not well-founded. Then there came to me the most fortunate thought of my life in the following form:
>
> Like the electric field generated by electromagnetic induction, ... the gravitational field only has a relative existence. *Because, for an observer freely falling from the roof of a house, during his fall there exists*—at least in his immediate neighborhood—*no gravitation field*. Indeed, if the observer lets go of any objects, relative to him they remain in a state of rest or uniform motion, independently of their particular chemical or physical composition [note by AE: air resistance is naturally ignored in this argument]. The observer is thus justified in interpreting his state as being at rest.
>
> Through these considerations, the unusually extraordinary experimental law, that all bodies fall with equal acceleration in the same gravitational field, immediately obtains a deep physical significance. For if there were just one single thing that fell differently from the others in the gravitational field, then with its help the observer could recognize that he was falling in a gravitational field. If such a thing does not exist—which experiment has shown with great precision—then there is no objective basis for the observer to regard himself as falling in a gravitational field. Rather, he has the right to regard his state as one of rest and, with respect to a gravitational field, his neighborhood as field free. The experimental fact of the material-independence of the acceleration due to gravity is thus a powerful argument for the extension of the relativity postulate to coordinate systems in non-uniform relative motion with respect to each other .... The generalization of the relativity principle thus indicates a speculative path towards the investigation of the properties of the gravitational field (pp. 24–25).

Einstein alludes here to his initial attempts to set up a special-relativistic theory of gravitation, but gives no details. In 1933 he gave the fullest account of how he "arrived at the equivalence principle by a detour [*Umweg*]"[8] through such attempts.[9]

---

6 "Grundgedanken und Methoden der Relativitätstheorie, in ihrer Entwicklung dargestellt," in (CPAE 7, Doc. 31).

7 I shall follow the common but anachronistic practice of referring to "the special theory of relativity." During this period, Einstein initially called it "the principle of relativity" [*das Relativitätsprinzip*] and then, following the practice of others, "the theory of relativity" [*die Relativitätstheorie*]. For details, see the discussion in the Editorial Headnote, "Einstein on the Theory of Relativity," (CPAE 2, 254); reprinted in (Stachel 2002, 192).

8 "Erinnerungen-Souvenirs" (Einstein 1955, 145–153) was reprinted as "Autobiographische Skizze," in (Seelig 1955, 9–17). Citation from "Autobiographische Skizze," p. 14.

9 "Einiges über die Entstehung der allgemeinen Relativitätstheorie," the German text of a lecture given at the University of Glasgow, 20 June 1933. The German text was published in (Einstein 1934, 248–256). Cited from the paperback edition edited by Carl Seelig: (Seelig 1981, 134–138).

After mentioning his doubts after 1905 about the privileged dynamical role of inertial systems, and his early fascination by Mach's idea that the acceleration of a body is not absolute, but relative to the rest of the bodies in the universe, he turns to the events of 1907:

> I first came a step closer to the solution of the problem when I attempted to treat the law of gravitation within the framework of special relativity. Like most authors at the time, I attempted to establish a field law for gravitation, since the introduction of an unmediated action at a distance was no longer possible, at least in any sort of natural way, on account of the abolition of the concept of absolute simultaneity.
>
> The simplest thing naturally was to preserve the Laplacian scalar gravitational potential and to supplement Poisson's equation in the obvious way by a term involving time derivatives, so that the special theory of relativity was satisfactorily taken into account. The equation of motion of a particle also had to be modified to accord with the special theory. The way to do so was less uniquely prescribed, since the inertial mass of a body might well depend on its gravitational potential. This was even to be expected on the basis of the law of the inertia of energy.
>
> However, such investigations led to a result that made me highly suspicious. For according to classical mechanics, the vertical acceleration of a body in a vertical gravitational field is independent of the horizontal component of its velocity. This is connected with the fact that the vertical acceleration of a mechanical system, or rather of its center of mass, in such a gravitational field turns out to be independent of its internal kinetic energy. According to the theory I was pursuing, however, such an independence of the gravitational acceleration from the horizontal velocity, or from the internal energy of a system, did not occur.[10]
>
> This did not accord with an old fact of experience, that all bodies experience the same acceleration in a gravitational field. This law, which can also be formulated as the law of equality of inertial and gravitational mass, now appeared to me in its deep significance. I was most highly amazed by it and guessed that in it must lie the key to the deeper understanding of inertia and gravitation (pp. 135–136).

Turning from later reminiscences, let us see how Einstein presented his approach to gravitation in 1907:[11]

> Up to now we have only applied the principle of relativity, i.e., the presupposition that the laws of nature are independent of the state of motion of the reference system, to *acceleration-free* reference systems. Is it conceivable that the principle of relativity also holds for systems that are accelerated relative to each other?
>
> This is not the place for an exhaustive treatment of this question. Since, however, it is bound to occur to anyone who has followed the previous applications of the relativity principle, I shall not avoid taking a position on the question here.

---

10 As Einstein later realized, a special-relativistic theory of gravitation that does justice to the equivalence principle is possible, and indeed one was developed a little later by Gunnar Nordström. For an account of Nordström's theories, and Einstein's reaction to them, see (Norton 1992).

11 "Über das Relativitätsprinzip und die aus demselbem gezogenen Folgerungen," (Einstein 1907); "Berichtigungen," in (Einstein 1908).

> Consider two systems in motion Σ_1 and Σ_2. Let Σ_1 be accelerated in the direction of its X-axis, and let γ be the magnitude (constant in time) of this acceleration. Let Σ_2 be at rest, but in a homogeneous gravitational field that imparts an acceleration $-\gamma$ in the direction of the X-axis to all objects. As far as we know, the laws of physics with respect to Σ_1 do not differ from those with respect to Σ_2; this is due to the circumstance that all bodies in a gravitational field are equally accelerated. So we have no basis in the current state of our experience for the assumption that the systems Σ_1 and Σ_2 differ from each other in any respect; and therefore in what follows shall assume the complete physical equivalence of a gravitational field and the corresponding acceleration of a reference system.
>
> This assumption extends the principle of relativity to the case of uniformly-accelerated translational motion of the reference system. The heuristic value of this assumption lies in the circumstance that it allows the replacement of a homogeneous gravitational field by a uniformly accelerated reference system, which to a certain extent is amenable to theoretical treatment.[12]

Some further comments on this equivalence in his next paper on gravitation in 1911[13] are illuminating. He notes that in both systems, objects subject to no other forces fall with constant acceleration:

> For the accelerated system K' [corresponding to the 1907 Σ_1–JS], this follows directly from the Galileian principle [of inertia–JS]; for the system K at rest in a homogeneous gravitational field [corresponding to the 1907 Σ_2–JS], however, it follows from the experimental fact that in such a field all bodies are equally strongly uniformly accelerated. This experience of the equal falling of all bodies in a gravitational field is the most universal with which the observation of nature has provided us; in spite of that, this law has not found any place in the foundations of our physical picture of the world. ... From this standpoint one can as little speak of the *absolute acceleration* of a reference system, as one can of the *absolute velocity* of a system according to the usual [special–JS] theory of relativity. [note by AE: Naturally, one cannot replace an *arbitrary* gravitational field by a state of motion of the system without a gravitational field; just as little as one can transform all points of an arbitrarily moving medium to rest by a relativity transformation.] From this standpoint the equal falling of all bodies in a gravitational field is obvious.
>
> As long as we confine ourselves to purely mechanical processes within the realm of validity of Newtonian mechanics, we are certain of the equivalence of the systems K and K'. Our point of view will only have a deeper significance, however, if the systems K and K' are equivalent with respect to all physical processes, i.e., if the laws of nature with respect to K agree completely with those with respect to K'. By assuming this, we obtain a principle that, if it really is correct, possesses a great heuristic significance. For by means of theoretical consideration of processes that take place relative to a uniformly accelerated reference system, we obtain conclusions about the course of processes in a homogeneous gravitational field. (CPAE 3, 487–488)

With hindsight, one can see that Einstein's attempt to find the best way to implement mathematically the physical insights about gravitation incorporated in the equivalence principle was hampered significantly by the absence of the appropriate mathematical concepts. His insight, as he put is a few years later, that gravitation and inertia are "essentially the same" [*wesensgleich*],[14] cries out for implementation by

---

12 See (CPAE 7, 476). Also see p. 495 for a discussion of the meaning of uniform acceleration.
13 "Über den Einfluß der Schwerkraft auf die Ausbreitung des Lichtes," (Einstein 1911b).

their incorporation into a single inertio-gravitational field, represented mathematically by a non-flat affine connection on a four-dimensional manifold. But the concept of such a connection was only developed *after*, and largely in response to, the formulation of the general theory. So Einstein had to make do with what was available: Riemannian geometry and the tensor calculus as developed by the turn of the century, i.e., based on the concept of the metric tensor, without a geometrical interpretation of the covariant derivative. As I have suggested elsewhere, this absence is largely responsible for the almost three-year lapse between the end of Act I and the close of the play.[15]

ACT II: THE METRIC TENSOR: "JUST WHAT ARE COORDINATES ACTUALLY SUPPOSED TO MEAN IN PHYSICS?"

In 1949, Einstein himself raised the question of what was responsible for this long delay:

> This [recognition that the relativity principle had to be extended to non-linear transformations–JS] took place in 1908. Why were a further seven years required for setting up the general theory of relativity? The principal reason is that one does not free oneself so easily from the conception that an immediate physical significance must be attributed to the coordinates.[16]

Both the question and answer thus concern the entire period between 1907 (or 1908) and 1915. In 1933 Einstein made the answer more precise, and confined it to a shorter period of time:

> I soon saw that, according the point of view about non-linear transformations required by the equivalence principle, the simple physical interpretation of the coordinates had to be abandoned; i.e., one could no longer require that coordinate differences be interpreted as signifying the immediate results of measurements with ideal measuring rods and clocks. This recognition tormented me a great deal because for a long time I was not able to see just what *are* coordinates actually supposed to mean in physics? The resolution of this dilemma was reached around 1912. (Seelig 1981, 137)[17]

Einstein reference to 1912 is a clear allusion to his introduction of the metric tensor. But, as his reference to "a further seven years" after 1908 in the previous quotation suggests, the problem of the meaning of coordinates in general relativity was by no

14 In 1912 Einstein regarded "the equivalence of inertial and gravitational mass" as being reducible to the "essential likeness [*Wesensgleichheit*] of both of these elementary qualities of matter and energy"; and asserted that his theory of "the static gravitational field" allows him to regard it "as physically the same in essence [*wesensgleich*] as an acceleration of the reference system." See (Einstein 1912c, 1063).

15 The entire complex of problems raised in this paragraph is discussed at length in "The Story of Newstein or: Is Gravity just Another Pretty Force?" (in vol. 4 of this series).

16 From Albert Einstein's "Autobiographical Notes," which, although published first in 1949 (Einstein 1949, 2–94), were actually written in 1947. Cited here from (Einstein 1979, 63).

17 See (Seelig 1981, 1).

means completely resolved with the introduction of the metric tensor. Only with the resolution in 1915 of the "hole argument" [*Lochbetrachtung*] against general covariance that Einstein developed in 1913, did Einstein fully solve this problem; but discussion of the post-1912 aspects of the question will be found in later chapters.[18]

Now let me return to the problem of coordinates as Einstein saw it in 1907–1908. It is worth emphasizing that Einstein attributed his success in formulating the special theory in 1905 in no small measure to his insistence on physically defining coordinate systems that allow one to attach direct physical significance to coordinate differences:

> The theory to be developed—like every other electrodynamics—rests upon on the kinematics of rigid bodies, since the assertions of each such theory concern relations between rigid bodies (coordinate systems), clocks and electromagnetic processes. Taking this into account insufficiently is the root of the difficulties, with which the electrodynamics of moving bodies currently has to contend.[19]

Little wonder that Einstein was "tormented" by the problem of "just what coordinates are actually supposed to mean in physics" once they lose their direct physical significance!

This problem arose in the course of the application of the equivalence principle to linearly accelerated frames of reference and the attempt to apply it to uniformly rotating frames, both considered within the confines of Minkowski space.[20] Its resolution came out of Einstein's work on a theory of the static gravitational field, in particular on the equations of motion of a particle in this field; and his attempt to generalize this static theory to non-static fields.

Both problem areas, accelerated systems of reference in Minkowski space and static gravitational fields, ultimately led Einstein beyond the confines of Minkowski space to the consideration of non-flat Riemannian spacetimes. For convenience of exposition, I shall discuss these two strands of the story as if they were the subject of two separate scenes of Act Two, culminating in a third scene that ends the act. While it is broadly true that events in Scene One precede those in Scene Two, and certainly true that they all precede the events in Scene Three, to the extent that this division suggests a strict chronological separation between events in the First and Second Scenes, it does a certain violence to the actual course of events. However, it seems preferable to run this risk rather than attempt to jump back and forth between events in each of the intertwined strands of the story.[21]

---

18 For a historical discussion of Einstein's hole argument, see my 1980 Jena paper, published as (Stachel 1989); and reprinted in (Stachel 2002, 301–337).

19 "Zur Elektrodynamik bewegter Körper," (Einstein 1905); reprinted in (CPAE 2, 276–306), citation from p. 277. For further discussion of this paper see the Editorial Note "Einstein on the Theory of Relativity," in (CPAE 2, 253–274); reprinted in (Stachel 2002, 233–244)

20 For a discussion of the development of Einstein's concept of the equivalence principle, see (Norton 1985); reprinted in (Howard and Stachel 1989, 5–47).

21 Since contemporary documents are cited with dates, the chronological sequence can easily be reconstructed.

SCENE I: "TO INTERPRET ROTATION AS REST":

I shall start by again citing Einstein's 1949 comments on coordinates:

> The change [from the viewpoint that coordinates must have an immediate metrical significance] came about in more-or-less the following way.
>
> We start with an empty, field-free space as it appears with respect to an inertial system in accord with the special theory of relativity, as the simplest of all conceivable physical situations. Now if we imagine a non-inertial system introduced in such a way that the new system (described in three-dimensional language) is uniformly accelerated in a (suitably defined) direction with respect to the inertial system; then, with respect to this system, there exists a static parallel gravitational field. In this case, the reference system may be chosen as a rigid one, in which three-dimensional Euclidean metric relations hold. But that time [coordinate–JS], in which the field appears static, is *not* measured by *equally constituted* clocks at rest [in that system–JS]. From this special example, one already recognizes that, when one allows non-linear transformations of any sort, the immediate metrical significance of the coordinates is lost. One *must* introduce such transformations, however, if one wants to justify the equality of gravitational and inertial mass by the foundations of the theory, and if one wants to overcome Mach's paradox concerning inertial systems.[22]

Examination of Einstein's 1907 paper[23] shows that this account correctly reflects its contents. Einstein first demonstrates that—at least to first order in the acceleration—the spatial coordinates in a uniformly accelerating frame of reference retain their direct physical significance in terms of measuring rods; and thus, by the principle of equivalence, they still do so in the equivalent gravitational field. He then goes on to show that what he calls "the local time σ" [he uses both "*Ortszeit*" and "*Lokalzeit*" as names], which is essentially the proper time as measured by an ideal clock at a fixed point of the frame, differs from the "time τ," which he later called the "universal" [*universelle*] time,[24] which must be used to define simultaneity of distant events if one wants a time coordinate expressing the static nature of the gravitational field that is equivalent to the uniformly-accelerated one.

Thus, by the end of 1907, Einstein knew that differences between the "universal" time coordinates of events in a uniform gravitational field do not correspond to differences in the readings of ideal clocks in that field. It is true that he had shown that, at least to first order in the field strength, spatial coordinate differences still correspond to the results of measurements with rigid rods. But the fact that he felt compelled to demonstrate this for uniform gravitational fields suggests that he anticipated the pos-

---

22 "Autobiographical Notes," (Einstein 1979, 62 and 64); see note 16. An idea of what he meant by "Mach's paradox concerning inertial systems" may be gathered from the citations of Mach in Einstein's article, "Ernst Mach," (Einstein 1916b). See also (Einstein 1916a); reprinted in (CPAE 6, 284–339, Section 2), "Über die Gründe, welche eine Erweiterung des Relativitätspostulates naheliegen," pp. 286–288.

23 "Über das Relativitätsprinzip," (CPAE 2, Section 18, 476–480); for the full references, see note 11.

24 Einstein did not actually introduce this term until 1912, in his first paper on the static gravitational field, in which he contrasts the "local time" and the "universal time" (for the full reference, see note 36).

sibility that similar problems might arise for the spatial coordinates in more complicated gravitational fields.

Einstein did not publish anything on gravitation between 1908 and 1911, but he continued to think about the subject:

> Between 1909–1912 while I had to teach theoretical physics at the Zurich and Prague Universities I pondered ceaselessly on the problem.[25]

The earliest surviving indication that Einstein contemplated an extension of the relativity principle beyond linearly accelerated systems dates from 1909:

> The treatment of the uniformly rotating rigid body seems to me to be of great importance on account of an extension of the relativity principle to uniformly rotating systems along lines of thought analogous to those that I attempted to carry out for uniformly accelerated translation in the last section of my paper published in the *Zeitschrift für Radioaktivität*.[26]

What he had in mind is made more explicit in 1912 in a letter to his friend Michele Besso. After a rather full account of his new static theory (to be discussed below), he concludes: "You see that I am still far from being able to interpret [*auffassen*] rotation as rest. Every step is devilishly difficult..."[27]

As his reference to a "uniformly rotating rigid body" suggests, a solution to the problem of "interpreting rotation as rest" seemed to him to depend on developing a theory of rigid bodies in special relativity. In 1910 he wrote of this

> child of sorrow [*Schmerzenskind*], the rigid body. ... one should attempt to devise hypotheses about the behavior of rigid bodies that would permit a uniform rotation.[28]

Born had provided a definition of a relativistic rigid body in 1909, but he only discussed the case of linearly accelerated motion in any detail.[29] Further clarification soon came:

> The latest relativity-theoretical investigations of Born and Herglotz interest me very much. It really seems that in the theory of relativity there does not exist a "rigid" body with 6 degrees of freedom.[30]

This was disturbing, but brought new hope: If rigid *bodies* are incompatible with the special theory, rigid *motions* are not. In 1911, Laue summarized the situation concisely:

---

25 "Autobiographische Skizze," (Seelig 1955, 14).

26 Einstein to Arnold Sommerfeld, 29 September 1909, (CPAE 5, 210). Einstein incorrectly names the title of the journal in which his earlier paper was published (see Einstein 1907). Einstein had described the main theme of the last section of this paper in an earlier letter to Sommerfeld, Einstein to Arnold Sommerfeld, 5 January 1908, (CPAE 5, 86).

27 Einstein to Michele Besso, 26 March 1912,(CPAE 5, 435–438); citation from p. 436.

28 Einstein to Arnold Sommerfeld,19 January 1910, (CPAE 5, 228–230); citation from p. 229.

29 See (Born 1909a). It was his report on this work at the 1909 Salzburg meeting of the *Versammlung deutscher Naturforscher und Ärzte*, (Born 1909b), that provoked the above-cited letter of 1909 from Einstein to Sommerfeld. See also (Born 1910), discussed below.

30 Einstein to Jakob Laub, 16 March 1910, (CPAE 5, 231–233); citation from p. 232.

> The limiting concept of a body that is rigid under all circumstances, which is so useful everywhere in classical mechanics, in my opinion cannot be taken over [to the special theory–JS] on account of the impossibility of indefinitely large velocities for the propagation of elastic deformations. However this does not exclude a body moving at times like a rigid one; even according to classical mechanics, under certain circumstances a drop of fluid can move as if it were rigid. (Laue 1911, note at bottom of p. 107)

In short, the problem had been transformed from a dynamical one (what is a relativistic "rigid body" and how does it behave when accelerated?) to a kinematical one (given Born's relativistic definition, what types of "rigid motion" are possible?). Progress on the kinematical problem was much easier.[31] Indeed, in the paper cited by Einstein, Herglotz showed:

> that, as soon as one of the points of a [rigid–JS] body in Mr. Born's sense is fixed, it can only rotate uniformly about an axis passing through this point, like the usual rigid body. (Herglotz 1910, 403)

In the course of classifying all solutions of Born's rigidity condition, Herglotz gave the explicit form of the solution for rigid rotation about the z-axis (Herglotz 1910, 412). So Einstein could discuss the kinematics of such an ideal rigid rotation, and the gravitational field which is equivalent to the inertial forces in such a rotating frame, without having to solve the dynamical problem of what types of physical system could actually undergo such a motion.

But for Einstein, there remained a second, Machian type of question: What distribution of matter could *induce* the gravitational field in a frame at rest that is equivalent to the inertial forces in an accelerated frame? Einstein first considered this question for linear acceleration, so we shall discuss it before returning to the problem of rotation. In a 1912 paper on gravitational induction,[32] Einstein showed that, as a consequence of the inertia of energy and the equivalence principle, a spherical shell of matter K accelerated linearly relative to an unaccelerated frame exerts such an inductive accelerating (gravitational) effect on a particle P enclosed in the shell.[33] He also showed that:

---

31 Of course, a dynamical problem remained: how to create the circumstances that would lead a non-rigid body of a particular constitution to execute a particular rigid motion. But such special dynamical problems could be attacked *after* the general kinematical problem was solved.

32 "Gibt es eine Gravitationswirkung, die der elektrodynamischen Induktionswirkung analog ist?"(Einstein 1912a); reprinted in (CPAE 4, 175–179). Einstein published the article in this journal because it was part of a *Festschrift* for his friend, Heinrich Zangger, an expert in forensic medicine.

33 Presumably, this would be Einstein's answer to Laue's objection to the equivalence principle: "For the gravitational field in the system K [at rest, with a uniform gravitational field] there must be present a body that causes gravitation, not however for the accelerated system K'. So a search for it must immediately decide whether there is a real gravitational field or only an accelerated reference system," Max Laue to Albert Einstein, 27 December 1911, (CPAE 5, 384). This letter is discussed further below. Einstein evidently tried to answer Laue's objection in a footnote to his next paper, submitted two months later: "The masses that produce this field must be thought of as at infinity" (Einstein 1912b, 356). The paper on gravitational induction followed almost immediately.

> the presence of the massive shell K increases the inertial mass of the particle P within it. This makes likely the assumption that the entire inertia of a massive particle is an effect of the presence of all the other masses, based on a sort of interaction with the latter. This is completely the same standpoint that E. Mach had upheld in his acute investigations on the subject (*E. Mach, The Development of the Principles of Dynamics. Chapter Two. Newton's Views on Time, Space and Motion*).[34] How far this conception is justified will be seen when we are in the happy possession of a usable dynamics of gravitation (p. 177).

Presumably, Einstein already had in mind the application of this induction idea to rotational acceleration. In 1921, discussing the development of the general theory, Einstein wrote:[35]

> Can gravitation and inertia be identical [*wesensgleich*]? The posing of this question leads directly to the General Theory of Relativity. Is it not possible for me to regard the earth as free from rotation, if I conceive of the centrifugal force, which acts on all bodies at rest relative to the earth, as being a "real" field of gravitation (or part of such a field)? If this idea can be carried out, then we shall have proved in very truth the identity of gravitation and inertia. For the same effect [*Wirkung*] that is regarded as *inertia* from the point of view of the system not taking part in the rotation can be interpreted as *gravitation* when considered with respect to the system that shares the rotation.

I believe that the phrase "or part of such a field" makes clear what Einstein had in mind as his ultimate goal. The total gravitational field of the earth in a frame in which it is at rest (i.e., a co-rotating frame) consists of two parts: a gravito-static term, which would be present even if the earth were not in rotation (this is the Newtonian gravitational field), and a gravito-stationary term. The latter is usually interpreted as an inertial field, consisting of centrifugal and Coriolis terms, which would exist even if the earth were massless, i.e., in *any* rotating frame of reference. But, in accord with the principle of equivalence, these terms may be interpreted as a gravito-stationary field in a non-rotating frame of reference.

In summary: Because of his attraction to Mach's program from the beginning of his search for a theory of gravitation based on the equivalence principle, the aim of interpreting rotation as rest-plus-a-gravitational-field appears to have loomed large in Einstein's motivation. This motive led him to consider uniformly rotating systems of reference soon after his 1907 treatment of uniformly linearly accelerated systems. But only after the clarification of the question of rigid motions do we find any signs of progress on the rotation problem.

The study of uniformly rotating reference systems then led him to the conclusion that, in this case, the spatial coordinates cannot be given a direct physical meaning. He announced this result in February 1912, in an uncharacteristically tentative tone, in the course of a discussion of the spatial coordinates in a linearly accelerated frame of reference K:[36]

---

34 Exactly the same words about Mach's book occur in "Einstein's Scratch Notebook," reproduced with transcription in (CPAE 3, "Appendix A," 564–596; see p. 592).

35 "A Brief Outline of the Development of the Theory of Relativity, (Einstein 1921, 783). A German draft, "Kurze Skizze zur Entwicklung der Relativitätstheorie," has been used to correct the English text. Both appear in (CPAE 7).

> The spatial measurement of K is done with measuring rods that—when compared with each other at rest at the same place in K— possess the same length; the theorems of [Euclidean–JS] geometry are assumed to hold for lengths measured in this way, and thus also for the relations between the coordinates x, y, z and other lengths. That this stipulation is allowed is not obvious; rather it contains physical assumptions that eventually could prove incorrect. For example, it is highly probable that they do not hold in a uniformly rotating system, in which, on account of the Lorentz contraction, the ratio of the circumference to the diameter, using our definition of lengths, must be different from π.

There is evidence suggesting that he had this rotating disk argument before 1912, but it is indirect and suggestive rather than conclusive. Einstein's letter of 1909 to Sommerfeld, cited above,[37] was written just a few days after the Salzburg meeting of the Society of German Natural Scientists and Physicians [*Deutsche Naturforscher und Ärzte*], at which Einstein had spoken. As the editors of the Einstein Papers note:[38]

> At the Salzburg meeting Max Born had presented a paper on rigid body motion in special relativity ..., on which Sommerfeld had commented in the discussion following the paper. Einstein and Born had discussed the subject and had discovered that setting a rigid disk into rotation would give rise to a paradox: the rim becomes Lorentz-contracted, whereas the radius remains invariant (see Born 1910, p. 233).[39] The existence of this paradox was first pointed out in print by Paul Ehrenfest (1880–1933) in a paper that was received on the date of this letter.[40]

While the line of argument about setting a rigid disk into rotation (which has come to be called "Ehrenfest's Paradox") is not the same as that in Einstein's treatment of an already rigidly-rotating disk,[41] the basic idea in both arguments is the same: Relative to an inertial frame, measuring rods at rest in a uniformly rotating frame of reference *do not* contract if aligned in a radial direction, but *do* contract if aligned orthogonally to a radial direction.

So it is reasonable to suppose that Einstein, already alerted to the possibility that coordinate differences in an accelerating frame might not be directly interpretable in terms of physical measurements and having read Herglotz's 1910 paper, realized that this was indeed the case for the spatial coordinates in a rigidly rotating frame of reference.[42]

Indeed there is evidence that, by the end of 1911 at the latest, Einstein saw an analogy between the gravitational field that, according to the equivalence principle (conceiving rotation as rest), is equivalent to the inertial field in a uniformly rotating

---

36 "Lichtgeschwindigkeit und Statik des Gravitationsfeldes," (Einstein 1912b); reprinted in (CPAE 4, 130–145); citation from p. 131. (This is his first paper on the static gravitational field, discussed at greater length below.) For a translation of a longer portion of this passage and a fuller discussion of the rotating disk problem, see (Stachel 1980); reprinted in (Howard and Stachel 1989, 48–62) and in (Stachel 2002, 245–260).

37 See note 28.

38 See (CPAE 5, 211, n. [5]).

39 The reference reads: "Mr. P. Ehrenfest ... showed in a very simple way that a body at rest can never be brought into uniform rotation; I had already discussed the same fact with Mr. A. Einstein in Salzburg."

40 "Gleichförmige Rotation starrer Körper und Relativitätstheorie," (Ehrenfest 1909).

41 For Einstein's way of avoiding Ehrenfest's paradox, see (Stachel 1980, 6–7 and 9).

frame of reference and the magnetostatic (or electro-stationary) field due to a stationary circular current distribution.[43] In the letter cited earlier,[44] Max Laue alludes to:

> your [i.e., Einstein's] question whether the gravitational field strength should be represented by a four-vector or a six-vector.

We shall return to this letter at some length below, but for the moment, consider the implications of Einstein's question. The most natural generalization of the Newtonian gravitational field strength would be a four-vector, since the force exerted by the Newtonian field depends only on the position and not the velocity of a mass in that field. On the other hand, the electric and magnetic field strengths together constitute a six-vector, and the force exerted on a charge in an electromagnetic field depends on the position (electric force) and velocity (magnetic force) of the charge. Around the turn of the century, H. A. Lorentz had suggested a gravitational theory modeled on electromagnetism, in which there were gravitational analogues of the electric and magnetic forces (Lorentz 1899–1900a).[45]

Thus, Einstein's question to Laue suggests that, by the end of 1911, he had reason to believe that the force exerted by the most general gravitational field might also be velocity dependent. In his 1911 paper, he had considered the gravitational analogue of a constant electrostatic field, and he was soon to consider the analogue of the general electrostatic field.[46] By 18 February 1912, Einstein was already communicating some of his results.[47]

There is also evidence that, by February 1912, Einstein was already considering the gravitational analogue of a magnetostatic field. Paul Ehrenfest visited Einstein in Prague during the last week of February, and Ehrenfest's diary entry for 24 February 1912 contains the following lines:

---

42 Curiously, neither he nor any other contemporary ever refers to a 1910 paper by Theodor Kaluza (of later five-dimensional Kaluza-Klein theory fame) solving the problem of the "proper geometry" [*Eigengeometrie*] of a Born rigidly-rotating body (Kaluza 1910). Kaluza was prevented by illness from presenting his work at the 1910 Königsberg meeting of the *Deutsche Naturforscher und Ärzte* which may help to explain its lack of impact on the rigid body discussion.

43 However, there is also an important difference between the two: The gravitational field equivalent to the inertial forces in an accelerated reference frame does not appear to correspond to any material sources, while the analogous electromagnetic fields are produced by a charged ring—rotating or not. As noted above, this was the purport of Max Laue's criticism of Einstein's treatment of the gravitational field equivalent to the inertial forces in a uniformly accelerated frame of reference (see note 33), to which Einstein hoped to provide a Machian answer.

44 Max Laue to Albert Einstein, 22 December 1911, (CPAE 5, 384–385); citation from p. 385.

45 English translation "Considerations on Gravitation," in (Lorentz 1899–1900b).

46 This was probably at least in part in response to Abraham's work on the problem of gravitation, which appeared early in 1912. Abraham's first two papers on the subject, dated "December 1911," were received 14 December and published in the issue of 1 January 1912 of *Physikalische Zeitschrift* (Abraham 1912a; Abraham 1912b). There is evidence that Einstein had corresponded with Abraham about his theory before publication (see below).

47 See Einstein to Hendrik Antoon Lorentz, (CPAE 5, 411–413); reference to gravitation on p. 413.

> Einstein told me about his gravitational work. [I omit some equations referring to the static case] Centrifuging of radiation.[48]

A subsequent letter from Ehrenfest makes clear the meaning of the final, rather cryptic phrase. He reports that a Russian colleague, Michael Frank, has "put me in a very uncomfortable situation" by asking Ehrenfest to translate into German a work concerning "the geometry of light rays in a uniformly rotating laboratory." After describing Frank's work, Ehrenfest adds:

> If one wanted to transform the acceleration field of uniform rotation into a corresponding force field at rest, as you do in your paper 'On the Influence of Gravitation ...' for uniform linear acceleration, then this substitute force-field would also have to give the proper Coriolis deflection for light rays.—That is the content of [Frank's] note. The thing is embarrassing [*peinlich*] for me since you had already communicated this argument to me. ... I told him that you had already told me about this (I remembered it naturally just at the moment when "Coriolis" was recognizable.)[49]

So "centrifuging of radiation" refers to "the geometry of light rays in a uniformly rotating laboratory," a problem on which Einstein had evidently worked before 24 February 1912 when he presented his results to Ehrenfest. Einstein's mention of "Coriolis" indicates that he had in mind a velocity-dependent gravitational force. It was presumably the publication of Frank's paper[50] that decided Einstein against publishing his own version of the results on light rays. He wrote Ehrenfest:

> Translate that work [of Frank-JS] in tranquility. I do not arrogate to myself any relativity-monopoly! Everything that is good is also welcome. You needn't send the proofs.[51]

So by February 1912, as seems probable on the basis of his own writings; and surely by April, on the basis of Ehrenfest's letter, Einstein was aware that the gravitational field equivalent to a rotating frame of reference would have to exert a force on a light ray that depends not only on its position but on (at least the direction of) its velocity—something that is incompatible with a scalar theory of gravitation. So, even while writing his first paper on the static gravitational field, he was aware that a scalar theory was not possible for more general gravitational fields. I shall return to this point below.

Ehrenfest continued to work on the kinematic aspects of the gravito-stationary problem. In a postcard, he writes that he has solved the "Problem: To determine the *most general* field of world-lines that is equivalent to a *stationary* gravitational field." He mentions two "special cases": "hyperbolic motion" (i.e., constant linear acceleration) and "uniform rotation."[52] Subsequent letters outline the proof his solution.[53] In

---

48 Rijksmuseum voor de Geschiednis der Natuurwetenschappen, Leiden: Ehrenfest Collection, Ehrenfest Notebook 4–11, Microfilm number 12.

49 Ehrenfest to Einstein, draft letter before 3 April 1912, (CPAE 5, 439–445), citation from p. 440.

50 "Bemerkung betreffs der Lichtausbreitung in Kraftfeldern," (Frank 1912). The paper is dated 7 March 1912. Frank does not suggest that the force field equivalent to a force-free rotating frame of reference could be gravitational.

51 Einstein to Ehrenfest, 25 April 1912, (CPAE 5, 450–451); citation from p. 450.

reply to one of these letters, Einstein says that he does not understand Ehrenfest's result,[54] but adds some comments on the problem:

> A rotating ring does not generate a static field in this sense [the sense of Papers I and II, see note 70–JS], although it is a time-independent field. In such a field the reversibility of light paths does not hold.[55] My case corresponds to the electrostatic field in electromagnetic theory, while the more general static case would also include the analogue of the static magnetic field. I am not yet that far along. The equations I have found only relate to the static case of masses at rest.[56]

I have now given reasons for believing that, at the earliest by 1909 and the latest by the end of 1911, Einstein was aware of problems with the interpretation of both temporal and spatial coordinates in accelerating frame of reference in Minkowski space; and that, by February 1912 at the latest, he had every reason to expect that the full theory of gravitation would have to pass beyond the limits of a four-dimensional scalar theory, on the one hand; and, at least spatially, beyond the limits of Euclidean geometry. Now I shall turn to Einstein's investigation of static gravitational fields, which ultimately led to the resolution of these problems.

SCENE II: "THE SPEED OF LIGHT IS NO LONGER CONSTANT"

It may well have been Max Laue who directed Einstein's attention to the crucial importance of the gravitational potential, and the possibility of its replacement by the variable speed of light. The letter Laue sent Einstein at the end of 1911 was cited above, but I must now quote from it at greater length. (Unfortunately we do not have Einstein's letter, if there was one, to which this is a reply.)[57] Discussing Einstein's 1911 paper,[58] he writes:

---

52 Ehrenfest to Einstein, 14 May 1912, (CPAE 5, 460–461). In later chapters, these two "special cases" will become very familiar to the reader since they are the two test cases that Einstein uses again and again to evaluate candidate gravitational field equations.

53 Only Ehrenfest's drafts of his letters have been preserved: Ehrenfest to Einstein: after 16 May 1912, (CPAE 5, 461–464, see 462–463); 29 June 1912, (CPAE 5, 487–496). Einstein to Ehrenfest, 25 April 1912, (CPAE 5, 450–451, see 451); 27 April 1912, (CPAE 5, 455); before 20 June 1912, (CPAE 5, 484–486, see 485–486). (Only letters containing references to gravitation are cited.) Ehrenfest later published a paper on this subject: "On Einstein's Theory of the Stationary Gravitation Field," (Ehrenfest 1913a); original version, (Ehrenfest 1913b).

54 Einstein to Ehrenfest, before 20 June 1912, (CPAE 5, 484–486, see p. 485).

55 This assertion is analogous to the result discussed above that Frank had published, but with a subtle difference. Like Einstein, Frank had discussed Minkowski spacetime as seen from a uniformly rotating frame of reference and the field (he does not specify it as gravitational) equivalent to the inertial forces present in such a frame. Here, Einstein is discussing the gravitational field generated by a rotating material ring, which he must have realized would be non-Minkowskian since this is true even for the field of a non-rotating material ring.

56 Einstein to Ehrenfest, before 20 June 1912, (CPAE 5, 484–486, see p. 486).

57 Since Einstein does not raise the question of whether the gravitational field strength is a four-vector or a six vector in his 1911 paper, I assume that the phrase "Your question" in Laue's letter refers to either a previous letter or conversation.

> It seems extraordinarily characteristic to me that the gravitational potential thereby acquires a physical significance, which is completely lacking for the electrostatic potential. One could, in principle, immediately determine the former by measurement of the velocity of light.

This comment may well have been the cue that prompted Einstein's replacement of the gravitational potential by the variable speed of light in his gravito-static theory.

But, as indicated in the last section, the letter contains more significant clues to the direction in which Einstein was heading. Laue continues:

> Your question, whether the gravitational field strength should be represented by a four-vector or a six-vector, is thereby settled. Not it [the field strength–JS] but rather the potential accordingly seems to me to be the primary concept, the four-dimensional representation of which must be investigated.[59]

A little background is helpful in assessing the full significance of this comment of Laue's. After initially slighting the significance of Minkowski's four-dimensional reformulation of the special theory, Einstein had started to study it in earnest around 1910,[60] probably at least in part in response to Sommerfeld's exposition of a four-dimensional vector algebra and analysis (Sommerfeld 1910a; Sommerfeld 1910b),[61] and its incorporation and further development in Laue's textbook—the first on special relativity.[62]

A further motive for this study was Einstein's decision to include the four-dimensional approach in a major review article that he agreed to write in 1911, and started work on by 1912.[63] In this review he notes that an anti-symmetric second rank tensor "is, following Sommerfeld, usually designated as a six-vector" [*Sechservektor*],[64]

58 See (Einstein 1911b).

59 Max Laue to Albert Einstein, 27 December 1911, (CPAE 5, 384).

60 In 1908, Einstein and Laub thought it worthwhile to publish a paper rederiving Minkowski's four-dimensional results on electrodynamics (see Minkowski 1908) in three-dimensional form because that "work makes rather great demands mathematically on the reader," see (Einstein and Laub 1908). In a review talk on special relativity given in January 1911, Einstein included a brief discussion of "the highly interesting mathematical development that the theory has undergone, primarily due to Minkowski who unfortunately died so young," noting that it had led to "a very perspicacious representation of the theory, which essentially simplifies its application" see (Einstein 1911a). For a discussion of Minkowski's work and the varying forms of its assimilation by the physics and mathematics communities, see (Walter 1999).

61 He states that the formalism he presents "is (aside from imaginary coordinates) an immediate generalization of the customary three-dimensional vector methods", and provides "a complete substitute for the matrix calculus used by Minkowski" (Sommerfeld 1910a, 749).

62 Laue notes that he has "taken into account extensively the mathematical development of the theory that Sommerfeld has recently given" (Laue 1911, vi). Einstein commented: "His book on relativity theory is a little masterpiece," Einstein to Alfred Kleiner, 3 April 1912, (CPAE 5, 445–446).

63 This article, prepared for Erich Marx's *Handbuch der Radiologie*, was completed but has been published only recently: See "Manuscript on Relativity," (CPAE 4, 9–108). For its history, see the Editorial Note, "Einstein's Manuscript on the Special Theory of Relativity," (CPAE 4, 3–8).

64 See (CPAE 4, 72). Sommerfeld had given the names "four-vector" and "six-vector" to what Hermann Minkowski had called "spacetime vectors of type I and II," respectively (Minkowski 1908, 65–68).

and notes that the electromagnetic field strength, the components of which are the electric (see p. 9) and magnetic (see p. 10) field strengths, is a six-vector (see p. 81).

So it is reasonable to assume that Einstein had already mastered the four-dimensional formalism by the time that he raised the question of whether the gravitational field strength is a six-vector (as in the electromagnetic case) or a four-vector (as would be the case of it were the gradient of a scalar field). Laue's reply assumes knowledge of the fact that the electromagnetic six-vector is the curl of the electromagnetic potential four-vector;[65] which breaks up into the electric potential (a three-scalar) and the magnetic (three-)vector potential with respect to any inertial frame. If we look only at the electrostatic field strength, it can be written as the three-gradient of the electric potential. But, Laue points out, the electrostatic potential is of no physical significance (because of the possibility of what are now called gauge transformations of the electromagnetic potentials);[66] while Einstein's 1911 work showed that the gravitational potential has an immediate physical significance because of its influence on the speed of light. Therefore, Laue suggests, the important question is: What is the four-dimensional representation of the gravitational *potential*? The fact that, in the static case, it reduces to a single quantity that behaves as a scalar under three-dimensional spatial transformations is *not* decisive for answering this question. The same is true of the electrostatic potential; yet the latter is known to be the fourth (i.e., timelike) component of a four-vector.

Thus, Laue's comment could have served to draw Einstein's attention away from the representation of the gravitational field strength, and toward the question that, within a few months, was to occupy him: What is the four-dimensional representation of the gravitational potential in the non-static case?

Laue's comment, like all of his and Sommerfeld's work on the four-dimensional formalism, is situated within the context of the special theory of relativity. But, as we have seen, with Einstein's interpretation of the equivalence principle as implying an enlargement of the relativity group, Einstein had already moved beyond that context; and he soon moved into the context of non-flat spacetimes. Indeed, there is a comment by Einstein himself dating from mid-1912, on the question of the four-versus six-vector representation of the gravitational field strength, that suggests the need for this shift of context:

> If the gravitational field can be interpreted within our present [i.e., special–JS] theory of relativity [*sich ... im Sinne unserer heutigen Relativitätstheorie deuten läßt*], then this can only happen in two ways. One can consider [*auffassen*] the gravitational vector either as a four-vector or as a six-vector. [In either case] one arrives at results that contradict the ... consequences of the law of the gravitational mass of energy, [namely] ... that gravitation acts more strongly on a moving body than on the same body in case it is at rest. ... It must

65 See Laue's book, *Das Relativitätsprinzip*, (Laue 1911, 99–100), which defines the "four-potential vector" [*Viererpotential*] as the four-vector, the four-curl of which is the electromagnetic field six-vector; and notes that the four-potential vector is only determined up to the four-gradient of a scalar.

66 This was, of course, long before discussions of the physical significance of the electromagnetic potentials, based on the Aharonov-Bohm effect, took place.

> be a task of the immediate future to create a relativistic-theoretical schema in which the equivalence of gravitational and inertial mass finds expression.[67]

In a similar vein, he wrote to Wien that:

> the [special–JS] relativity theory imperatively demands a further development since the gravitational vector cannot be fitted into the relativity theory with constant c if one demands the *gravitational* mass of energy[68]

As it turned out, such a theory would not only pass beyond the bounds of the special theory; it would involve spacetimes with non-flat line elements, as we shall soon see.

To summarize: There are good reasons to suggest that, by the beginning of 1912, Einstein already realized that he would ultimately have to go beyond a scalar theory of gravitation. His strategy was to proceed in a step-by-step fashion towards a full dynamical theory. The first step in the program was to consider what I have called above the gravito-static case, the gravitational analogue of electrostatics; but he was already thinking about the next step, the gravito-stationary case, the gravitational analogue of magnetostatics. His ultimate goal was to develop a theory for time-dependent gravitational fields.

Let us look at the first step, gravito-statics. By March 1912 he was able to write Paul Ehrenfest:

> The investigations of gravitational statics (point mechanics electromagnetism gravito-statics) are complete and satisfy me very much. I really believe that I have found a part of the truth. Now I am considering the dynamical case, again also proceeding from the more special to the more general [case–JS].[69]

Einstein was referring to his two papers on the static gravitational field (hereafter cited as "Paper I" and "Paper II"), completed in February and March 1912 respectively.[70] These papers center on the gravitational potential, as Laue had suggested, but effect a crucial transformation of the problem in line with Laue's comment that the gravitational potential could "in principle be determined by measurement of the speed of light." In his 1911 paper, Einstein had already shown that, with a certain definition of the universal time:[71]

> in a static gravitational field a relation between c [the speed of light–JS] and the gravitational potential exists, or in other words, that the field is determined by c. (Einstein 1912b, 360)

In Papers I and II $c(x,y,z)$, the spatially variable but temporally constant and direction-independent speed of light, completely replaces the gravitational potential.

---

67 "Relativität und Gravitation. Erwiderung auf eine Bemerkung von M. Abraham," (Einstein 1912c). The paper was received on 4 July 1912. The citations are from pp. 1062–1063.

68 Einstein to Wilhelm Wien, 17 May 1912, (CPAE 5, 465).

69 Einstein to Paul Ehrenfest, 10 March 1912, (CPAE 5, 428).

70 Paper I:"Lichtgeschwindigkeit und Statik des Gravitationsfeldes," (Einstein 1912b); reprinted in (CPAE 4, 130–145); Paper II: "Zur Theorie des statischen Gravitationsfeldes," (Einstein 1912d); reprinted in (CPAE 4, 147–164). For a discussion of these papers, see the Editorial Note "Einstein on Gravitation and Relativity: The Static Field," (CPAE 4, 122–128).

Most of Paper I is concerned with establishing the gravitational field equation that c obeys, and the equations of motion of a (test) particle in the static gravitational field described by c. Much of Paper II is concerned with a revision of the field equation of Paper I (Section 4), and a crucial mathematical reformulation of the equations of motion (The "Supplement" to the Proofs).[72]

Einstein's introduction of a variable speed of light brought down much scorn upon him at the time,[73] but it was absolutely crucial in initiating the sequence of steps that lead to the culmination of Act II: Einstein's leap from a scalar to a tensorial gravitational potential, in which $c(x,y,z)$ becomes one of the ten components of the metric tensor used to construct the line element of a non-flat spacetime.

Paper I also contains another step in the process. Einstein shows that, if one uses a light clock, for example, to measure:

> the local time, which Abraham denotes by l, then this stands to the universal time $[t]$ in the relation $dl = c\ dt$. (Einstein 1912b, 366)

In retrospect (remembering that here c is non-constant), we recognize in this equation the relation between the differential element of the proper time dl between two events at the same place (i.e., $x,y,z = \text{const}$) in a static field, and the coordinate differential dt between the times of the two events, using the preferred static time coordinate t. This equation begins to answer the question of the relation between coordinates and physical measurements in a gravitational field that had been puzzling Einstein for almost five years.

But before expanding on this point, let me turn to a further step in the process, contained in Paper II. The equations of motion of a particle in a static gravitational field, developed in Paper I, were rewritten in Lagrangian form in a "Supplement to

---

71 "The time in the field [that is] defined by the stipulation that the speed of light c depends indeed upon the position but not on direction" as Einstein explained to Michele Besso, 26 March 1912, (CPAE 5, 435). This definition of the time is given in more detail in Paper I: "We think of the time in the [uniformly accelerated–JS] system K as measured by clocks of such a nature and such a fixed arrangement at the spatial points of K that the time intervals—measured with them—that a light ray takes to go from a point A to a point B of the system K does not depend on the moment of emission of the light ray at A [static condition]. Further it turns out that simultaneity can be defined without contradiction in such a way that, with respect to the settings of the clocks, the stipulation is satisfied that all light rays passing a point A of K have the same speed of propagation, independent of their direction" [isostropy condition] (Einstein 1912b, 357–358).

72 The revision of the gravitational field equation is discussed in some detail in "Pathways out of Classical Physics …" (in this volume). The remainder of Paper II is concerned with electromagnetism (Sections 1 and 2) and thermodynamics (Section 3) in a static gravitational field, topics I shall not consider.

73 Abraham, for example, exulted: "Einstein ... had already given up his postulate of constancy of the velocity of light at the turn of the year, which was so essential for his earlier theory; in a recent work he abandons the requirement of the invariance of the equations of motion under Lorentz transformations, thereby delivering the *coup de grace* to the theory of relativity. Those who, like the author, have repeatedly had to warn against the siren song of this theory, can only greet with satisfaction the fact that its originator has now convinced himself of its untenability" (Abraham 1912e).

the Proofs" of Part II. This step proved to be so significant that it soon led to the final resolution of the problem of the correct representation of the gravitational potentials.

> It is noteworthy that the equations of motion of a particle [*materielle Punkt*] in the gravitational field take a very simple form when they are given the form of Lagrange's equations. Namely, if one takes
>
> $$H = -m\sqrt{c^2 - q^2},$$
>
> then ... For a particle moving in a static gravitational field without the action of external forces, there holds accordingly
>
> $$\delta\left\{\int H\ dt\right\} = 0,$$
>
> or
>
> $$\delta\left\{\int\sqrt{c^2dt^2 - dx^2 - dy^2 - dz^2}\right\} = 0.$$
>
> Here too—as was proved by Planck for the usual [i.e., special] theory of relativity—it is seen that the equations of analytical mechanics possess a significance that extends far beyond Newtonian mechanics. Hamilton's equation as finally written down lets us anticipate [*ahnen*] the structure of the equations of motion of a particle in a dynamical gravitational field. (Einstein 1912d), 458)

Einstein's lecture notes on mechanics, which he had been teaching since 1909,[74] show his familiarity with the use of variational techniques to derive the Lagrangian equations of motion (CPAE 3, 91–95, 116–117). Even more important for present purposes, he stressed the coordinate-invariant nature of the resulting equations of motion:

> The Cartesian coordinates of the particle no longer enter into the [variational–JS] principle. It holds independently of whatever coordinates we use to determine the position of the particles of the system (p. 93).

Now we have in hand all the strands, the interweaving of which finally allowed Einstein to take the great leap forward to a metric theory of gravitation. But let me emphasize that, however much they may help us in retrospect to understand the process, there remains something almost uncanny in how Einstein made the choices that led him so far from the path trodden by other physicists in the search for a relativistic theory of gravitation. He was about to enter an entirely new land, in which the space-time structure becomes a dynamical field.

74 "Lecture Notes for Introductory Course on Mechanics at the University of Zurich, Winter Semester 1909–1910," (CPAE 3, 11–129). See also the Editorial Note, "Einstein's Lecture Notes," (CPAE 3, 3–10, especially Section II, 4–6). Einstein also taught mechanics in Prague during the winter semester of 1911; see "Appendix B: Einstein's Academic Courses," (CPAE 3, 598–600).

SCENE III: "TEN SPACETIME FUNCTIONS":

Just what did the variational reformulation of the equations of motion lead Einstein to "anticipate"? I propose to answer this question with the help of the first three sections of his following paper on the topic, the "*Entwurf*" paper, published early in 1913.[75] After introductory comments discussing the equivalence principle, Section 1 treats "The equations of motion of a particle in a static gravitational field." Except for one small but significant detail, it is just an expanded version of the "Supplement." The detail is notational: he writes ds as an abbreviation for $\sqrt{c^2dt^2 - dx^2 - dy^2 - dz^2}$ for both the case of constant c ("the usual theory of relativity") and the spatially variable but static $c(x,y,z)$.

The significance of this notation does not emerge until Section 2, which treats "Equations for the motion of a particle in an arbitrary gravitational field. Characterization of the latter":

> With the introduction of a spatial variability of the quantity c we have passed beyond the framework of the theory that is now designated as "the theory of relativity"; for the expression designated by ds no longer behaves as an invariant under linear orthogonal transformations of the coordinates. ...
>
> If we introduce a new spacetime system $K'(x', y', z', t')$ by means of an arbitrary substitution.
>
> $$\begin{aligned} x' &= x'(x, \mathrm{y}, z, t) \\ y' &= y'(x, y, z, \mathrm{t}) \\ z' &= z'(x, y, z, t) \\ t' &= t' = (x, y, z, t) \end{aligned}$$
>
> and if the gravitational field in the original system K is static, then under this substitution equation (1) goes over into an equation of the form
>
> $$\delta\left\{\int ds'\right\} = 0,$$
>
> where
>
> $$ds'^2 = g_{11}dx'^2 + g_{22}dy'^2 + \ldots + 2g_{12}dx'dy' + \ldots$$
>
> and the quantities $g_{\mu\nu}$ are functions of x', y', z', t' ...
>
> Thus we arrive at the interpretation that, in the general case, the gravitational field is characterized by ten spacetime functions ... . (Einstein and Grossmann 1913, 6)

75 "Entwurf einer verallgemeinerten Relativitätstheorie und einer Theorie der Gravitation I. Physikalischer Teil von Albert Einstein," (Einstein and Grossmann 1913); reprinted in (CPAE 4, 303–323). For a discussion of this paper, see the Editorial Note "Einstein on Gravitation and Relativity: The Collaboration with Marcel Grossmann," (CPAE 4, 294–301). See also the Editorial Note "Einstein's Research Notes on a Generalized Theory of Relativity,"(CPAE 4, 192–199, especially Section II, 193–195).

Einstein proceeds to derive the equations of motion from the generalized Hamiltonian function (which we would now call the Lagrangian):

$$H = -m\frac{ds}{dt} = -m\sqrt{g_{11}\dot{x}_1^2 + \dots + \dots + 2g_{12}\dot{x}_1\dot{x}_2 + \dots + \dots + 2g_{14}\dot{x}_4 + \dots + g_{44}}$$

[where $\dot{x}_1 = \frac{dx_1}{dt}$, etc. -JS].

He writes the three Lagrange equations for the three spatial coordinates x_i, derives from them expressions for the momentum of the particle and the force exerted on it by the gravitational field; and then derives the energy of the particle by performing the usual Legendre transformation on H. He closes this discussion by noting that:

> In the usual relativity theory only linear orthogonal substitutions are permitted. It will be shown that we are able to set up equations for the influence of the gravitational field on material processes that behave covariantly under arbitrary substitutions. (Einstein and Grossmann 1913, 7)

Up to this point, there is nothing in Einstein's results that depends on the interpretation of the $g_{\mu\nu}$ as anything more than a generalization of his 1912 results for the static gravitational field to their form in an arbitrary reference frame. (Remember that, for Einstein, a change of coordinates amounts to a change of spacetime reference frame.) In such a frame, the single static gravitational potential c is transformed into ten functions $g_{\mu\nu}$.

It is only in the next section of the paper, "Significance of the Fundamental Tensor $g_{\mu\nu}$ For the Measurement of Space and Time," that he proceeds to the geometrical interpretation of these functions in spacetime. After recalling that the time coordinate had already lost its immediate physical significance in a static gravitational field, he continued:

> In this connection, we remark that ds is to be understood as an invariant measure for the interval [*Abstand*] between two neighboring spacetime points. (Einstein and Grossmann 1913, 8)

Presumably, this is what his results in the "Supplement" suggested to him almost immediately: if the integrand is interpreted as the interval ds between neighboring points, then the variational principle can be interpreted as giving rise to the equation for a geodesic in the resulting non-flat space time. In a later reminiscence (Einstein 1955),[76] Einstein stated:

> The equivalence principle allows us ... to introduce non-linear coordinate transformations in such a [four-dimensional] space [with (pseudo)-Euclidean metric]; that is, non-Cartesian ("curvilinear") coordinates. The pseudo-Euclidean metric then takes the general form:
>
> $$ds^2 = \sum g_{ik}dx_i dx_k$$

76 Cited from those given in (Seelig 1955).

> summed over the indices i and k (from 1–4). These g_{ik} are then functions of the four coordinates, which according to the equivalence principle describe not only the metric but also the "gravitational field." ... This formulation so far applies only to the case of pseudo-Euclidean space. It indicates clearly, however, how to attain the transition to gravitational fields of a more general type. Here too the gravitational field is to be described by a type of metric, that is a symmetric tensor field g_{ik} ... The problem of gravitation was thereby reduced to a purely mathematical one. Do differential equations exist for the g_{ik} that are invariant under non-linear coordinate transformations? Such differential equations and *only* such could be considered as field equations for the gravitational field. The equation of motion of a particle was then given by the equation for a geodesic line.
>
> With this task in mind, I turned to my old student friend Marcel Großmann, who had in the meantime become Professor of Mathematics at the ETH (pp. 14–15).

Einstein here makes a rather precise claim about what he had accomplished before turning to Grossmann upon his move back to Zurich at the end of July 1912. Earlier, in 1923, he had made a similar claim, but with a significant addition—a reference to Gauss:

> I first had the decisive idea of the analogy between the mathematical problems connected with the theory and the Gaussian theory of surfaces in 1912 after my return to Zurich, initially without knowing Riemann's and Ricci's or Levi Civita's investigations.[77]

In the 1955 reminiscence, Einstein noted that Carl Friedrich Geiser's course on differential geometry at the ETH played an important role in his thinking; it was in that course that he learned about Gauss' theory of surfaces, based on analysis of the distance ds between neighboring points on a surface, expressed in terms of an arbitrary coordinate system on the surface, thereafter often called Gaussian coordinates.[78]

The spacetime interval ds Einstein introduced represents a generalization of what was often referred to in differential geometry as the "line element." The term "element" appears to go back to Monge, who speaks of the "elements" [*élémens*] of a curve in space. Coolidge comments: "An élémen is an infinitesimally short chord."[79] Gauss[80] makes the concept of what he calls a "line element" [he uses both "*Linienelement*" and "*Linearelement*"—see (Gauss 1881, 341–347)] connecting a pair of points on a two-dimensional surface central to his theory of surfaces, and

77 In the Preface to the Czech edition of his popular book on relativity. The German text is in (CPAE 6, 535, n. [4]).

78 In (Einstein 1955), Einstein described the lectures as "true masterpieces of pedagogical art, which later helped me very much when wrestling with general relativity" (pp. 10–11). In a letter of 24 April 1930 to Walter Leich, Einstein wrote: "Geiser was dry only in the large lectures, otherwise I owe him the most of all." For an outline of the contents of Geiser's course on "Infinitesimalgeometrie" given in the Winter Semester of 1897/1898 and the Summer Semester of 1898, based on Grossmann's lecture notes, see (CPAE 1, 365–366). Grossmann's notes are preserved in the ETH Bibliothek, Hs 421: 15 & 16.

79 See (Coolidge 1940), "Book III Differential Geometry," Chapters I, II and III, 318–387. The citations are from p. 322.

shows that it may be used to define the intrinsic properties of the surface, such as its curvature. Bianchi-Lukat (Bianchi 1910, 60–61)[81] defines ds as the "element of a curve" [*Bogenelement*], and comments:

> Since the expression for ds given by [$ds^2 = Edu^2 + 2Fdudv + Gdv^2$, the right hand side being the square of Gauss' expression for the line element] holds for any arbitrary curve on the surface, it is designated the *line element* of the surface. Gauss' ideas were generalized to n- dimensional manifolds by Riemann. (Riemann 1868)[82]

The basic ideas of Gauss' theory of surfaces are reproduced in Grossmann's notes on Geiser's course on differential geometry [*Infinitesimalgeometrie*].[83] Reich indicates some of the high points:

> Geiser treats the line element and its special form in different coordinate systems especially intensively,

indicating that he used the notation ds^2 for it.

> The curvature of surfaces and especially the Gaussian measure of curvature, which Geiser derives in Gauss' fashion and with Gauss' notation, is a particular theme of the semester. The result reads: 'if
>
> $$ds^2 = E\ dp^2 + 2F\ dp\ dq + G\ dq^2,$$
>
> then the measure of curvature depends solely on E, F, G and their derivatives.'... A further important point are geodesic lines. After a longish introduction, Geiser goes into the 'differential equation for geodesic lines,' ... It appears important to me that Geiser offers not only the geometrical aspect but also argues invariant-theoretically, as in the case of the metric, the measure of curvature and of geodesic lines. (Reich 1994, 164–165)

Geiser included a derivation of the equation for a geodesic on a surface by variation of the integral $\int ds$, where

$$ds^2 = \sqrt{E\ dp^2 + 2F\ dp\ dq + G\ dq^2},$$

along some curve $q = \psi(p)$ to find its minimum.[84] Comparison of this with Einstein's variational principle for the equations of motion of a particle in a static gravitational field could have suggested the analogy between Gauss' theory of surfaces and Einstein's theory of the static gravitational field.

---

80 "Disquisitiones generales circa superficies curvas," (Gauss 1828); reprinted in (Gauss 1881, 217–258). A valuable notice ["*Anzeige"*] by Gauss appeared in (Gauss 1827); reprinted in (Gauss 1881, 341–347). The basic idea of using the line element to investigate the properties of a surface had already been used in Gauss' 1822 Prize Essay, (Gauss 1825); reprinted in (Gauss 1881, 189–216). For discussions, see (Coolidge 1940, ch. III, sec. 1, 355–359) and (Stäckel 1918, Heft V, 25–142, sec. V, "Die allgemeine Lehre von den krummen Flächen," 104–138).

81 As we shall see below, this work was consulted by Marcel Grossmann.

82 This refers to the posthumous publication of Riemann's *Habilitationsschrift* of 1854.

83 These notes are described in (Reich 1994, 163–166).

84 Grossmann notes for 10 June 1898.

So much for the mathematical literature. In the physics literature, the equivalent proper time element $d\tau$ had been introduced by Minkowski for timelike worldlines.[85] But neither the concept of, nor the notation for, the line element ds or the proper time element $d\tau$ occurs in the works of Sommerfeld or Laue developing special-relativistic vector and tensor analysis (which, as we have seen Einstein studied) until after 1912.[86] Nor does it occur in Einstein's summary of vector and tensor analysis in his unpublished review of the special theory (discussed earlier).[87]

However, the concept and even the term, were beginning to appear in the physics literature in connection with the discussion of rigid bodies.[88] Herglotz seems to have introduced them (Herglotz 1910, 394). After pointing out that Minkowski had introduced the idea of representing the spatial and temporal coordinates "as the four coordinates of a point in a fourfold-extended manifold $R_4(x, y, z, t)$," he goes on:

> Similarly a measure relation [*Maßbestimmung*] is introduced in this R_4, according to which (the velocity of light being set equal to 1) the square of the distance of two infinitely neighboring points is:
>
> $$ds^2 = dx^2 + dy^2 + dz^2 - dt^2.$$
>
> Line elements of real length $(ds^2 > 0)$ are called spatial, however those of purely imaginary length $(ds^2 < 0)$ are called timelike.

Born's next paper on the definition of a rigid body (Born 1910, 233) speaks of "a four-dimensional space $x\ y\ z\ t$ ... in which a measure relation [*Maßbestimmung*] with the line element $dx^2 + dy^2 + dz^2 - c^2dt^2$ is introduced" (p. 233).

As noted above, Einstein refers to these papers in a 1910 letter, so we may assume that he was familiar with them.[89] And, it was in response to a criticism by Einstein that Abraham wrote:[90]

> On lines 16, 17 of my note "On the Theory of Gravitation," an oversight is to be corrected, of which I became aware through a friendly communication of Mr. A. Einstein. One should read "let us consider dx, dy, dz and $du = i\ dl = i\ cdt$ as components of a displacement ds in four-dimensional space." Thus,

---

85 See Section III, p. 108 of (Minkowski 1909a); it was soon reprinted as a separate booklet, (Minkowski 1909b); and then in Minkowski's *Gesammelte Abhandlungen*, vol. 2, (Minkowski 1911, 431–434).

86 Laue only introduces the proper time and uses it to define the four-velocity in the second edition of *Das Relativitätsprinzip* (Laue 1913, 57 and 69).

87 Einstein does define the four-velocity vector by the following equation:

$$G_\mu = \left\{ \frac{dx_\mu}{\sqrt{-\sum dx_\sigma^2}} \right\},$$

but without any explanation of the expression in the denominator, which does not occur anywhere else in the paper (CPAE 4, 84).

88 See (Maltese and Orlando 1995).

89 "The latest relativity-theoretical investigations of Born and of Herglotz interest me very much ..." Albert Einstein to Jakob Laub, 16 March 1910, (CPAE 5, 231–233; citation from 232).

90 "Berichtigung," (Abraham 1912c). This correction was submitted to issue no. 4, which had a closing date of 2 February 1912, but was published on 15 February 1912.

$$ds^2 = dx^2 + dy^2 + dz - c^2 dt^2$$

> is the square of the four-dimensional line element, in which the velocity of light c is determined by equation (6) [Abraham's relation between the gravitational potential and the speed of light–JS].[91]

Indeed, as noted above, in Paper I on the static field Einstein had written that, in using a light clock:

> we operate with a sort of local time, which Abraham designates with l. This stands in the relation
>
> $$dl = c\ dt$$
>
> to the universal time.

This is the earliest indication (the end of February 1912) that Einstein realized the need to use differentials of the two quantities in order to relate a coordinate time t to a physically measured time l (in this case, the proper time between two events at the same spatial point).

To summarize: on the basis of the mathematical and physical resources at his command, at some point in mid-1912, after generalizing the single gravitational potential c to the array of ten gravitational potentials g_{ik}, Einstein realized that they formed the coefficients of a quadratic form $\sum g_{ik} dx_i dx_k$, which could be regarded as the square of the invariant line element ($ds^2 = \sum g_{ik} dx_i dx_k$) of a four-dimensional spacetime manifold; and that the interval ds represents a physically measurable quantity—the proper time if the interval between two events were time-like, the proper length if it were space-like (of course it would vanish for null intervals).

I suggest that it was at this point that he turned to Grossman. Continuing the quotation from the 1955 reminiscence:

> I was made aware of these [works by Ricci and Levi-Civita—JS] by my friend Großmann in Zurich, when I put to him the problem to investigate generally covariant tensors, whose components depend only on the derivatives of the coefficients of the quadratic fundamental invariant.
>
> He at once caught fire, although as a mathematician he had a somewhat skeptical stance towards physics. ... He went through the literature and soon discovered that the indicated mathematical problem had already been solved, in particular by Riemann, Ricci and Levi-Civita. This entire development was connected to the Gaussian theory of curved surfaces, in which for the first time systematic use was made of generalized coordinates. (Seelig 1955, 15, 16)

91 Abraham reiterated this point in his next paper "Die Erhaltung der Energie und der Materie im Schwerkraftfelde," (Abraham 1912d): "As a result of the variability of c, the Lorentz group only holds in the infinitely small, so that dx, dy, dz and $du = i\ cdt$ represent the components of an infinitely small displacement in a four-dimensional space" (p. 312). Note that, in contrast to Einstein, Abraham's c may be a function of all four spacetime coordinates.

In short: While the analogy to Gaussian surface theory had occurred to Einstein *before* he consulted Grossmann, probably including the role of the line element; the connection between this theory and the later line of development from Riemann to Ricci and Levi-Civita only became clear to Einstein *after* consulting Grossmann.

Louis Kollross, another student friend of Einstein, who was also Professor of Mathematics at the ETH during this time, adds another name that belongs between those of Riemann and of Ricci and Levi-Civita:

> [Einstein] spoke to Großmann about his troubles and said to him one day: "Großmann, you must help me, otherwise I'll go crazy!" And Marcel Großmann succeeded in showing him that the mathematical instrument that he needed had been created precisely in Zurich in the year 1869 by Christoffel in the paper "On the Transformation of Homogeneous Differential Expressions of the Second Degree," published in volume 70 of "Crelle's Journal" for pure and applied mathematics.[92]

A look at Grossmann's Part II of their joint paper, confirms Kollros's recollection:

> The mathematical tool [*Hilfsmittel*] for the development of the vector analysis of a gravitational field that is characterized by the invariance of the line element
>
> $$ds^2 = \sum_{\mu\nu} g_{\mu\nu} dx_\mu dx_\nu$$
>
> goes back to the fundamental paper of Christoffel on the transformation of quadratic differential forms.[93]

So it was Marcel Grossmann, who introduced Einstein to the work of Ricci and Levi-Civita after Einstein's return to Zurich in early August 1912.[94] However, in his exposition Grossmann plays down the geometrical significance of vector and tensor analysis:

> In it I have purposely left geometrical methods [*Hilfsmittel*] aside, since in my opinion they contribute little to the visualization [*Veranschaulichung*] of the concepts constructed in vector analysis (p. 325).

This distinction between tensor analysis and geometrical methods is based on the distinction Ricci and Levi-Civita make between the "fundamental quadric or form" (p. 13),which they denote by ϕ, and the line element (they never use these words), denoted by ds^2, of an n-dimensional manifold, denoted by V_n (see, e.g., pp. 128, 153). They assert that: "The methods of the absolute differential calculus depend essentially on consideration of" the fundamental form (p. 133); but the geometrical interpretation of it "as the ds^2 of a surface" (p. 162) is merely one possibility.

---

92 "Erinnerungen-Souvenirs," (Kollross 1955b); reprinted as "Erinnerungen eines Kommilitionen," (Kollross 1955a). Citation from p. 27.

93 "Entwurf einer verallgemeinerten Relativitätstheorie und einer Theorie der Gravitation II. Mathematischer Teil von Marcel Grossmann," (Einstein and Grossmann 1913, 23–38); reprinted in (CPAE 4, 324–339; citation from p. 324). The paper cited is (Christoffel 1869).

94 "Méthodes de calcul différentiel absolu et leurs applications," (Ricci and Levi-Civita 1901).This paper has been translated into English in (Hermann 1975).

Grossmann's exposition of tensor analysis is based on Chapter I, "Algorithm of the Absolute Differential Calculus" (pp. 128–144), which includes discussions of covariant differentiation and of the Riemann tensor that do not depend at all upon the geometrical interpretation of the fundamental form, but rather on the theory of algebraic and differential invariants of the fundamental form and other functions (see pp. 127 of the "Preface" and Section 1 of the first chapter, "Point transformations and systems of functions," pp. 128–130).[95] This is entirely in the spirit of Christoffel's exposition of the differential invariants of a quadratic differential form in n independent variables. Only in the last paragraph of his paper does he mention "a posthumous paper of Riemann" on "the square of the line element in a space of three dimensions" (Christoffel 1869, 70). And indeed, until Levi-Civita developed the concept of parallel displacement in a manifold with metric, geometrical methods did not contribute much to the interpretation of the covariant derivative and the Riemann tensor.[96] Only starting with Chapter II of Ricci and Levi-Civita, on "Intrinsic geometry as a calculational tool," are geometrical applications to n-dimensional manifolds introduced.

Bianchi-Lukat (Bianchi 1910), another source that Grossmann mentions,[97] also separates the invariant-theoretical treatment of "Binary Quadratic Forms" in Chapter II from the geometrical treatment of "Curvilinear Coordinates on Surfaces" in Chapter III, which includes the introduction of "The Line Element of Surfaces" in Section 33.

—But I have already begun to encroach on the opening scene of Act III. If, so far, Einstein's intuition led him almost without mis-step along the highway in the search for dynamical field equations governing the behavior of the metric tensor field, the last act will depict our hero's wanderings along many a curious by-way, before regaining the high road.

---

95 Only in Section 4 of this chapter, "Applications to vector analysis," is "the ds^2 of [Euclidean three-] space as the fundamental form" introduced (p. 135). The paper has a number of lapses: for example, the fundamental form is introduced on p. 130 without the name, and the notation ϕ for it is used on p. 132, before both are defined on p. 133. Most serious, the Christoffel symbols of the second kind are introduced and used to define the covariant derivative on p. 138, without ever being defined or related to the symbols of the first kind, which are defined and then used to define the covariant form of the Riemann tensor (called "the covariant system of Riemann) on p. 142.

96 For a discussion of this question, see "The Story of Newstein or: Is Gravity just Another Pretty Force?" (in vol. 4 of this series).

97 See ""Entwurf einer verallgemeinerten Relativitätstheorie und einer Theorie der Gravitation II. Mathematischer Teil von Marcel Grossmann," (CPAE 4, 330).

REFERENCES

Abraham, Max. 1912a. "Zur Theorie der Gravitation." *Physikalische Zeitschrift* 13: 1–4. (English translation in vol. 3 of this series.)

———. 1912b. "Das Elementargesetz der Gravitation." *Physikalische Zeitschrift* 13: 4–5.

———. 1912c. "Berichtigung (zu: Zur Theorie der Gravitation, 1912)." *Physikalische Zeitschrift* 13: 176.

———. 1912d. "Die Erhaltung der Energie und der Materie im Schwerkraftfelde." *Physikalische Zeitschrift* 13: 311–314.

———. 1912e. "Relativität und Gravitation. Erwiderung auf eine Bemerkung des Hrn A. Einstein." *Annalen der Physik* 38: 1056–1058.

Bianchi, Luigi. 1910. *Vorlesungen über Differentialgeometrie,* 2 ed. translated by M. Lukat. Leipzig/Berlin: Teubner.

Born, Max. 1909a. "Die Theorie des starren Elektrons in der Kinematik des Relativitätsprinzips." *Annalen der Physik* (30) 11: 1–56.

———. 1909b. "Über die Dynamik des Elektrons in der Kinematik des Relativitätsprinzips." *Physikalische Zeitschrift* (10) 22: 814–817.

———. 1910. "Über die Definition des starren Körpers in der Kinematik des Relativitätsprinzips." *Physikalische Zeitschrift* (11) 6: 233–234.

Christoffel, Elwin Bruno. 1869. "Ueber die Transformation der homogenen Differentialausdrücke zweiten Grades." *Journal für die reine und angewandte Mathematik* 70: 46–70.

Coolidge, Julian Lowell. 1940. *A History of Geometrical Methods*. Oxford: Clarendon Press.

CPAE 1: John Stachel, David C. Cassidy, Robert Schulmann, and Jürgen Renn (eds.), *The Collected Papers of Albert Einstein*. Vol. 1. *The Early Years, 1879–1902*. Princeton: Princeton University Press, 1987.

CPAE 2: John Stachel, David C. Cassidy, Jürgen Renn, and Robert Schulmann (eds.), *The Collected Papers of Albert Einstein*. Vol. 2. *The Swiss Years: Writings, 1900–1909*. Princeton: Princeton University Press, 1989.

CPAE 3: Martin J. Klein, A. J. Kox, Jürgen Renn, and Robert Schulmann (eds.), *The Collected Papers of Albert Einstein*. Vol. 3. *The Swiss Years: Writings, 1909–1911*. Princeton: Princeton University Press, 1993.

CPAE 4: Martin J. Klein, A. J. Kox, Jürgen Renn, and Robert Schulmann (eds.), *The Collected Papers of Albert Einstein*. Vol. 4. *The Swiss Years: Writings, 1912–1914*. Princeton: Princeton University Press, 1995.

CPAE 5: Martin J. Klein, A. J. Kox, and Robert Schulmann (eds.), *The Collected Papers of Albert Einstein*. Vol. 5. *The Swiss Years: Correspondence, 1902–1914*. Princeton: Princeton University Press, 1993.

CPAE 6: A. J. Kox, Martin J. Klein, and Robert Schulmann (eds.), *The Collected Papers of Albert Einstein*. Vol. 6. *The Berlin Years: Writings, 1914–1917*. Princeton: Princeton University Press, 1996.

CPAE 7: Michel Janssen, Robert Schulmann, József Illy, Christoph Lehner, and Diana Kormos Buchwald (eds.), *The Collected Papers of Albert Einstein*. Vol. 7. *The Berlin Years: Writings, 1918–1921*. Princeton: Princeton University Press, 2002.

CPAE 8: Robert Schulmann, A. J. Kox, Michel Janssen, and József Illy (eds.), *The Collected Papers of Albert Einstein*. Vol. 8. *The Berlin Years: Correspondence, 1914–1918*. Princeton: Princeton University Press, 1998. (Part A: 1914–1917, pp. 1–590. Part B: 1918, pp. 591–1118.)

Ehrenfest, Paul. 1909. "Gleichförmige Rotation starrer Körper und Relativitätstheorie." *Physikalische Zeitschrift* 10: 918.

———. 1913a. "On Einstein's Theory of the Stationary Gravitation Field." *Proceedings Amsterdam Academy* 15: 1187–1191.

———. 1913b. "Over Einstein's Theorie van het stationaire gravitatieveld." *Verslagen Akademie Amsterdam* 21: 1234–1239.

Einstein, Albert. 1905. "Zur Elektrodynamik bewegter Körper." *Annalen der Physik* 17: 891–921, (CPAE 2, Doc. 23).

———. 1907. "Über das Relativitätsprinzip und die aus demselben gezogenen Folgerungen." *Jahrbuch der Radioaktivität und Elektronik* 4: 411–462, (CPAE 2, Doc. 47).

———. 1908. "Berichtigungen." *Jahrbuch der Radioaktivität und Elektronik* 5: 98–99, (CPAE 2, Doc. 49).

———. 1911a. "Die Relativitäts-Theorie." *Naturforschende Gesellschaft in Zürich, Vierteljahresschrift* 56: 1–14, (CPAE 3, Doc.17).

———. 1911b. "Über den Einfluß der Schwerkraft auf die Ausbreitung des Lichtes." *Annalen der Physik* 35: 898–908, (CPAE 3, Doc. 23).

———. 1912a. “Gibt es eine Gravitationswirkung, die der elektrodynamischen Induktionswirkung analog ist?” *Vierteljahresschrift für gerichtliche Medizin und öffentliches Sanitätswesen* 44: 37–40, (CPAE 4, Doc. 7).
———. 1912b. “Lichtgeschwindigkeit und Statik des Gravitationsfeldes.” *Annalen der Physik* 38: 355–369, (CPAE 4, Doc. 3). (Cited in text as Paper I.)
———. 1912c. “Relativität und Gravitation. Erwiderung auf eine Bemerkung von M. Abraham.” *Annalen der Physik* 38: 1059–1064, (CPAE 4, Doc. 8).
———. 1912d. “Zur Theorie des statischen Gravitationsfeldes.” *Annalen der Physik* 38: 443–458, (CPAE 4, Doc. 4). (Cited in text as Paper II.)
———. 1916a. “Die Grundlage der allgemeinen Relativitätstheorie.” *Annalen der Physik* 49: 769–822, (CPAE 6, Doc. 30).
———. 1916b. “Ernst Mach.” *Physikalische Zeitschrift* 17: 101–104, (CPAE 6, Doc. 29).
———. 1920a. *Relativity, the Special and the General Theory, A Popular Exposition*. London: Methuen.
———. 1920b. *Über die spezielle und allgemeine Relativitätstheorie* (*Gemeinverständlich*), 5 ed. Braunschweig: Vieweg.
———. 1921. “A Brief Outline of the Development of the Theory of Relativity.” *Nature* 106: 782–784.
———. 1934. Einiges über die Entstehung der allgemeinen Relativitätstheorie. In *Mein Weltbild*, 248–256. Amsterdam: Querido.
———. 1949. “Autobiographical Notes.” In P. A. Schilpp (ed.), *Albert Einstein: Philosopher-Scientist*, 2–94. Open Court: La Salle/Cambridge University Press, London.
———. 1955. “Erinnerungen-Souvenirs.” *Schweizerische Hochschulzeitung* (Sonderheft) 28: 145–153.
———. 1979. *Autobiographical Notes: A Centennial Edition*. Open Court: La Salle and Chicago.
Einstein, Albert, and Marcel Grossmann. 1913. *Entwurf einer verallgemeinerten Relativitätstheorie und einer Theorie der Gravitation*. Leipzig: Teubner, (CPAE 4, Doc. 13).
Einstein, Albert, and Jakob Laub. 1908. “Über die elektromagnetischen Grundgleichungen für bewegte Körper.” *Annalen der Physik* 26: 532–540, (CPAE 2, 509–517).
Frank, Michael L. 1912. “Bemerkung betreffs der Lichtausbreitung in Kraftfeldern.” *Physikalische Zeitschrift* 13: 544–545.
Gauss, Carl Friedrich. 1825. “Allgemeine Auflösung der Aufgabe die Theile einer gegebenen Fläche auf einer anderen gegebenen Fläche so abzubilden dass die Abbildung dem Abgebildeten in den kleinsten Theilen ähnlich wird.” *Astronomische Abhandlungen* Heft 3, H. C. Schumacher (ed.). Altona.
———. 1827. “Anzeige: Disquisitiones generales circa superficies curvas.” *Göttingische gelehrte Anzeigen* Stück 177: 1761–1768.
———. 1828. “Disquisitiones generales circa superficies curvas.” *Commentationes societatis regiae scientarium Gottingensis recentiores* 6: 99–146.
———. 1881. Disquisitiones generales circa superficies curvas. In *Werke*, vol. 4. Göttingen: Königliche Gesellschaft der Wissenschaften.
Herglotz, Gustav. 1910. “Über den vom Standpunkt des Relativitätsprinzips aus als “starr” zu bezeichnenden Körper.” *Annalen der Physik* 31: 393–415.
Hermann, Robert. 1975. “Ricci and Levi-Civita’s Tensor Analysis Paper,” translation and comments by R. Hermann. In *Lie Groups: History, Frontiers and Applications*. Series A, vol. 2. Brookline: Math Sci Press.
Howard, Don, and John Stachel (eds.). 1989. *Einstein and the History of General Relativity.* (*Einstein Studies*, Vol. 1) Boston: Birkhäuser.
Kaluza, Theodor. 1910. “Zur Relativitätstheorie.” *Physikalische Zeitschrift* (11) 21/22: 977–978.
Kollross, Louis. 1955a. “Erinnerungen eines Kommilitionen.” In (Seelig 1955).
———. 1955b. “Erinnerungen-Souvenirs.” *Schweizerische Hochschulzeitung* (Sonderheft) 28: 169–173.
Laue, Max von. 1911. *Das Relativitätsprinzip*. Braunschweig: Vieweg.
———. 1913. *Das Relativitätsprinzip*, 2. ed. Braunschweig: Vieweg.
Lorentz, Hendrik Antoon. 1899–1900a. “Beschouwingen over de zwaartekracht.” *Verslagen van de Gewone Vergaderingen der Wis- en Natuurkundige Afdeeling, Koninklijke Akademie van Wetenschappen te Amsterdam* 8: 603–620.
———. 1899–1900b. “Considerations on Gravitation.” *Proceedings of the Section of Sciences, Koninklijke Akademie van Wetenschappen te Amsterdam* 2: 559–574.
Maltese, Giulio, and Lucia Orlando. 1995. “The Definition of Rigidity in the Special Theory of Relativity and the Genesis of the General Theory of Relativity.” *Studies in the History and the Philosophy of Modern Physics* 26B: 263–303.
Minkowski, Hermann. 1908. “Die Grundgleichungen für die elektromagnetischen Vorgänge in bewegten Körpern.” *Nachrichten der Königliche Gesellschaft der Wissenschaften zu Göttingen. Mathematisch-physikalische Klasse* 53–111. (English translation of excerpts in vol. 3 of this series.)
———. 1909a. “Raum und Zeit.” *Physikalische Zeitschrift* (10) 3: 104–111.

———. 1909b. *Raum und Zeit*. Leipzig: Teubner.
———. 1911. *Gesammelte Abhandlungen*, D. Hilbert (ed.), vol. 2. Leipzig: Teubner.
Norton, John. 1985. "What was Einstein's Principle of Equivalence?" *Studies in the History and Philosophy of Science* 16: 203–246.
———. 1992. "Einstein, Nordström and the Early Demise of Scalar, Lorentz Covariant Theories of Gravitation." *Archives for History of Exact Sciences* 45: 17–94.
Pais, Abraham. 1982. *'Subtle is the Lord ...': the Science and the Life of Albert Einstein*. Oxford: Clarendon Press /New York: Oxford University Press.
Reich, Karin. 1994. *Die Entwicklung des Tensorkalküls: vom absoluten Differentialkalkül zur Relativitätstheorie*. *Science Networks: Historical Studies*, vol. 11. Basel/Boston/Berlin: Birkhäuser.
Ricci, Gregorio, and Tullio Levi-Civita. 1901. "Méthodes de calcul différentiel absolu et leurs applications." *Mathematische Annalen* 54: 125–201.
Riemann, Bernhard. 1868. "Ueber die Hypothesen, welche der Geometrie zugrunde liegen." *Abhandlungen der königlichen Gesellschaft der Wissenschaften zu Göttingen* 13: 133–150.
Seelig, Carl (ed.). 1955. *Helle Zeiten - Dunkle Zeiten: In Memoriam Albert Einstein*. Zürich/Stuttgart/Wien: Europa Verlag.
———. 1981. *Mein Weltbild*. Frankfurt am Main/Berlin/Wien: Ullstein.
Sommerfeld, Arnold. 1910a. "Zur Relativitätstheorie I. Vierdimensionale Vektoralgebra." *Annalen der Physik* 32: 749–776.
———. 1910b. "Zur Relativitätstheorie II. Vierdimensionale Vektoranalysis." *Annalen der Physik* 33: 649–689.
Stachel, John. 1980. "Einstein and the Rigidly Rotating Disk." In A. Held (ed.), *General Relativity and Gravitation One Hundred Years After the Birth of Albert Einstein*, vol. 1, 1–15. New York/London: Plenum Press.
———. 1982. "Einstein." *Science* 218: 989–990.
———. 1989. "Einstein's Search for General Covariance, 1912–1915." In D. Howard and J. Stachel (eds.), *Einstein and the History of General Relativity: based on the Proceedings of the 1986 Osgood Hill Conference, North Andover, Massachusetts, 8–11 May 1986*, 63–100. (*Einstein Studies,* Vol. 1) Boston: Birkhäuser.
———. 2002. *Einstein from 'B' to 'Z'*. (*Einstein Studies,* Vol. 9). Boston: Birkhäuser.
Stäckel, Paul. 1918. "Gauß als Geometer." In F. Klein et. al. (eds.), *Materialien für eine wissenschaftliche Biographie von Gauß*. Leibzig: Teubner.
Walter, Scott. 1999. "The Non-Euclidean Style of Minkowskian Relativity." In J. Gray (ed.), *The Symbolic Universe: Geometry and Physics 1890–1930*, 91–127. Oxford: Oxford University Press.

JÜRGEN RENN AND TILMAN SAUER

PATHWAYS OUT OF CLASSICAL PHYSICS

Einstein's Double Strategy in his Search for the Gravitational Field Equation

CONTENTS

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 1
Einstein's Zurich Notebook: *Introduction and Source*.

1. INTRODUCTION

1.1 The Incomplete Revolution

The relativity revolution was far from complete when Einstein published his path-breaking paper on the electrodynamics of moving bodies in 1905. It started with his reinterpretation of Lorentz's theory of electromagnetism in what may be called a "Copernicus process" in analogy to the transition from the Ptolemaic to the Copernican world system or to the transition from preclassical to classical mechanics.[1] In such a transition the formalism of an old theory is largely preserved while its semantics change.[2] Einstein's special theory of relativity of 1905 had altered the semantics of such fundamental concepts like space and time, velocity, force, energy, and momentum, but it had not touched Newton's law of gravitation. Since, however, according to special relativity, physical interactions cannot propagate faster than light, Newton's well-established theory of gravitation, based on instantaneous action at a distance, was no longer acceptable after 1905. The relativity revolution was completed only when this conflict was resolved ten years later in November 1915 with Einstein's formulation of the general theory of relativity.

Neither the emergence of the special theory of relativity nor that of the general theory of relativity were isolated achievements. The virtual simultaneity of the beginning of the relativity revolution with Einstein's other breakthrough discoveries of 1905 indicate that his non-specialist outlook and, in particular, his youthful pursuit of atomistic ideas enabled him to activate the hidden potentials of highly specialized nineteenth-century physics that others, such as Henri Poincaré, had also exposed.[3] In 1907, Einstein first attempted to address the issue as to how to modify Newton's law of gravitation according to the new kinematic framework of special relativity, as did others, like Hermann Minkowski and Henri Poincaré.[4] But Einstein began to transcend the very special-relativistic framework in light of Galileo's insight that in a vacuum all bodies fall with the same acceleration. In 1912, to the amazement of colleagues like Max Abraham,[5] he abandoned the scalar gravitational potential of Newtonian physics in favor of a ten-component object—the metric tensor—the mathematics of which he subsequently began to explore with the help of his mathematician friend Marcel Grossmann. And he was able to formulate clear-cut criteria which a field equation for the metric tensor acting as a gravitational potential would have to satisfy. However, in the winter of 1912–1913, Einstein and Grossmann dis-

1 Cf. (Damerow et al. 2004, Renn 2004).
2 Such a change of semantics may be illustrated with the example of Lorentz's concept of local time. Originally merely a peripheral aspect of his theory, this auxiliary variable was reinterpreted by Einstein as the time actually measured by clocks in a moving reference system, thus assuming a central role in the new kinematics of special relativity. For an extensive treatment of the first phase of the relativity revolution compatible with this view, see, e.g., (Janssen 1995).
3 See (Renn 1993, 1997). For the parallelism between Einstein and Poincaré, see also (Galison 2003).
4 See Scott Walter's "Breaking in the 4-vectors ..." (in vol. 3 of this series) and (Katzir 2005).
5 See (Cattani and De Maria 1989a) and "The Summit Almost Scaled ..." (in vol. 3 of this series).

carded generally-covariant field equations based on the Riemann tensor, an expression that included second-order derivatives of the metric tensor. Einstein even believed to have a proof that such field equations had to be ruled out, although in hindsight these were the only acceptable mathematical solution. In spite of the skepticism of many of his physics colleagues but supported by the critical sympathy of mathematicians like Tullio Levi-Civita and David Hilbert,[6] Einstein stood by his original agenda and in late 1915 returned to field equations based on the Riemann tensor, finally formulating the general theory of relativity, a theory which became the basis of all subsequent developments in physics and astronomy.

Einstein's Zurich Notebook represents a uniquely valuable and, as it turns out, surprisingly coherent,[7] record of his thinking in an intermediate phase of the emergence of general relativity. The entries begin in mid-1912 and end in early 1913. His aim during this period was to create a relativistic theory of gravitation that makes sense from a physical point of view and that, at the same time, corresponds to a consistent mathematical framework based on the metric tensor. Central to his thinking was the problem of interpreting the physical knowledge on gravitation in terms of a generalization of the mathematical representation associated with Minkowski's four-dimensional spacetime. The main challenge he faced was to construct a field equation, on the one hand, that can be reduced by an appropriate specialization to the familiar Newtonian law of gravitation, and, on the other hand, that satisfies the requirements resulting from his ambitious program to formulate a relativistic theory of gravitation.

There is perhaps no single episode that better illustrates the conceptual turn associated with the genesis of general relativity than the fact that, in the Zurich Notebook, Einstein first wrote down a mathematical expression close to the correct field equation and then discarded it, only to return to it more than three years later. Why did he discard in the winter of 1912–1913 what appears in hindsight to be essentially the correct gravitational field equation, and what made this field equation acceptable in late 1915?[8] Our analysis of the Zurich Notebook has made it possible not only to answer these questions but, more generally, to resolve what might be called the three epistemic paradoxes raised by the genesis of general relativity:

The paradox of missing knowledge. How was it possible to create a theory such as general relativity that was capable of accounting for a wide range of phenomena which were only later discovered in the context of several revolutions of observational astronomy? If neither the expansion of the universe, black holes, gravitational lenses, nor gravitational radiation were known when Einstein set up the gravitational field equation, how could he nevertheless establish such a firm foun-

---

6 See Einstein's correspondence with Levi-Civita and Hilbert in (CPAE 8). For further discussion, see also (Cattani and De Maria 1989b) and (Corry 2004).

7 See the "Commentary ..." (in vol. 2 of this series) and especially our reconstruction in section 6 of the present chapter.

8 For further discussion of these two questions, see also "Commentary ..." sec. 5, and "Untying the Knot ..." (both in vol. 2 of this series).

dation for modern cosmology? Which knowledge granted such stability to a theory that did not initially seem superior to its competitors, since no phenomena were known at the time which could not also be explained with traditional physics?

The paradox of deceitful heuristics. After a tortuous search in the course of which he even temporarily abandoned hope of ever solving his problem, how was Einstein able to formulate the criteria for a gravitational field equation years before he established the solution? How could he establish a heuristic framework that would quickly lead him to a correct mathematical expression, and then to the conclusion that it was unacceptable, only to bring him back to essentially the same expression three years later?

The paradox of discontinuous progress. How could general relativity with its non-classical consequences—such as the dependence of space and time on physical interactions—be the outcome of classical and special-relativistic physics although such features are incompatible with their conceptual frameworks?

Addressing the challenges which these paradoxes formulate requires taking into account all of the following dimensions that are crucial to a historical epistemology of scientific knowledge: the long-term character of knowledge development, the complex architecture of knowledge, and the intricate mechanisms of knowledge dynamics. In order to resolve these paradoxes and to adequately describe the reorganization of knowledge occurring between 1912 and 1915, we shall, in particular, make use of concepts from cognitive science, adapted to the description of the structures of shared knowledge resources such as those Einstein adopted from classical and special-relativistic physics. These concepts will be used to analyze the architecture of the knowledge relevant to Einstein's search for a gravitational field equation and to explain its restructuring as a result of the interaction with the mathematical representation of this knowledge.

We intend to show in the following that the history of Einstein's search for a gravitational field equation can, against the background of the Zurich Notebook, be written as that of a mutual adaptation of mathematical representation and physical meaning. The eventual success of this adaptation becomes intelligible only if it is conceived of as part of a long-term process of integrating intellectual resources relevant to Einstein's problem that were rooted in the shared knowledge of classical and special-relativistic physics.

Only by analyzing the complex architecture of these shared knowledge resources is it possible to understand in which sense classical and special-relativistic knowledge about gravitation and inertia, energy and momentum conservation, and the relation between different reference frames, was turned into a heuristic framework for Einstein's search. In the course of his work, elements of this heuristic framework crystallized into a double strategy that shaped his search in an essential way until he succeeded in formulating the definitive field equation of general relativity in November 1915.

The identification of the two components of this double strategy has not only allowed us to reconstruct Einstein's notes and calculations in the Zurich Notebook as traces of a surprisingly coherent research process, but also to analyze the dynamics of this process. It has become clear, in particular, how a combination of knowledge resources rooted in classical and special-relativistic physics could give rise to the theory of general relativity whose conceptual foundation is no longer compatible with the knowledge that formed the starting point of Einstein's search. In this way, the genesis of general relativity can be understood as resulting from a transformation of shared resources of knowledge, while Einstein's search for the gravitational field equation appears as an investigation of pathways out of classical physics.

In this introduction, we shall briefly recapitulate the essential elements of our story.[9] We begin with a review of the principal steps taken by Einstein towards a relativistic theory of gravitation between the years 1907 and 1912 before his research is documented in the Zurich Notebook.[10] Here our aim is to show that each of these steps highlighted knowledge resources that were relevant for addressing the challenge of constructing a relativistic theory of gravitation. The heuristics at work in the Zurich Notebook were the result of this prior research experience. We shall then offer a first description of the crucial role played by Einstein's double strategy for his heuristics and finally introduce the epistemological framework for our analysis of how exactly this strategy worked.

In the second section, we shall discuss what we will call Lorentz model, as the conceptual framework for Einstein's construction of a relativistic field theory of gravitation. In the third section, we shall examine the essential elements of his heuristics, showing in which sense these elements turned knowledge resources of classical and special-relativistic physics into key components of Einstein's search. In the fourth section, we shall analyze how this search process was structured by the way in which the Lorentz model functioned as a mental model in the sense of cognitive science. In the fifth section, we shall examine how the candidates for a gravitational field equation that Einstein considered in the course of his search fared in the light of the heuristic criteria he had established on the basis of his prior research experience. This discussion will help to understand why one and the same candidate fared differently depending on the depth to which Einstein had explored the implications of the mathematical representation. In the sixth section, we shall reconstruct Einstein's pathway, as documented in the Zurich Notebook, as a learning experience in which he passed from one candidate field equation to the other, building up strategic devices that would guide him until he reached his final result in 1915. In the seventh section, we

---

9 For Einstein's own account, see (Einstein 1933). The history of general relativity has been an intensive subject of research in the last decades, see, in particular, the contributions in (Stachel and Howard 1989–2006). Especially with respect to Einstein's own path, early contributions were (Hoffmann 1972, Lanczos 1972, Mehra 1974, Earman and Glymour 1978, Vizgin and Smorodinski 1979, Pais 1982, sec. IV., Stachel 1980, 1982), a groundbreaking paper was (Norton 1984). See also (Capria 2005, Howard and Norton 1992, Janssen 1999, 2005, Maltese 1991, Maltese and Orlando 1995, Miller 1992, Norton 1992a, 1992b, 1999, 2000, Renn 2005b, 2005c, Renn and Sauer 1996, 1999, 2003a, Sauer 2005b, Stachel 1987, 1989b, 1995, 2002, Vizgin 2001).

10 For detailed analyses of this part of the story, see "The First Two Acts" and "Classical Physics in Disarray …" (both in this volume).

shall turn to Einstein's elaboration of the so-called *Entwurf* theory, published in 1913 as the result of the research documented in the Zurich Notebook. It will be shown, in particular, how the work on this problematic theory created the preconditions for the conceptual changes of the final theory of general relativity. In the concluding eighth section, we shall review our reconstruction with a view to pinpointing the essential structures of this scientific revolution.

1.2 The Emergence of a Heuristic Framework

The incompatibility between Newton's theory of gravitation and the special theory of relativity of 1905 presented Einstein and his contemporaries with the task of constructing a relativistic theory of gravitation. Special relativity, for the purpose of our account, arose from the confrontation of classical mechanics and classical electrodynamics as two major knowledge blocks, i.e. from the confrontation of two highly elaborated, individually consistent, and empirically well-confirmed systems of knowledge whose simultaneous validity had nevertheless produced inconsistencies and contradictions. The newly established mathematical and conceptual framework of special relativity added to the physical knowledge available for dealing with the problem of a relativistic theory of gravitation. The knowledge blocks of classical mechanics and electrodynamics and of special relativity offered various points of departure for the continuation of the relativity revolution in coming to terms with the problem of gravitation.

The new spatio-temporal framework of special relativity suggested a plausible mathematical procedure for adapting the classical theory of gravitation to the requirements of a relativistic field theory. In classical physics, the Poisson equation determines the Newtonian gravitational potential by a given distribution of the masses that act as the sources of the gravitational field (which in turn can be derived from the gravitational potential).[11] This equation is not invariant with respect to the Lorentz transformations of special relativity. But the Poisson equation can easily be extended in a formal way to a relativistic field equation by adding a differential operator involving the time coordinate. The problem with this obvious generalization was that the resulting theory of gravitation no longer incorporates Galileo's principle according to which all bodies fall with the same acceleration. The most obvious way of bringing gravitation within the purview of the relativity revolution therefore came at the price of having to give up one of the fundamental insights of classical mechanics.

At this point, classical mechanics provided knowledge resources that were turned into an alternative heuristic starting point for the continuation of the relativity revolution. In 1907 Einstein formulated his principle of equivalence as a heuristic device

11 The Poisson equation, being an equation for the gravitational potential, should properly be called a potential equation. However, since the Einstein equations are commonly referred to as "field equations" rather than "potential equations," we will in the following loosely also refer to the Poisson equation as a "field equation."

that allowed him to incorporate Galileo's principle into a relativistic theory of gravitation.[12] The equivalence principle asserts that it is not possible to distinguish between a uniformly and rectilinearly accelerated reference frame without gravitational fields and an inertial system with a static and homogeneous gravitational field. Accordingly, the problem of a revision of the classical theory of gravitation became associated with that of a generalization of the relativity principle to accelerated motion, which henceforth constituted another heuristic guideline for Einstein's further research.[13]

Between 1907 and 1911 Einstein used the equivalence principle to derive several consequences of his yet to be formulated new gravitation theory.[14] By the spring of 1912, he made a first attempt at formulating a theory for a static but otherwise arbitrary gravitational field.[15] The gravitational field equation of this theory was a straightforward modification of the Poisson equation of classical physics. Since the Poisson equation embodies the classical knowledge of gravitation from Newtonian theory, it formed a crucial asset for Einstein's heuristics. The further elaboration of Einstein's theory of the static field met with great difficulties. He found that this theory was incompatible with the conservation of energy and momentum, another pillar of classical physics. This led to another key element of his heuristic framework, the requirement that the conservation laws must be fulfilled.

The various heuristic requirements serving as different starting points for the search for a relativistic theory of gravitation could lead into different directions, confronting it with different obstacles and different intermediate results, as well as leading perhaps to different solutions to the original problem. After finding a more or less satisfactory theory of the static field, Einstein further pursued the heuristics embodied in the equivalence principle and in the knowledge about field theory available in classical physics. This approach led him to consider uniformly rotating reference frames.[16] As with linearly accelerated motion, he sought to interpret the inertial forces occurring in such reference frames as generalized gravitational forces. This interpretation was made plausible by Mach's critical analysis of classical mechanics. But the conceptual and technical difficulties implied by the inclusion of rotating reference frames prevented, for the time being, the formulation of a gravitation theory that covered this more general case as well. In hindsight, it is clear that a response to the difficulties which Einstein encountered required the introduction of more sophisticated mathematical tools. The heuristics based on the equivalence principle led to

12 See (Einstein 1907). For historical discussion, see (Miller 1992).

13 For a discussion of the problematic relation between the equivalence principle and the generalization of the relativity principle, see, for instance, secs. 1.1.1–1.1.2 of "Commentary" (in vol. 2 of this series) and (Janssen 2005, 61–74).

14 See (Einstein 1911).

15 See (Einstein 1912b) and, for historical discussion, (CPAE 4, 122).

16 The crucial role of rotating reference frames in recognizing the role of non-Euclidean geometry was first discussed in (Stachel 1980), see also (Maltese and Orlando 1995).

substantial but isolated physical insights, and not to the kind of coherent mathematical framework necessary for formulating a relativistic field theory of gravitation.

A different path had meanwhile been followed by Max Abraham, who exploited heuristic clues of the four-dimensional mathematical framework established by Minkowski for special relativity.[17] Abraham succeeded in developing a comprehensive theory of gravitation through an *ad-hoc* modification of this framework. Einstein soon discovered weaknesses in Abraham's theory. After a controversy with Abraham, he realized that a successful application of Minkowski's formalism to the problem of gravitation called for a mathematical generalization of this formalism. In late spring 1912 Einstein found the appropriate starting point for such a generalization of Minkowski's formalism. In the appendix to the last paper he published before the considerations documented in the Zurich Notebook, he formulated the equation of motion in a static gravitational field in a form that suggested that a generalization of his theory of gravitation would involve non-Euclidean geometry as had been formulated by Gauss for curved surfaces. As early as summer 1912 Einstein succeeded in formulating a generally-covariant equation of motion for a test particle in an arbitrary gravitational field. In this equation, the gravitational potential is represented by a four-dimensional metric tensor, which became the key object for Einstein's further research in the following years.

The search for a relativistic gravitational field equation, which occupied Einstein for the following three years, also involved a new role of the heuristic clues that had so far guided the research of Einstein and his contemporaries. Initially, these heuristic clues were more or less isolated hints. They gradually turned into elements of a more systematic research program, characterized by what we have called Einstein's "double strategy." This double strategy allowed him to attack the problem of finding a gravitational field equation by bringing to bear on this problem the entire range of knowledge resources embodied in the various heuristic elements sketched above.

1.3 The Double Strategy

The mathematical difficulty of finding a field equation for the ten-component metric tensor representing the gravitational potential showed Einstein that he needed much more sophisticated mathematical methods than those available to him at that point. A mathematical formalism providing what Einstein's generalized theory of relativity required had been developed in the second half of the 19th century by Gauss, Riemann, and Christoffel. In a paper published in 1901 by Ricci and Levi-Civita on the so-called absolute differential calculus, the work of these mathematicians had been extended to an elaborate mathematical apparatus.[18] However, among physicists, the absolute differential calculus remained largely unknown for a considerable time. Einstein was certainly not familiar with it until mid-1912.[19] Only after his move from

17 For a more detailed treatment, see "The Summit Almost Scaled ..." (in vol. 3 of this series).
18 See (Ricci and Levi Civita 1901).

Prague to Zurich did he gain access to these mathematical methods through his contact with Marcel Grossmann. In October 1912, he wrote to Arnold Sommerfeld:

> I am now working exclusively on the gravitation problem and believe that I can overcome all difficulties with the help of a mathematician friend of mine here. But one thing is certain: never before in my life have I troubled myself over anything so much, and I have gained enormous respect for mathematics, whose more subtle parts I considered until now, in my ignorance, as pure luxury! Compared with this problem, the original theory of relativity is child's play.[20]

The mathematical difficulty of finding a satisfactory relativistic field equation also gave a new role to the physical requirements that such an equation had to satisfy. These physical requirements had to be translated into mathematical conditions to be satisfied by candidate field equations. They were thus also brought into systematic relations with each other. This translation was by no means unambiguous, since Einstein was exploring an as yet largely unknown territory of knowledge. At the same time, the theory had to preserve the physical knowledge on gravitation already available, and its relation to other parts of physics, and it had to be formulated as a mathematically consistent theory built, according to Einstein's insight of 1912, around the four-dimensional metric tensor. This combination of relatively clear-cut conditions and the incompleteness of the information needed to turn the situation into a fully determined mathematical problem was characteristic of Einstein's situation when he began the search for a field equation as documented in the Zurich Notebook. The search strategy that gradually emerged enabled a mutual adaptation of mathematical representation and physical concepts, and provided a heuristic device that eventually turned out to be the adequate response to this situation.

What conditions to be imposed on a relativistic gravitational field equation for the metric tensor had emerged from Einstein's prior research experience? From the mathematical point of view, the task was to find a differential operator of second order for the metric tensor covariant with respect to the largest possible class of coordinate transformations. The requirement that the candidate differential operator has to be of second order follows from the analogy with the classical theory of gravitation: the Poisson equation for the Newtonian gravitational potential is a differential equation of second order. The requirement of the covariance of this differential operator under a broad class of coordinate transformations represented for Einstein the goal of a gen-

19 The works by Bianchi (1910) and Wright (1908) probably served as Einstein's mathematical reference books. For historical discussion, see (Reich 1994).

20 "Ich beschäftige mich jetzt ausschliesslich mit dem Gravitationsproblem und glaube nun mit Hilfe eines hiesigen befreundeten Mathematikers aller Schwierigkeiten Herr zu werden. Aber das eine ist sicher, dass ich mich im Leben noch nicht annähernd so geplag[t] habe, und dass ich grosse Hochachtung für die Mathematik eingeflösst bekommen habe, die ich bis jetzt in ihren subtileren Teilen in meiner Einfalt für puren Luxus ansah! Gegen dies Problem ist die ursprüngliche Relativitätstheorie eine Kinderei." Einstein to Arnold Sommerfeld, 29 October 1912, (CPAE 5, Doc. 421). Unless otherwise noted, all translations are based on the English companion volumes to the *Collected Papers of Albert Einstein*.

eralized relativistic theory in which, if possible, all reference frames would be equivalent.[21] A further requirement was that the classical field equation emerge as a special case of the relativistic field equations under appropriate restrictive conditions, such as for weak and static fields. The heuristic framework furthermore included general physical principles such as Galileo's principle and the laws of energy and momentum conservation applying to the energy and momentum of the gravitational field as well.

These requirements formed the relatively stable framing conditions shaping Einstein's search for the gravitational field equation from its beginning in summer 1912 to the formulation of the eventual solution in late 1915. His main problem was to ensure the compatibility of these different heuristic components by integrating them into a coherent gravitation theory represented by a consistent mathematical framework. It turned out that again and again, in the course of his investigations, only some of Einstein's heuristic goals could be fully realized while others had to be given up or at least modified. If not all of his goals could be satisfied, the appropriate balance between the different heuristic requirements for a gravitational field theory could not be decided *a priori*. Their relative weight could only be judged by their concrete embodiment in candidate gravitational field theories.

Physical properties or mathematical statements could each be looked upon either as principles of construction for the building blocks of the theory or as criteria by which the acceptability of such building blocks could be checked. It is this double perspective that provided the basis for the double strategy that emerged in the course of Einstein's search for the gravitational field equation, as documented in the Zurich Notebook. Earlier the choice between physically or mathematically motivated expressions had been a choice between entirely different approaches to the problem of gravitation. Einstein's 1912 theory of static gravitational field was, for instance, motivated by physical considerations based on the equivalence principle, while Abraham's theory started from mathematical considerations related to Minkowski's formalism. In the course of Einstein's work documented in the Zurich Notebook, the two approaches gradually grew closer and turned into complementary strategies of a more or less systematic research program. In this research program the two approaches were distinguished mainly by the sequence in which the building blocks of the theory come into play. Einstein's "physical strategy" took the Newtonian limiting case as its starting point, then turned to the problem of the conservation of energy and momentum and only then examined the degree to which the principle of relativity is satisfied. His "mathematical strategy," took the principle of relativity as its starting point and only then turned to the Newtonian limiting case and the conservation of energy and momentum. The reconstruction of Einstein's notes in the Zurich Notebook has made it evident that his search for the gravitational field equation is to a large extent determined by the exploration of the possibilities offered by these alternatives.

21 See (Norton 1999) for a discussion about Einstein's ambiguity regarding the difference between invariance and covariance during this period.

Einstein's oscillation between these two strategies is characteristic not only of his approach in the notebook but of his entire struggle with the problem of gravitation between 1912 and 1915, a struggle that brought him from his 1912 static theory, via the *Entwurf* theory of 1913, to the final theory of general relativity.[22]

This oscillation between the physical and the mathematical strategy suggests that his search for the gravitational field equation was not just a matter of resolving a well-defined mathematical problem, but involved an interaction between mathematical representation and physical concepts that affects the structures of the mathematical and physical knowledge. Why else did Einstein's first attempts along the mathematical strategy in the winter of 1912–1913 fail, while his pursuit of the physical strategy seemed to be essentially successful, at least until the demise of the *Entwurf* theory in late 1915?[23] As we will show in detail, the completion of the general theory of relativity required, in addition to the appropriation of the available mathematical knowledge, a revision of foundational concepts of physics, the extent of which Einstein could hardly have foreseen at the beginning of his search. He initially believed that classical physics would provide the appropriate context for the theory to be found and attempted to formulate a gravitational field equation by immediate generalization of familiar Newtonian concepts. It eventually turned out to be more successful to construct a field equation corresponding to Einstein's program of integrating gravitation and relativity than to relate it to the conceptual foundations of classical theory.

1.4 The Epistemological Framework of the Analysis

How was it possible for Einstein to formulate a theory involving conceptual novelties on the basis of knowledge that was still anchored in the older conceptual foundation of classical physics? Such a development can hardly be described in terms of formal logic. As Einstein's investigative pathway illustrates, scientific conclusions can result in a reconceptualization of the premises on which these conclusions were based. Even in cases involving major restructuring of knowledge, science never starts from scratch. In fact, not only scientific knowledge but also the knowledge of large domains of human experience transmitted over generations is not simply lost when new scientific theories replace the old ones. In the case at hand, the knowledge of classical physics had to be preserved and exploited in a conceptual revolution, the outcome of which was a relativistic theory of gravitation whose far-reaching physical implications were largely unknown when it was created. But they eventually changed our understanding of the universe. An adequate description of the cognitive dynamics of the genesis of general relativity therefore requires an account of the knowledge that

22 Very similar characteristics of a physical and mathematical double strategy have also been identified in Einstein's later work on unified field theory, see (van Dongen 2002, 2004) and (Sauer 2006) for further discussion. See also the discussion in (Norton 2000).

23 And, as is argued in "Untying the Knot …" (in vol. 2 of this series), beyond the demise of this theory as well.

makes it understandable. We have to understand, first, how past experiences can enter inferences about matters for which only insufficient information is available, and, second, how conclusions can be corrected without eventually having to start from scratch each time a premise is found to be wanting, with the possibility that the whole deductive structure changes in the process. Such an approach is offered by an historical epistemology that integrates the methodology of historical analysis with a theoretical framework informed by philosophical epistemology and cognitive science.

In order to adequately account for the features of Einstein's search for the gravitational field equation described above, we will in the following make use in particular of the concept of a "mental model" and the concept of a "frame."[24] A mental model for us is an internal knowledge representation structure serving to simulate or anticipate the behavior of objects or processes. It possesses "terminals" or "slots" that can be filled with empirically gained information, but also with default assumptions resulting from prior experience. The default assumptions can be replaced in light of new information, so that inferences based on the model can be corrected without abandoning the model as a whole. Information is assimilated to the slots of a mental model in the form of "frames." These are chunks of knowledge which themselves are equipped with terminals and which have a well-defined meaning anchored in a given body of shared knowledge.

Mental models can, as a rule, be externally represented by material models which also serve as the element of continuity in their transmission from one generation to the next. The basic features of the field-theoretical model of distant causation, which will play a central role in our analysis, may, for instance, be represented by the material model of a magnet setting a piece of iron into motion by affecting the state of its environment. In addition, it may be represented by symbolic representations making use of natural and formal language. The internal architecture of a system of knowledge is constituted by a network of mental models and frames that can be linked by operations in the sense of mental acts typically corresponding to handling external representations, be they material arrangements or symbolic expressions. A sequence of such operations constitutes a procedure which typically has a goal, for instance of creating an *ad hoc* knowledge representation structure, which is called a "real-time construction" in cognitive science. A real-time construction may be exemplified by the geometrical construction typically accompanying the Euclidean proof of a geometrical theorem or by a set of mathematical expressions corresponding to checking a candidate field equation according to one of Einstein's heuristic principles.

Two fundamentally important types of mental acts are "chunking" and "reflection." By chunking different knowledge representation structures are combined into a unity. This often leads to a linguistic representation of the resulting chunk by a technical term designating, for instance, a particular procedure. By reflection, the usage

24 For the concepts of frame and mental model, see (Minsky 1975, 1987; Damerow 1996; Gentner and Stevens 1983; and Davis 1984). For a view on the potential of cognitive science and cognitive psychology for the history of science to which the present work is much indebted, see (Damerow 1996).

of knowledge representation structures becomes the object of reasoning; it typically presupposes an external representation of these structures, for instance by a technical term. Reflection obviously plays a crucial role in accommodating a system of knowledge to new experiences by changing its architecture. An example is what we call a "Copernicus process" in which the internal network of a system of knowledge is essentially preserved while originally peripheral elements take on a central role in the deductive structure. The status of such elements as being either peripheral or central is prescribed by a "control structure." A control structure is constituted by any knowledge representation structure serving to control the operation and to order other such structures. In this sense, Einstein's heuristic principles as well as his double strategy may be considered examples of such control structures. These elements of the architecture of knowledge can partially be captured by traditional epistemological terminology. A concept, for instance, may be understood as the linguistic representation of a mental model, a frame, or a particular terminal of a frame, while a theory is just one example of many conceivable control structures. We shall also use of these traditional terms, specifying their meaning in the context of the epistemological framework we have introduced whenever appropriate.

We claim that the shared knowledge of classical and special-relativistic physics can be conceived of in terms of this richer epistemological framework, and that it then becomes understandable how this knowledge could serve as a resource for Einstein's search for the gravitational field equation. We will argue that essential relations between fundamental concepts such as that between field and source remain the same to a great extent even though the concrete applications of these concepts differ considerably from their applications to a classical or a relativistic field equation. This structural stability turned the concepts and principles of classical and special-relativistic physics into guiding principles when Einstein entered unknown terrain, for instance, when he encountered a new expression generated by the elaboration of a mathematical formalism. None of these expressions by themselves constituted a new theory of gravitation. Only by complementing them with additional information based on the experience accumulated in classical and special-relativistic physics, as well as in the relevant branches of mathematics did such expressions become candidates for a gravitational field equation embedded in a full-fledged theory of gravitation. In the language of mental models, such past experience provided the default assumptions necessary to fill the gaps in the emerging framework of a relativistic theory of gravitation. Because of their nature as default assumptions, they could be given up again in the light of novel information without making it necessary to abandon the underlying mental models, which thus continue to play their heuristic role.

In this way we hope to render understandable how a gradual process of knowledge accumulation could overcome the very conceptual foundations that had formed its starting point. The concepts of classical physics shaped Einstein's search at its beginning, and made the physical strategy the most natural approach to exploit his heuristic principles for finding the gravitational field equation. In the context of the physical strategy, the default assumptions of the relevant mental models were supplied by the knowledge of classical and special-relativistic physics. The mathematical strategy, on the other hand, drew on default assumptions based on prior mathematical knowledge and led to candidate field equations whose compatibility with established physical knowledge was problematic. The gradual accumulation of knowledge fostered by both of these approaches enriched the network constituted by the mental models and frames relevant to a relativistic theory of gravitation. Eventually, the reorganization of this network by a Copernicus process became feasible.

In the following, we shall discuss in detail the essential aspects and phases of this process, Einstein's heuristic framework, the gradual accumulation of knowledge in the course of his research, the successive replacement of one candidate gravitational field equation by another, the switches back and forth between the physical and the mathematical strategy, and finally the reinterpretation of the results acquired in this way as aspects of one and the same transformation leading from the system of knowledge of classical to that of general-relativistic physics.

2. THE MENTAL MODEL OF FIELD THEORY

2.1 The Poisson Equation of Classical Mechanics and the Field Equation of General Relativity

The revision of Newton's theory of gravitation confronted Einstein with two fundamental problems. He needed to find an equation of motion for bodies in a gravitational field (the analogue of Lorentz's equation of motion of a charged body in an electromagnetic field) and to find a field equation determining the gravitational field itself (the generalization of the Poisson equation and the analogue of Maxwell's equations relating the electromagnetic field to its sources). These two problems presented themselves in terms of basic concepts and structures of classical physics and the special theory of relativity. These concepts and structures also provided an essential part of the intellectual resources for solving these problems.

The most fundamental structures of knowledge relevant to Einstein's search for a new theory of gravitation were incorporated in the understanding of an equation of motion and of a field equation in classical physics and in the special theory of relativity. This understanding involves concepts such as force, energy, momentum, potential, field, source, and mass.[25] Above all, however, it involves the mental model of field theory which had emerged, in its most mature, successful, and widely accepted form, in Lorentz's electron theory of electrodynamics and hence may also be referred to as the Lorentz model.[26] This model actually comprises two mental models with more ancient roots in the history of physics; one for a field equation and one for the equation of motion. The Lorentz model of a field equation, which will be at the center

---

25 For Einstein's account of the emergence of fundamental concepts of physics, including that of field, see (Einstein and Infeld 1938).

26 See (Lorentz 1895), and for historical discussion (Whittaker 1951, ch. XIII, 1953, ch. II, Buchwald 1985, Janssen 1995, Darrigol 2000, Janssen and Stachel 2004).

of our analysis of Einstein's search for the gravitational field equation, has slots for the source, the potential, and a differential operator acting on the potential. Default settings for these slots are provided by the classical theory of gravitation which describes the relation between gravitational source and gravitational potential in terms of the Poisson equation. In the classical case, the source-slot and the potential-slot of the frame are filled by scalar functions that can be subsumed under what we might call the potential-frame and the mass-density-frame, respectively. The default setting for the differential-operator-slot is the Laplace operator.

Before we come back to a more detailed examination of the structure of the Lorentz model, we want to justify the introduction of this model by examining some of the basic concepts and knowledge structures relating the Poisson equation in classical mechanics to the Einstein field equation of general relativity. We claim that these common features played an important role in the historical development linking the two equations so that their description by an overarching structure makes historical sense.

The Poisson equation of classical gravitation theory describes how gravitating matter generates a gravitational potential. This potential can then be related to the gravitational field and to the force acting on material particles exposed to it. The Poisson equation is

$$\Delta\varphi = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right)\varphi = 4\pi\kappa\rho \tag{1}$$

where the gravitational potential is denoted by $\varphi = \varphi(x, y, z)$, which is a function of spatial coordinates x, y, z, where $\rho = \rho(x, y, z)$ denotes the density of gravitating matter, and where κ is a constant. Δ is a linear second-order differential operator, known as the Laplace operator.

The gravitational interaction between material bodies in classical physics can, of course, also be treated directly on the basis of Newton's law of gravitation. This law states that an attractive force between two point particles acts instantaneously along the direction defined by the two bodies and its strength varies inversely proportional to the squared distance between the particles. This action-at-a-distance force can also be calculated from a local potential function φ which is then determined by the Poisson equation introduced above.[27] While the Poisson equation thus appears only as an alternate description of the same physical content as Newton's law, this equation suggests, at the same time, a different physical interpretation of gravitation. According to this interpretation, gravitation—represented by the potential φ and produced by some matter distribution ρ which acts as its source—fills the entire space and exerts its influence on matter locally as a force. By virtue of this interpretation, the Poisson equation can be considered as a first hint at a gravitational field theory, in particular at a time when the field theoretic framework established by Maxwell's electrodynamics

27 Recall (see note 11) that we are loosely referring to the Poisson equation as a "field equation" even though it should properly be called a "potential equation."

suggested a field-theoretic revision of Newtonian gravitation. Nevertheless, the Newtonian gravitational potential lacks two essential features required by a genuine physical field theory. First, the gravitational field does not propagate with a limited speed, a field-theoretical feature that became mandatory after the advent of the theory of special relativity. It also does not describe some expected dynamical effects of gravitation such as dragging effects due to moving masses ("gravitational induction") or gravitational waves.

The Einstein equation stands at the end of a historical process in which the wish to conceive of the gravitational interaction in a truly field-theoretic manner played a significant heuristic role. The Poisson equation and the Einstein equation share a number of common features, in spite of the long and sometimes circuitous discovery process separating the two. In general relativity, the gravitational interaction is also determined by a second-order partial differential equation —the Einstein equation— which relates the gravitational potential to its source.

The Einstein field equation

$$G^{\mu\nu} = R^{\mu\nu} - \frac{1}{2} g^{\mu\nu} R = \bar{\kappa} T^{\mu\nu} \tag{2}$$

written in terms of the Ricci tensor $R^{\mu\nu}$ and the Riemann scalar R can also be written explicitly as (where summation over repeated indices is understood)

$$\begin{aligned} G^{\mu\nu} = \sum_{i\kappa} \Big(g^{\mu i} g^{\nu\kappa} - \frac{1}{2} g^{\mu\nu} g^{i\kappa} \Big) \\ \Bigg\{ \frac{1}{2} g^{lm} \Bigg(\frac{\partial^2 g_{im}}{\partial x^\kappa \partial x^l} + \frac{\partial^2 g_{\kappa m}}{\partial x^i \partial x^l} - \frac{\partial^2 g_{i\kappa}}{\partial x^m \partial x^l} - \frac{\partial^2 g_{lm}}{\partial x^i \partial x^\kappa} \Bigg) \\ + \frac{1}{2} \frac{\partial g^{lm}}{\partial x^l} \Bigg(\frac{\partial g_{im}}{\partial x^\kappa} + \frac{\partial g_{\kappa m}}{\partial x^i} - \frac{\partial g_{i\kappa}}{\partial x^m} \Bigg) - \frac{1}{2} \frac{\partial g^{lm}}{\partial x^\kappa} \frac{\partial g_{lm}}{\partial x^i} \\ - \frac{1}{4} g^{jm} g^{ln} \Bigg[\Bigg(\frac{\partial g_{ml}}{\partial x^i} + \frac{\partial g_{im}}{\partial x^l} - \frac{\partial g_{li}}{\partial x^m} \Bigg) \Bigg(\frac{\partial g_{jn}}{\partial x^\kappa} + \frac{\partial g_{\kappa n}}{\partial x^j} - \frac{\partial g_{\kappa j}}{\partial x^n} \Bigg) \\ - \Bigg(\frac{\partial g_{im}}{\partial x^\kappa} + \frac{\partial g_{\kappa m}}{\partial x^j} - \frac{\partial g_{i\kappa}}{\partial x^m} \Bigg) \frac{\partial g_{ln}}{\partial x^j} \Bigg] \Bigg\} = \bar{\kappa} T^{\mu\nu} \end{aligned} \tag{3}$$

where the gravitational potentials are denoted by $g_{\mu\nu}$ and $g^{\mu\nu}$, which are functions of the spacetime coordinates x_i, $(i) = 1, 2, 3, 4$, and where $\bar{\kappa} = -\frac{8\pi\kappa}{c^4}$ is a constant.

Like the left-hand side of the Poisson equation, the Einstein tensor $G^{\mu\nu}$ is a second-order differential operator applied to the gravitational potential, even though the operator in this case is much more complicated than the Laplace operator. $T^{\mu\nu}$ denotes the so-called stress-energy or energy-momentum tensor and corresponds to another element familiar from the Poisson equation, the role of matter as gravitating source. The mass density ρ which functions as the gravitating source in the Poisson equation reappears, for instance, in the following example of an energy-momentum tensor:

$$T^{\mu\nu} = \rho \frac{dx^{\mu}}{ds}\frac{dx^{\nu}}{ds}, \tag{4}$$

which describes a dust-like cloud of material particles acting as the source of the gravitational field where ρ stands for the mass-density of the swarm and dx^{μ}/ds denotes the special-relativistic four-velocity of the dust particles.

2.2 The Lorentz Model of a Field Equation

In our introduction of the Poisson equation as the point of departure in classical physics for a development eventually leading to the Einstein equations, we have emphasized their common features. One such basic feature is that both equations establish a relation between matter and gravitational potential; a second feature is that both equations relate the action of gravitation to its source by second-order partial differential equations. Such common features are more than distant mathematical similarities or analogies perceived only in hindsight. We claim that such similarities guided the historical development linking the two equations. These similarities, we believe, correspond to structural properties following from the basic mental model shaping the thinking process connected with this development. This interpretation is corroborated by the historical observation that the development from the Poisson to the Einstein equation went through a number of intermediate field equations of the same fundamental structure. We will show that they can all be interpreted as instantiations of the mental model of a field equation, which was modified, again and again, in response to inconsistencies by replacing a minimal number of specific features, while all other components retained their "default" settings. In spite of the inherently conservative structure of this development its outcome entailed fundamental changes in the conceptual structure of classical physics including the original mental model of a gravitational field equation itself.

We will write the basic structure of the mental model of a field equation implemented in the context of gravitational theory symbolically as:

OP(POT) = SOURCE. **(I)**

This equation is meant to symbolize a structure of shared physical knowledge according to which a source **SOURCE** generates a potential **POT**, related to each other by a differential equation with a second-order differential operator **OP** acting on the potential. We justify the introduction of our symbolic notation by the observation that the same knowledge structure can be found in such different cases as the Poisson equation, Einstein's intermediate equations for the gravitational potential, the Laplace equation for the electrostatic potential, and the four-dimensional potential formulation of Maxwell's equations. Correspondingly, **OP**, **POT**, and **SOURCE** can be instantiated in many different ways, such as the Laplace or the d'Alembert operators for **OP**, mass density or the energy-momentum tensor for **SOURCE**, Newton's gravitational potential or the metric tensor for **POT**. Notwithstanding the different contexts for each of those instantiations, we find an overarching conceptual structure relevant for each of them. It is the role of these overarching structures in guiding physical reasoning in a qualitative way that we wish to describe in terms of mental models and frames and that we wish to capture in our symbolic notation. We discuss the relevant instantiations for the frame of the field equation in somewhat greater detail.

Before the crucial phase of Einstein's search for a gravitational field equation in the years 1912 – 1915, the mental model of a field equation essentially covered two physical structures, that shaped Einstein's conceptual background in his search: the Poisson equation of classical mechanics and the potential equations of electrodynamics. In the latter, the structure even appears twice, once in electrostatics, in a simple form analogous to that in classical mechanics, and once in a more complex version extended to cover the dynamical aspects of the electromagnetic field as well. In electrostatics, the electrostatic potential φ_e is generated by an electric charge density ρ_e according to[28]

$$\Delta\varphi_e = -4\pi\rho_e. \tag{5}$$

The more extended version, which covers this equation as a special case under certain conditions, is the four-dimensional potential formulation of Maxwell's equations at the core of classical electrodynamics. The four-dimensional, special-relativistic formulation of electrodynamics was developed beginning in 1908 by Minkowski, Laue, and Sommerfeld[29] and was quickly established as a standard.[30] In this framework, the inhomogeneous Maxwell equations can also be written in a potential formulation:[31]

$$\Box\phi^{\mu} = \left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}-\frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\phi^{\mu} = -\frac{4\pi}{c}j^{\mu} \tag{6}$$

---

28 The minus sign which does not appear in the Poisson equation of classical mechanics given in eq. (1) reflects the fact that the gravitational interaction is attractive whereas the electrostatic interaction of two charges of equal sign is repulsive.

29 See (Minkowski 1908, Laue 1911, and Sommerfeld 1910a; 1910b).

30 For historical studies, see (Reich 1994, Walter 1999).

31 See, e.g., (Laue 1911; 1913 § 19). The potential formulation of Maxwell's equation given in eq. (6) presupposes a gauge fixing of the form

$$\partial_{\mu}\varphi^{\mu} = \mathrm{div}\boldsymbol{A} + \frac{1}{c}\frac{\partial\varphi_e}{\partial t} = 0$$

(Lorentz gauge). Together with this gauge condition eq. (6) represents a fully equivalent representation of Maxwell's equations.

where $\Box$ is the d'Alembert operator, $\phi^{\mu} = (\varphi_e, \boldsymbol{A})$ the electromagnetic four-potential composed of a scalar electric potential φ_e, and a vector magnetic potential $\boldsymbol{A} = (A^x, A^y, A^z)$, and $j^{\mu} = (\rho_e c, \rho_e \boldsymbol{v})$ is the four-current, composed of the electric charge density ρ_e and the velocity vector $\boldsymbol{v} = (v^x, v^y, v^z)$ acting as a source of the potential.

As we shall discuss in more detail below,[32] the relation between electrostatics and electrodynamics provided Einstein and his contemporaries with a basis for an understanding of how Newton's theory of gravitation might be elaborated into a field theory satisfying the requirements of the relativity theory of 1905. Einstein explicitly compared the task of building a relativistic theory of gravitation to the task of developing the entire theory of electromagnetism knowing only Coulomb's law, and found it just as formidable.[33]

Concrete instantiations of the general structure (I) make it clear that there are profound differences between them that are not represented by the simple symbolic equation. A major difference between the gravitational or electrostatic Poisson equation (1) resp. (5) and the full electrodynamic wave equation (6) concerns, for instance, the behavior of the equations under coordinate transformations. The simple mathematical form of those two equations is valid only if specific systems of coordinates are used. The same equations rewritten for a different coordinate system would, in general, change their appearance, unless the new system of coordinates is related to the old one by a coordinate transformation of the appropriate covariance group. This group of admissible coordinate transformations is a mathematical feature of the equation, that is, of the differential operator as well as of the source-term appearing in the equation. It also expresses the validity of a relativity principle for the relevant physical theories, be they those of classical or special-relativistic physics. Coordinate systems can be associated with observers in different locations and in different states of motion, and covariance with regard to coordinate transformations can be associated with the independence of physical phenomena of the perspectives of these different observers.[34]

The Laplace operator, appearing in the electrostatic as well as in the gravitational Poisson equation, retains its form if Galilean coordinate transformations are used

32 See section on "correspondence principle," p. 148.

33 See (Einstein 1913). For more evidence that Einstein conceived the problem of gravitation in analogy with electrodynamics, consider e.g. the title of (Einstein 1912a): "Is there a Gravitational Effect Which Is Analogous to Electrodynamic Induction?" ("Gibt es eine Gravitationswirkung, die der elektrodynamischen Induktionswirkung analog ist?") or a number of references to the analogy with electrodynamics in Einstein's contemporary correspondence; see, e.g., Einstein to Paul Ehrenfest, before 20 June 1912: "A rotating ring does not generate a static field in this sense, even though it is a temporally invariant field. [...] My case corresponds to the electrostatic field in the theory of electricity, whereas the more general static case would also include the analogue of the static magnetic field. I haven't got as far as that yet." ("Ein sich drehender Ring erzeugt nicht ein statisches Feld in diesem Sinne, obwohl es ein zeitlich unveränderliches Feld ist. [...] Mein Fall entspricht in der Elektrizitätstheorie dem elektrostatischen Felde, wogegen der allgemeinere statische Fall noch das Analogon des statischen Magnetfeldes mit einschliessen würde. So weit bin ich noch nicht.") (CPAE 5, Doc. 409).

which relate the inertial reference frames of classical mechanics to each other; these inertial transformations express a symmetry of Newtonian spacetime. The d'Alembertian operator appearing in special relativistic electrodynamics, on the other hand, is invariant under the Lorentz transformations which relate inertial reference frames of special relativity to each other and express a symmetry of four-dimensional Minkowski spacetime.

This illustrates that there can still be profound differences between various instantiations of the structure (I). But whatever these differences may be, insofar as the relation between **OP**, **POT**, and **SOURCE** is cancelled, the corresponding frames enter the same network of relatively stable relations to other physical concepts such as field and force. The concept of a field, in particular, is related to the concept of potential appearing in this mental model by a structure according to which a field **FIELD** is derived from a potential **POT** by some differential operation **GRAD**. This relation can be written in symbolic notation as

$$\textbf{FIELD} = -\,\textbf{GRAD(POT)}. \qquad \textbf{(II)}$$

In classical mechanics the equation relating the gravitational potential φ to the gravitational field $\boldsymbol{g}$ is given by

$$\boldsymbol{g} = -\mathrm{grad}\varphi. \tag{7}$$

In electrostatics a similar equation holds for the electric field $\boldsymbol{E}$ derived from the electrostatic potential φ_e by

$$\boldsymbol{E} = -\mathrm{grad}\varphi_e. \tag{8}$$

In the case of electrodynamics the same structural relation reappears, albeit in a somewhat more complex form. The components of φ_e and $\boldsymbol{A}$ of the four-potential φ^μ are related to the electric field $\boldsymbol{E}$ and the magnetic field $\boldsymbol{B}$ by

$$\boldsymbol{E} = -\,\mathrm{grad}\varphi - \frac{d}{dt}\boldsymbol{A}, \tag{9}$$

and

$$\boldsymbol{B} = \mathrm{curl}\boldsymbol{A}. \tag{10}$$

These equations can be combined in a tensor equation for the electromagnetic field tensor $F_{\mu\nu}$

34 We add a note of caution here. From the point of view of modern coordinate-free descriptions of physical theories, the covariance of a particular equation under a specific group of coordinate transformations can be understood as expressing a symmetry property of the underlying spacetime manifold, if the coordinate systems related by the transformations are those associated with so-called geodesic observers. In this case, the geodesic observer field is also a Killing vector field of the manifold, see (O'Neill 1983, 358–362). See also the discussion in (Salmon et al. 1999, chap. 5) and (Norton 1992b). At the time, however, the validity of a physical principle of relativity was directly associated with the covariance of the corresponding equations under coordinate transformations, without considering the symmetry properties of a manifold independently from its coordinate representation.

$$F_{\mu\nu} = \frac{\partial\phi_\mu}{\partial x^\nu} - \frac{\partial\phi_\nu}{\partial x^\mu}, \tag{11}$$

which may be considered as another instantiation of (II).

The discussion of the foregoing examples should make it clear that any concrete meaning of our symbolic equations is context-dependent as is only fitting for relations between frames in the sense introduced above. Entities such as **OP**, **POT**, **SOURCE**, **FIELD**, **GRAD** but also operations such as multiplication can take on entirely different mathematical meanings in different contexts. These symbolic operators may inherit different default-settings from different frameworks of reasoning. There is no *a priori* guarantee that the resulting concrete expressions can still be subsumed under one overarching theory. If one looks, however, at the function of these frames as heuristic devices that guided Einstein's pathway out of classical physics, this obviously was precisely their strength.

In summary, the stability of the mental model of a field equation is a consequence of its embedding in a network of physical concepts covering a broad spectrum of physical knowledge. Specifically, the concepts of potential and mass have stable relations to such concepts as field, force, energy, momentum, and motion. Furthermore, all instantiations of the Lorentz model we have encountered in classical and special-relativistic physics include a second-order differential operator **OP** and can be characterized by symbolic relations between the associated physical concepts such as (II). For each instantiation, the mental model of a field equation acquires local stability also through the representation in terms of mathematical concepts that in themselves are interconnected in an elaborated network allowing for formal manipulations of the mathematical expressions using well-known formal rules.

The various slots in our symbolic equations can be filled with objects of very different mathematical character, **POT** and **SOURCE** may be instantiated by scalar or vectorial objects, which behave differently under coordinate transformations; the corresponding differential operator **OP** may be the Laplacian or the d'Alembertian operator. Physically, potential and mass enter the stable conceptual relation described above, but are at the same time connected with quite different physical concepts and hence quite different physical phenomena. Thus, the potential **POT** could be instantiated to the potential of gravitational, electrostatic, or electrodynamic interaction, and the source-term **SOURCE** could be gravitating mass-density, electric charge-density, or electric current.

The Einstein equation introduced in the beginning of this section emerged, as we shall see, in a process that started from the Poisson equation of classical mechanics and proceeded via intermediate field equations that are *all* structured by what we have called the Lorentz model of a field equation. It is therefore no accident that the Einstein equation also displays features of this model. We shall show that the Einstein equation came about only as the result of a complicated process of adaptations of the original mental model demanding a number of variations ("changes of default settings") that at each step had to fulfill different, and often conflicting requirements. A consistent solution for meeting those requirements was reached only with the final theory of general relativity. In this theory, however, the field equation has implications that, as we shall see, challenge the original mental model.

2.3 The Lorentz Model of an Equation of Motion

The field-theoretic model comprises not only a structure shaping the understanding of a field equation but also a scheme determining the meaning of an equation of motion. In classical physics a field equation must be complemented by an equation of motion. Their complementarity derives from the way in which interactions are split into cause and effect in the Lorentz model. In classical mechanics, the concept of force allows one to separate the generic features of the action of some agent, to be described in terms of a general force law, from its specific effect on a given physical object, to be described in terms of a change of its state of motion. A similar structure is characteristic of Maxwellian electrodynamics, especially in Lorentz's electron theory. The field equation describes how sources, represented in our symbolic equation by the **SOURCE**-frame, affect the state of the surrounding space, represented by the **POT**-frame or the **FIELD**-frame. The equation of motion describes the effect of the thus affected space on physical objects in it. From the perspective of classical mechanics, a field equation is therefore nothing but a specific way of prescribing a general force law. What is required is a bridge between the concept of field and that of force.

According to classical mechanics, the effect of a force is a deviation (to be observed within an inertial frame of reference) from a state of rest or a state of uniform rectilinear motion described in terms of an **ACCELERATION**-frame. The magnitude of the acceleration depends not only on the force (characterized in the following by a **FORCE**-frame) but also on the reactive properties of the physical object exposed to it; these properties will be summarily described by the inertial mass frame, $\mathbf{MASS_{IN}}$. In short, an equation of motion according to classical and special-relativistic physics, complies with a mental model of causation that may be called the "acceleration-implies-force model" and takes the form:

$$\mathbf{FORCE = MASS_{IN}\ x\ ACCELERATION.} \qquad \mathbf{(III)}$$

In classical mechanics this relation corresponds to Newton's

$$\boldsymbol{F} = m \cdot \boldsymbol{a}, \tag{12}$$

where m is the inertial mass of a material particle, a its acceleration in three-space and F a classical force. The special relativistic generalization of this relation is

$$F^{\mu} = m\frac{du^{\mu}}{ds}, \tag{13}$$

where m is the rest mass, u^{μ} the four-velocity and s the proper time. Here F^{μ} denotes the force as a four-vector.

The structure of the symbolic equation (III) also complies with that of a much more general and much older mental model of causation rooted in intuitive physics, the "force-implies-motion model," which thus serves as a "higher-order model" for the Newtonian relation (III):[35] According to this higher-order model of causation, the effect of an action (here **ACCELERATION**), depends on the strength of the action (here **FORCE**) as well as on the resistance to the action (here $\mathbf{MASS_{IN}}$).

How can the acceleration-implies-force model belonging to the core of Newtonian mechanics be integrated with the concept of field at the center of the field-theoretical model? In order to bridge the two models one needs a specification of the relation between **POT** or **FIELD**, describing the local state of the surrounding space, and **FORCE**, describing the role of this space as an agent determining the motion of matter. In classical field theory, this bridge relation is given by the notion that the field is tantamount to a local force. The force experienced by a particle in a field is proportional to the strength of the field at the point of the particle in space and time. It is also proportional to that quality of the particle that responds to the particular field, be it its gravitational mass, its electric charge, or its magnetic moment. We capture the relation between **FORCE** and **FIELD** by the symbolic relation

$$\mathbf{FORCE = CHARGE\ x\ FIELD} \qquad \mathbf{(IV)}$$

At this point, our symbolic relations allow us to describe a possible inference on the level of qualitative physical reasoning. We may use relation (II) between **FORCE** and **POT**, to derive a relation between **FIELD** and **POT**

$$\mathbf{FORCE = - CHARGE\ x\ GRAD(POT).} \qquad \mathbf{(V)}$$

From a different perspective, one may also look at the set of relations (II), (IV), and (V) as the expression for a conceptual network on the level of qualitative physical reasoning in which the frames **FORCE**, **FIELD**, **CHARGE** and **POT** are related to each other.

In electrostatics the **CHARGE**-frame is instantiated by the charge density ρ_e ,

$$\mathbf{CHARGE} \xrightarrow{\text{electrostatics}} \rho_e\,, \qquad \mathbf{(VI)}$$

and the force density acting on ρ_e and determined by an electric field $\boldsymbol{E}$ derived from the electrostatic potential φ_e is given by:

$$\boldsymbol{F}_e = \rho_e \boldsymbol{E} = -\rho_e \mathrm{grad}\varphi_e. \qquad (14)$$

In Newtonian gravitational theory the default setting of the **CHARGE** frame is the so-called "passive gravitational mass":

35 For the force-implies-motion model, see (Gentner and Stevens 1983; Renn 2000), and also "Classical Physics in Disarray …" (in this volume).

$$\textbf{CHARGE} \xrightarrow[\text{gravitation}]{\text{Newtonian}} m_p \text{ resp. } \rho_p . \qquad \textbf{(VII)}$$

Accordingly, the force density acting on a mass density ρ due to a gravitational field $\boldsymbol{g}$ that can be derived from a gravitational potential φ is given by:

$$\boldsymbol{F} = \rho \boldsymbol{g} = -\rho \operatorname{grad}\varphi . \qquad (15)$$

In the case of electrodynamics the same structural relation holds in terms of the electromagnetic field tensor $F^{\mu\nu}$ expressed in terms of a generalized electrodynamic potential in (11). The four-force density K^{μ} is given by

$$K^{\mu} = j_{\nu} F^{\mu\nu} . \qquad (16)$$

This equation again exhibits the structure **FORCE = CHARGE x FIELD**, even though the multiplication of our symbolic equation is realized in this case by a four-dimensional contraction.

The discussion of the bridge relation required to integrate the acceleration-implies-force model with the field concept makes the intrinsic complexity of the Lorentz model particularly evident. This complexity stands in striking contrast to certain elementary features of gravitational interactions. It is mainly due to the fact that the Lorentz model results from the integration of mental models referring to two kinds of physical substances, the model of an extended, space-filling physical medium traditionally labelled as "aether" and the model of matter constituted by particles. The relation between field and force given by (IV) mediates between these models and at the same time points to the conceptual intricacies resulting from their integration. For instance, what at first sight merely seems to be a problem of two bodies moving about their common center of gravity, say of the sun and a planet, appears, from the perspective of the Lorentz model as the consequence of a field generated by one body which is then felt by the other body as a force that in turn is the cause of its motion.

As a consequence of this construction, both the concept of force and the concept of mass take on connotations, which they did not possess independently in the more elementary models. In classical mechanics, for instance, the concept of force comprises actions at a distance, typically between particles. In the context of the field-theoretical model, it applies exclusively to local interactions, a rather artificial limitation from the point of view of Newtonian physics. Similarly, while the Newtonian concept of force entails a reciprocity of the interaction it describes, expressed in Newton's *actio = reactio*, such a reciprocity is less evident for an interaction that is conceived to relate a state of space, characterized by the **FIELD**-frame or the **POT**-frame, to changes of the state of motion of a physical object, characterized by the **ACCELERATION**-frame. As a matter of fact, theories such as Lorentz's electron theory violate this reciprocity and *actio = reactio* no longer holds for the interaction between ether and charged matter.

The conceptual intricacies implied by the Lorentz model for the concept of mass are even more serious. Mass may be conceived as a "source" causing changes of the state of space or of the "aether" according to (**I**). We thus have "active gravitational mass" in the case of gravitational interaction:

$$\textbf{SOURCE} \xrightarrow{\text{gravitation}} m_a \text{ resp. } \rho_a. \qquad \textbf{(VIII)}$$

Mass may also, according to (V) and (VII), be conceived as a passive property of a physical object exposed to the resulting field, determining the degree to which the field locally acts as a force (**CHARGE** or "passive gravitational mass" in the case of gravitation); it may finally be conceived, according to (III), as "inertial mass" $\mathbf{MASS_{IN}}$, i.e., as resistance to **ACCELERATION**. In classical electrostatics, these magnitudes are represented by electrical charge and inertial mass, respectively, and can vary independently from each other. In classical gravitation theory, gravitational and inertial mass happen to coincide empirically. In this case we are thus entitled to introduce a generic **MASS**-frame for which we have:

$$\mathbf{MASS = MASS_{IN}}, \qquad \textbf{(IX)}$$

which may hence be instantiated by inertial, or active gravitational, or passive gravitational mass.

The integration of different mental models within the field-theoretical model produces conceptual distinctions that may actually not be warranted by the available knowledge of the interactions it describes. The emergence of conceptual distinctions as an artefact of a theoretical framework was visible, in the case of the gravitational interaction, even from a less sophisticated perspective than that offered by the field-theoretical model. When the gravitational action is described not in terms of a field theory but simply using the Newtonian force law, the distinction between mass as a property of matter that *causes* gravitation and mass as a reactive property of matter that *resists* the acceleration caused by a gravitational force is rather artificial. Indeed, it has long been known that all bodies fall with the same acceleration in a gravitational field whatever their mass (Galileo's principle). Within the context of the field-theoretical model this insight suggests far-going consequences for the understanding of an equation of motion.

In fact, since in classical mechanics the **CHARGE**-frame and the **SOURCE**-frame instantiate to the passive and active gravitational mass resp. mass density according to (VII) and (VIII), we may identify these two frames with each other and with the general **MASS**-frame:

$$\textbf{SOURCE = CHARGE = MASS} \qquad \textbf{(X)}$$

Recalling the relations that the **FORCE**-frame enters with the **ACCELERATION**-frame and the **FIELD**-frame according to (III) and (IV), our symbolic equations entail

$$\textbf{ACCELERATION = FIELD.} \qquad \textbf{(XI)}$$

This symbolic equation translates Galileo's principle to the assertion that in a gravitational field theory the local acceleration actually represents the gravitational field. This makes it possible to interpret the effect of gravitation, namely **ACCELERATION**, directly as a representation of the local force, i.e. as a **FIELD**, independently of the properties of the object exposed to it. We emphasize that we introduced our symbolic notation in order to be able to represent this kind of inference which can be made largely on the level of qualitative physical reasoning independent of any concrete representation. It also expresses the fact that the identification of the **ACCELERATION**-frame and the **FIELD**-frame is a general relation between two frames that is not tied to the concrete conceptualization of the gravitational interaction. It may hence guide the physical reasoning also in situations where new ways of mathematical representation or else new conceptual relations within a gravitational theory are being explored.

Eq. (XI) no longer contains **FORCE**. This insight crucial for the development of general relativity. It suggests that it should be possible to set up a theory where field phenomena are equivalent to acceleration phenomena. This, of course, is exactly the idea at the core of Einstein's equivalence principle.[36] In such a theory Galileo's principle would find the conceptual justification it lacked in classical mechanics, where it appeared as a mere empirical coincidence.

The insight that in a gravitational field theory the acceleration is directly equivalent to the field, symbolically represented by eq. (XI), also suggests the formulation of an equation of motion in a gravitational field that does not make use of the intermediate concept of force. The idea of eliminating the concept of force was familiar from classical physics and had been elaborated in the context of the Lagrange formalism of analytical mechanics. In elementary situations of classical mechanics the Langrangian or Lagrange function at the center of this formalism is simply the difference between the kinetic and the potential energy of a physical system:

$$L = T - V. \tag{17}$$

The Lagrange formalism provides an alternative way of obtaining equations of motion. In this formalism the trajectory of a material body is selected from the set of all kinematically possible trajectories satisfying given constraints. The criterion for the selection is that the action, defined as the same integral of the Langrangian along a given trajectory is stationary, i.e., takes on either a maximum or a minimum value, for the actual trajectory. This criterion is known as Hamilton's principle. Saying that the action is stationary is the same as saying that its variation vanishes:

36 See the discussion below. Cf. in this context Einstein's use of the word "Beschleunigungsfeld" (acceleration field) in (Einstein 1912b): "the hypothesis that the "acceleration field" is a special case of the gravitational field [...]" ("die Hypothese, daß das 'Beschleunigungsfeld' ein Spezialfall des Gravitationsfeldes sei [...]" (p. 355).

$$\delta\left\{\int L dt\right\} = 0. \tag{18}$$

The situation is similar to the problem of finding the shortest path connecting two points on a curved surface. This problem can be solved by looking for an extremal value among the lengths of all possible paths connecting these points.

The Lagrange formalism yields the explicit equation of motion for a particle in the form of the so-called Euler-Lagrange equations, which follow from Hamilton's principle:

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}_i}\right) - \frac{\partial L}{\partial x_i} = 0. \tag{19}$$

Under appropriate circumstances, these equations may be assimilated to the relation

$$\boldsymbol{F} = d\boldsymbol{p}/dt \tag{20}$$

between a force F and the change of momentum p familiar from classical mechanics. We capture the qualitative physical content of equations (19) and (20) by the symbolic equation

DIFF(MOMENTUM) – FORCE = 0, **(XII)**

introducing, at the same time, a **MOMENTUM**-frame and a **DIFF**-frame, the latter being in the present case instantiated by time-derivatives. The advantage of the Lagrange formalism compared to the explicit specification and use of forces becomes clear if one considers motion under geometrical constraints, such as the motion of a point particle on the surface of a sphere. Using the acceleration-implies-force model such geometrical constraints are realized by constraining forces which are defined only by their effect, i.e., if a body moving under geometric constraints departs from uniform rectilinear (inertial) motion this deviation is assumed to be caused by the constraining forces. The precise magnitude and direction of these forces are generally unknown and hence cannot be explicitly specified in order to obtain an equation of motion by instantiating the acceleration-implies-force model. The Lagrangian formalism uses the fact that these constraining forces do not perform work. They play no role in the interplay between kinetic and potential energy as captured by the Lagrange function if the latter is expressed solely in terms of those generalized coordinates that describe possible motions under the given constraints, without losing information about the physical situation.

The significance of the Lagrange formalism for Einstein's research on gravitation was twofold. For one, it represented a generalizable formalism that was applicable in cases where a force or a momentum was not easily identified. It therefore also applied in a geometrized theory by expressing the Lagrange function in terms of a geometry adapted to the physical situation at hand.[37] Describing motion in a gravitational field with the help of the Lagrange formalism, as we shall see, naturally suggests to conceive of gravitation as a consequence of geometry, rather than of force. And this conception of gravitation in turn suggests adopting the Lagrange formalism as the natural framework for formulating the equation of motion in a gravitational field.

3. THE ELEMENTS OF EINSTEIN'S HEURISTICS

In the course of the first period of Einstein's research on gravitation—between 1907 and 1912—specific components of the heuristics had crystallized as relatively stable structures which would guide his search for the gravitational field equation in the second period—documented in the Zurich Notebook. To some of them Einstein even attached labels, such as the "equivalence hypothesis,"[38] making their outstanding role for his heuristics evident, while other heuristic requirements were so obviously interwoven with canonical expectations from classical physics, such as the requirement of energy conservation, that they did not receive a special name. For ease of reference, we shall nevertheless introduce standard names.

Einstein's *equivalence principle*, which in the first period mainly served to find properties of special cases of gravitational fields, became in the second period a standard criterion for checking whether or not candidates for a general gravitational field equation incorporated his earlier insights about the intimate relation between gravitation and inertia. The *generalized relativity principle*, a closely related result of Einstein's research in the first period, was applied in the second period either as a starting point in the context of the mathematical strategy for choosing appropriate candidates for the gravitational field equation or as a validation criterion by which a candidate constructed in the context of the physical strategy was examined. The *conservation principle*, inherited from classical physics, played a crucial role in developing the theory of the static gravitational field and was similarly used in the second period both as a touch stone and as a building block. This was also the case for what we will call the *correspondence principle*. This principle represents the demand to incorporate in a new theory of gravitation the knowledge about Newtonian gravitation by requiring that the basic relations of the latter be recovered from the former in some approximation or as some special case. Its implementation as a component of Einstein's heuristics took clues from the relation between electrostatics and electrodynamics. Einstein thus expected that the generalized theory should be connected to the Newtonian theory via the intermediate case of the weak and static gravitational field.

---

37 Its significance for expressing the equation of motion in special relativity was realized by Max Planck (Planck 1906, 1907).

38 Although Einstein referred to the equivalence principle as a "hypothesis" in (Einstein 1907) and in (Einstein 1911), the terms "Aequivalenzhypothese" and "Aequivalenzprinzip" were used for the first time in (Einstein 1912b).

3.1 The Equivalence Principle and the Generalized Relativity Principle

According to Einstein's equivalence principle the effects of a homogeneous static gravitational field are equivalent to those in a uniformly and linearly accelerated reference frame. The equivalence principle, which establishes a connection between the gravitational field and inertial forces, is closely related to Galileo's principle that all bodies fall with the same acceleration in a gravitational field, independent of their constitution. While neither Galileo's principle nor the equivalence principle are part of the foundational structure of classical physics, they are part of the knowledge contained in it, as expressed by our symbolic equation (XI). Einstein established a meaningful connection between acceleration and the gravitational field by integrating two mental models of classical physics which originally belonged to different domains of knowledge, the mental model of a system with a homogeneous static gravitational field, familiar from everyday physics in local terrestrial laboratories, and the model of a system in uniformly accelerated motion (Einstein's famous elevator experiment), which was analyzable using standard tools of classical mechanics. The indistinguishability of motions in these two systems makes it possible to identify the terminals of these models and thus to establish an equivalence between gravitational and inertial forces as well as between an accelerated frame of reference and an inertial frame. These identifications turned out to have far-going consequences for the organization of physical knowledge. Such consequences can be spelled out if further elements of the knowledge of classical and special relativistic physics are taken into account and are combined, for instance, with simple mental models of ray optics leading to the conclusion that light is curved in a gravitational field.[39]

Einstein's "elevator model," admits an extension to a more general class of gravitational fields and accelerated motions. Such an extension was suggested, in particular, by the Machian idea to interpret the inertial forces occurring within a uniformly rotating system as due to the interaction with distant masses rather than due to "absolute space." Mach had compared an accelerated system—Newton's famous rotating bucket—with a system at rest in which an interaction with distant masses, the stars revolving around the bucket, accounts for the same physical phenomena as are produced by the inertial forces in the accelerated system. This thought experiment provided a blueprint for the elevator-thought-experiment, which is at the heart of Einstein's "principle of equivalence." In analogy to the "elevator model," a "bucket model" could thus be conceived as one in which the inertial forces occurring in a rotating reference frame are interpreted as the effects of a generalized gravitational field.

The elevator and the bucket models may both be considered as special cases of a general "gravito-inertial model" in which inertial forces resulting from arbitrarily accelerated motions are interpreted as coming from a "dynamic" gravitational field. This gravito-inertial model made it plausible to assume that inertial frames of reference play no privileged role in a theory that adequately describes such a generalized

39 See the discussion in "Classical Physics in Disarray …" (in this volume).

gravitational field. It also suggested that the generic properties of gravitational fields that can be thought of as resulting from accelerated motions are shared by arbitrary gravitational fields. It suggests, for instance, that the laws governing the motion of bodies are the same in both types of fields. More generally, the gravito-inertial model made it plausible that physical interactions taking place in a gravitational field are essentially equivalent to those taking place in a gravitation-free system that is described from the point of view of an accelerated observer. In hindsight, the equivalence principle—and the gravito-inertial model structuring the reasoning on which this principle is based—thus introduced four more or less distinct requirements into the search for a theory of general relativity:

– the theory should satisfy a "generalized principle of relativity" and eliminate as much as possible the privileged *a priori* structures which in the classical theory are associated with such notions as absolute space and inertial frames of reference;
– the theory should describe motion in a gravitational field as a "free fall" independent of the structure of the moving body;
– the theory should treat gravitation and inertia as aspects of one more general interaction; and
– the theory should describe non-gravitational physical interactions essentially in the same way as special relativity if an appropriate reference frame (local inertial frame) is chosen for that description.

These requirements are directly related to general relativity as we know it today. Historically, the impact of the equivalence principle on the search for a new theory of gravitation was much less straightforward than it may appear in hindsight. A number of conceptual and technical problems had to be resolved or at least disentangled before such a clear relation could emerge.[40] In particular, Einstein was convinced that the demand for a generalized relativity principle could be satisfied by requiring the equations of his theory to be generally covariant (Norton 1994, 1999). He lacked the modern notion of spacetime symmetries. Similarly, the description of motion in a gravitational field as "free fall" along a geodesic trajectory is closely related today to the understanding of the affine structure of spacetime. But Einstein did not have the concept of affine connection at his disposal and still saw the need to interpret the equation of motion in terms of a classical gravitational force.[41]

The equivalence principle and the generalized relativity principle did not give rise to requirements which the new theory had to satisfy as a set of fixed axioms; they acted in a more general and diffuse way as heuristic guiding principles which, in different contexts, had a variety of concrete implications not necessarily covered by their modern counterparts. The generalized principle of relativity, in particular, motivated Einstein to consider the absolute differential calculus as the appropriate lan-

40 See (Norton 1985), "Classical Physics in Disarray …" and "The First Two Acts" (both in this volume).
41 See "The Story of Newstein …" (in vol. 4 of this series).

guage for his new theory of gravitation but also to construct mathematical objects which are covariant merely under much more limited classes of coordinate transformations. The equivalence principle led him to identify qualitative consequences of general relativity such as light deflection even before the formulation of the definitive theory, albeit with numerically different results. It led Einstein to adopt the geodesic equation of motion as the law of motion appropriate for general gravitational fields but also to systematically check whether candidate field equations are covariant at least under transformations to linearly accelerated systems and to uniformly rotating systems.

In the context of Einstein's systematic search for the gravitational field equation documented in the Zurich Notebook the adoption of the generalized relativity principles amounted to a check of the covariance properties of a candidate field equation. But even the way in which this check was implemented—by the introduction of a generally-covariant differential operator along the mathematical strategy or by explicitly checking the behavior of a candidate under coordinate transformations along the physical strategy—depended on the specific perspective guiding the implementation. At the beginning of Einstein's search it was not at all clear whether he would eventually succeed in finding a generally-covariant field equation of gravitation incorporating the equivalence principle. From the outset it was unclear whether the ambitious aim of a generalized relativity principle and perhaps even the equivalence principle would be realizable or whether these postulates had to be restricted or modified in order to be able to satisfy other requirements to be imposed on such a field equation, such as the conservation principle.

3.2 The Conservation Principle

According to the conservation principle as it functioned in Einstein's heuristics, it should be possible to establish a balance of energy and momentum in a gravitational field, resulting in a conservation law if all contributions to the balance, including that of the gravitational field itself, are taken into account. This expectation was motivated by the experience of classical physics where such a balance of energy and momentum could indeed be obtained for all physical processes if only appropriate concepts of energy and momentum were identified for all relevant subdomains, such as mechanics, thermodynamics, and electrodynamics. This expectation had been both amplified and modified by the advent of special relativity, and in particular that of special relativistic continuum physics, which had shown that several distinct conservation laws of classical physics, such as those of mass, energy, and momentum actually had to be integrated into a single all-encompassing conservation law referring to a complex new entity, the stress-energy or energy-momentum tensor. Against this background, the conservation principle, understood as part of the heritage of classical and special-relativistic physics, introduced three more or less distinct requirements into the search for a theory of general relativity:

– the theory should take into account the close relation between mass and energy established by special relativity and consider not just mass but more generally mass and energy as embodied in the energy-momentum tensor (or some entity derived from it) as the source of the gravitational field;
– the theory should contain some generalization of the special-relativistic law for the conservation of energy and momentum; and, in particular,
– the gravitational field equation should be compatible with this generalized requirement of energy and momentum conservation.

From the perspective of today's understanding of general relativity, these requirements considerably restrict the choice of an acceptable gravitational field equation. But historically, just as with the generalized relativity principle, Einstein's heuristic expectations could not simply be turned into iron-clad axioms for the formulation of his new theory. Precisely because the requirements listed above were rooted in the knowledge of classical and special-relativistic physics, they were still embedded in a conceptual framework that was eventually overturned by general relativity. Furthermore, there were, at the outset of his search, still numerous possibilities for instantiating the general relations suggested by Einstein's classical expectations. It was, for instance, conceivable that not the energy-momentum tensor itself but its trace acts as the source of the gravitational field. For some time, Einstein assumed that he had to find a generally-covariant energy-momentum tensor of the gravitational field in analogy to the one for matter, while such a tensor does not exist according to the final theory. He also assumed that the conservation principle would play the role of an additional postulate of the theory, whereas it is implied by the correct gravitational field equations. Such conceptual novelties of general relativity could not have been anticipated on the basis of the knowledge of classical physics informing Einstein's heuristics. They were the eventual outcome of his heuristic schemes in the course of concrete and often futile attempts to identify a gravitational field equation compatible with criteria such as the conservation principle.

The effect on Einstein's search of the requirements here summarized under the label "conservation principle" depended on the specific questions he pursued and on the level of sophistication of the techniques at his disposal. At one point he erroneously convinced himself, for instance, that a gravitational theory based on a single scalar potential was incompatible with the conservation principle but then had to retract that argument in the light of a closer analysis of such a scalar theory.[42] The clear-cut function which the conservation principle eventually assumed as a compatibility requirement for an acceptable field equation in his search for such an equation documented in the Zurich Notebook was the result of his learning experience with the theory for static gravitational fields in 1912.[43] This experience demonstrated to Einstein the crucial significance of the conservation principle for his search. He became aware step by step of the full scope of the network of relations it implies. In the

42 See (Norton 1992a).
43 See "The First Two Acts" (in this volume) and the discussion below.

course of his research documented in the Zurich Notebook, involving the mathematically much more complex tensorial formalism, these relations combined to form a set of standard expectations for a field equation that had to be systematically checked for each candidate. Only towards the very end of this phase of his research did Einstein recognize the possibility of turning this network into a recipe for constructing a gravitational field equation satisfying the conservation principle—albeit as a requirement essentially still conceived within a classical framework.

3.3 The Correspondence Principle

The correspondence principle requires that the new relativistic theory of gravitation incorporate the empirically well-founded knowledge about gravitation contained in the classical Newtonian theory. Ideally, it should be possible to obtain the Newtonian theory as a limiting or special case from the new theory under appropriate conditions, such as low velocities and weak fields.[44] In contrast to the generalized relativity principle of which it was not clear at the outset to what extent it could be implemented in the new theory, the correspondence principle was a much less negotiable, if not absolutely necessary requirement for any acceptable theory of gravitation. It also seemed clear from the beginning how this principle would have to be implemented in concrete attempts to create a relativistic theory of gravitation. The classical theory offered a model for a gravitational field equation, the Poisson equation, even if this model does not take into account the relativistic demand of a finite speed of propagation of the gravitational action as would a field equation based on the d'Alembertian operator as in (6). But the Poisson equation did not only serve as a model for the structure of the new field equation. Einstein also expected it to emerge from a limiting process by which a relativistic field equation should touch base, via the intermediate case of a special-relativistic field equations based on the d'Alembertian operator, with the classical Newtonian theory. Einstein's theory of the static gravitational field provided another such base-line. Since it represents an intermediate situation between the full relativistic theory and the Newtonian case, he expected that the general theory would, under appropriate limiting conditions, first reproduce the results of the static special case and then, under further constraints, those of the Newtonian theory. A relativistic theory with this limiting behavior clearly would cover the full range of physical knowledge covered by the more specialized theories. Since the constraints imposed by the correspondence principle were embodied not just in abstract requirements but in well-developed theories, it follows that this heuristic principle could act not only as a compatibility condition for an acceptable gravitational field equation but also as a starting point for its construction.

---

44 For a discussion of the Newtonian limit of general relativity from a modern point of view, see (Kuenzle 1976, Ehlers 1981, 1986). For a discussion of the relation between Newtonian gravitation theory and general relativity from an axiomatic point of view as a case of reduction, see (Scheibe 1997, 1999, esp. ch. VIII).

As with the other criteria, there is a perspective from which the correspondence principle, together with a few other conditions, singles out general relativity as the only acceptable solution to Einstein's problem. But as we saw with the other heuristic principles, this hindsight-perspective tends to obscure rather than clarify the actual role of the correspondence principle in the creation of general relativity. This process involved conceptual innovations that could not have been anticipated on the basis of classical physics. From hindsight, we would rather have to say it could not even be anticipated *that* (let alone *how*) the definitive solution of his problem would yield the Newtonian theory since the classical limit of the final theory—which in some sense must exist for the reasons pointed out above—might not resemble the familiar Newtonian formulation of the classical knowledge about gravitation. Vice versa, the classical expectations concerning the relation between Newtonian and relativistic theory might impose restrictions on the choice of admissible candidates that could effectively rule out a satisfactory realization of Einstein's other heuristic requirements, in particular, of the generalized principle of relativity. The dilemma, in short, was that the correspondence principle represented, in view of its roots in the classical knowledge about gravitation, the most weighty of Einstein's heuristic principles but also the one most likely to be entangled with physical assumptions that would have to be given up if the new, relativistic theory of gravitation were to challenge those classical roots.

This dilemma could hardly be avoided. At the beginning of his search, Einstein sought to extrapolate the classical knowledge about gravitation into the new territory of a relativistic field theory. From his perspective, that territory, fortunately, was mapped out nicely by the implications of the Lorentz model. As we have seen, his model also determined the conceptualization of the relation between a generic field theory and the special case of a static field. In Maxwell's theory of the electromagnetic field that relation was well understood, so it could serve as a guide for exploring the analogous relation in the case of the relativistic gravitational field. The mental model of a field theory and the knowledge of classical physics it incorporates had governed Einstein's seemingly inductive procedure all along in examining special cases such as that of the static field.[45] It was clear to him from the outset that the static gravitational field corresponds to the electrostatic field while the field of a rotating reference frame corresponds to the magnetostatic field. The theory of electromagnetism also suggested that and how a many-component tensorial object representing the field in general turns into a much simpler object for the special case of a static field, which can be derived from a scalar potential. The fact that in classical physics both the electrostatic and the gravitational potential are represented by a scalar potential lent support to the assumption that a reduction to a scalar potential also takes place in a relativistic theory of gravitation, at least in the limit of weak static fields. Although this assumption eventually turned out to be wrong, it was backed by a long tradition in classical field theory to which no alternative was known and it initially prevented Einstein from accepting the Einstein tensor as a viable candidate for the left-hand side of the field equation.

Just as Einstein's other heuristic principles the correspondence principle did not act as an isolated axiom which in the end turned out to be either compatible or not with general relativity as we know it today. It was not an isolated statement at all but part of a network of arguments, affecting his heuristics in the context of a variety of considerations. The correspondence principle comprised, in particular, the demands that:

– the differential operator on the left-hand side of the gravitational field equation should, for weak fields, reduce to the d'Alembertian operator as in (6);
– the field equation, for weak static fields, should reduce to the Poisson equation for the scalar potential of classical physics;
– the same scalar potential should determine the behavior of a particle in a gravitational field, via the equation of motion.

The correspondence principle was also subject to modifications as Einstein's experience with attempts to implement this principle in concrete candidate field equations grew. The paradoxical fluid yet firm character of Einstein's qualitative reasoning on the level of his heuristic principles, which we have tried to grasp by describing it in terms of mental models and frames, allows it to first exclude and then support the correct field equations of general relativity. The correspondence principle thus left room for learning experiences as when Einstein found out that it was possible to meet the requirements of this principle with the help of additional constraints on the choice of the coordinate system.

While the technicalities of its implementation were subject to reconsideration and improvement, the basic structure of Einstein's understanding of the correspondence principle was stabilized by a wider context of arguments rooted in classical physics.

---

45 Compare the following equations from Einstein's correspondence: "I finished the investigations on the statics of gravitation (point mechanics electromagnetics gravitostatics) and am very satisfied with them. I really believe that I discovered a piece of truth. Now I ponder the dynamic case, going again from the more special to the more general." ("Die Untersuchungen über die Statik der Gravitation (Punktmechanik Elektromagnetik Gravitostatik) sind fertig und befriedigen mich sehr. Ich glaube wirklich, ein Stück Wahrheit gefunden zu haben. Nun denke ich über den dynamischen Fall nach, auch wieder vom spezielleren zum Allgemeineren übergehend.") Einstein to Ehrenfest, 10. March 1912, (CPAE 5, Doc. 369); "Lately I have been working like mad on the gravitation problem. Now I have gotten to the stage where I am finished with the statics. I do not know anything yet about the dynamic field, that will come only now. [...] You see that I am still far from being able to conceive of rotation as rest! Each step is devilishly difficult, and what I have derived so far is certainly still the simplest of all." ("In der letzten Zeit arbeitete ich rasend am Gravitationsproblem. Nun ist es soweit, dass ich mit der Statik fertig bin. Von dem dynamischen Feld weiss ich noch gar nichts, das soll erst jetzt folgen. [...] Du siehst, dass ich noch weit davon entfernt bin, die Drehung als Ruhe auffassen zu können! Jeder Schritt ist verteufelt schwierig, und das bis jetzt abgeleitete gewiss noch das einfachste.") Einstein to Michele Besso, 26. March 1912, (CPAE 5, Doc. 377); "My case corresponds to the electrostatic field in the theory of electricity, whereas the more general static case would also include the analog of the static magnetic field." ("Mein Fall entspricht in der Elektrizitätstheorie dem elektrostatischen Felde, wogegen der allgemeinere statische Fall noch das Analogon des statischen Magnetfeldes mit einschliessen würde.") Albert Einstein to Paul Ehrenfest, Prague, before 20 June 1912, (CPAE 5, Doc. 409).

Precisely because of this wider context, the modifications of Einstein's understanding in the course of his search for the field equation had the potential of challenging not only the technical aspects but also the conceptual framework of his heuristics.

3.4 Einstein's Heuristic Principles and his Double Strategy

Einstein's heuristic principles, as we have seen, did not constitute a set of axioms from which a theory of gravitation could be derived in a straightforward way. These principles yielded both too much and too little knowledge to find a new theory of gravitation—too little, because they were not sufficient to determine the new theory uniquely, too much, because they imposed requirements on the new theory that could not be maintained all at once. As we mentioned in the introduction, these principles initially even acted as competing approaches toward a relativistic theory of gravitation. In addition, their interpretation in concrete attempts to realize such a theory depended on the specific formalism applied and on the form other requirements took within that formalism. In the course of Einstein's work documented in the Zurich Notebook, these principles nonetheless developed together to become elements of a heuristic double strategy. Earlier research on the problem of a relativistic theory of gravitation, Einstein's own as well as that of others, had not only suggested the mathematical tools to be employed but had also circumscribed the requirements such a theory had to satisfy. As a consequence, the problem of identifying an acceptable gravitational field equation had become the task of constructing, as if in a theoretical laboratory, a more or less well-defined but never-tried device from a set of given building blocks.

Each of Einstein's heuristic principles against which constructions would have to be checked could be used either as a construction principle or as a criterion for their validity. The sequence in which the heuristic principles were used essentially determined their function. The approach we have labelled the "physical strategy" starts from the correspondence principle, i.e., from a candidate field equation which by inspection is seen to yield the Newtonian limit in the expected way. Such a candidate field equation is thus firmly rooted in classical physics. Typically, only mathematical knowledge familiar from the context of classical and special-relativistic physics was used in its construction. The compatibility of such a "physical candidate" with other criteria was, as a rule, less obvious and needed to be checked explicitly. If the primary goal was to stay as close to the familiar territory of classical physics as possible, the first thing to check was the conservation principle, which could turn out to be satisfied, give rise to modifications, or lead to the rejection of the candidate altogether. If the candidate survived this test, it was to be explored to what extent it complied with the generalized relativity principle, i.e., under how broad a class of coordinate transformations it would retain its mathematical form. For candidates that were not generally covariant, it had to be determined under which class of transformations the candidate was covariant and whether or not the restriction of this class was acceptable on physical grounds. In particular, it made sense to check whether at least the situations at the core of the equivalence principle, i.e., transformations to reference frames in uniform linear and rotational accelerated motion, were included in this class.

The approach we have labelled the "mathematical strategy" starts from the generalized relativity principle, i.e. from a candidate field equation which by inspection is seen to be covariant under a broad enough class of coordinate transformations. Since such a general principle of relativity was not part of classical physics, it was much less obvious than in the case of the correspondence principle what "by inspection" meant in this case. The expert mathematical knowledge of the time, however, provided him with a certain reservoir of suitable objects. Their relation to any meaningful physics was much less obvious than for a candidate of the physical strategy. It had to be checked explicitly whether such a "mathematical candidate" could be brought into agreement with the requirements of the correspondence principle. Failure to comply with the correspondence principle could lead to immediate rejection of the candidate, or generate additional conditions amounting to a restriction of the relativity principle. It could even trigger the discovery of a new way to obtain the Newtonian limit. It could also suggest how a given candidate was to be modified in order to pass the test. The situation was similar for the conservation principle, which represented another necessary condition for a physically meaningful theory. Since both the correspondence and the conservation principles could neither be circumvented nor substantially weakened, they tended, in turn, to impose restrictions on the generalized relativity principle or suggest modifications of the candidates.

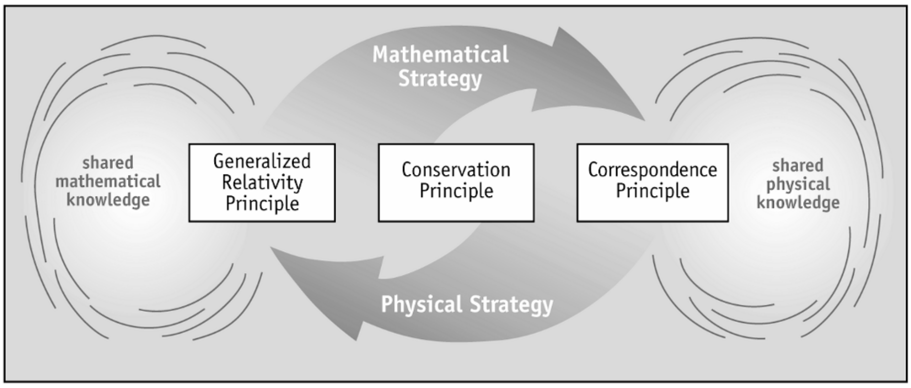


Figure 1: Einstein's double strategy arose from the different roles of the heuristic requirements of Generalized Relativity, Conservation, and Correspondence.

The two strategies are illustrated in Fig. 1 above. The physical and the mathematical strategy work with the same heuristic principles, draw on the same knowledge base of classical and special-relativistic physics, and essentially use the same mathematical representations. Why did they nevertheless produce, as we will see, different results in the course of Einstein's research? An answer is suggested by noting that the candidate solutions he examined are not determined directly by the two strategies but only through the concrete representations in which he attempted to embody his heuristic criteria. The two strategies did not act as an algorithm for producing solutions but rather as different channels for filling the Lorentz model with concrete mathematical and physical content. In other words, the two strategies constituted alternative ways for bringing to bear the available physical and mathematical knowledge on the problem of finding a gravitational field equation.

The notions of mental models and frames are helpful, we believe, for describing this process of knowledge assimilation. Because of its character as a mental model, the Lorentz model does not just represent an abstract scheme, but carries with it the experience of previous implementations. This prior experience includes the model's default settings enabling it to generate concrete candidate field equations even in the absence of sufficient knowledge about the properties of a relativistic gravitational field. The default settings make it possible to deal with the problem of insufficient knowledge by supplementing missing information drawn from prior experience. Moreover, since the experience of classical and special-relativistic physics entered the Lorentz model in the form of default settings, Einstein could give up prior assumptions in the course of his research without shattering his entire heuristic framework.

One and the same mental model may come with different sets of default settings, depending on prior experience, applications, knowledge resources, and higher-order models in which it is embedded. Default-settings depend on knowledge contexts. Classical field theory, the knowledge about Newtonian gravitation, the insights opened up by the elevator and the bucket models, the Machian interpretation of classical mechanics—all constitute different knowledge contexts relevant to the default assumptions of the Lorentz model when implemented in attempts to create a relativistic field theory of gravitation. The same is true for the mathematical resources of Gaussian surface theory, vector calculus, the theory of invariant forms, and the absolute differential calculus. Einstein's double strategy can be understood as a way of dealing with this problem of overabundant knowledge by consciously selecting alternate knowledge contexts dominating the default settings of the model. In this sense, the physical strategy, in particular, starts not just from the correspondence principle but from candidates embodying the classical knowledge about gravitation. The mathematical strategy likewise starts, not just from the generalized relativity principle but from candidates embodying the prior mathematical knowledge, in particular about generally-covariant, second-rank tensors of second order in the derivatives of the metric.

The selection of such different approaches dominating the default settings of the Lorentz model occurred initially, of course, in the hope that one or the other knowledge context would be more relevant or turn out to be more suitable to yield a full solution of the problem. Effectively, however, the alternation between different knowledge contexts led to a systematic exploration of resources that could not have been assimilated to the model all at once. The double strategy was not an astute plan for attacking the problem of finding field equations from two sides, the physical and the mathematical side. It emerged only gradually as a result of learning more about the implications of the field-theoretical model for gravitational field equations by varying its default settings. To understand Einstein's search for the gravitational field equation, it is therefore not enough to examine his heuristic principles as we have done in this chapter. We also have to reconstruct the default settings for the Lorentz model in different contexts of this search.

4. DEFAULT SETTINGS AND OPEN SLOTS IN THE LORENTZ MODEL FOR A GRAVITATIONAL FIELD EQUATION IN 1912

In this section we will introduce the principal entities figuring in Einstein's search for the gravitational field equation in the period documented by the Zurich Notebook. His research in this period focused on the problem of formulating a field equation for gravito-inertial phenomena, which had to satisfy all heuristic requirements, both those embodied in the mental model of a field equation and those that had emerged from his work between 1907 and 1912. The experience of these years had largely shaped the default assumptions that formed the starting point of Einstein's exploration of the mental model represented by the symbolic equation **OP(POT) = SOURCE**, cf. (I).

In particular, the metric tensor, which we will represent by the frame **METRIC**, was adopted as the representation of the gravitational potential and became the canonical instantiation of **POT** in the Lorentz model:

$$\mathbf{POT} =_{\mathbf{DEFT}} \mathbf{METRIC}, \tag{XIII}$$

where "$=_{\mathbf{DEFT}}$" is meant to express that the right-hand side of the equation represents the default-setting of the left-hand side. Similarly, we will refer to the energy-momentum tensor of matter and of the electromagnetic field, by the frame **ENEMO**, which became the new standard setting for **SOURCE**:

$$\mathbf{SOURCE} =_{\mathbf{DEFT}} \mathbf{ENEMO}. \tag{XIV}$$

These two key components of the gravitational field equation were generally-covariant tensors and thereby nurtured the expectation that the field equation itself would take the form of a generally-covariant tensorial equation, thus allowing Einstein to realize his ambition of creating a generalized relativity theory.

For the third component of the Lorentz model, the differential operator **OP**, the situation was more complicated. At the beginning of his search, Einstein was largely ignorant of the mathematical techniques necessary for constructing suitable candidates. The many requirements to be imposed on acceptable candidates prevented the selection of an obvious default assumption for the differential operator **OP** compatible with all these requirements.

We summarize the situation in the following figure which we will further elaborate in the following sections:

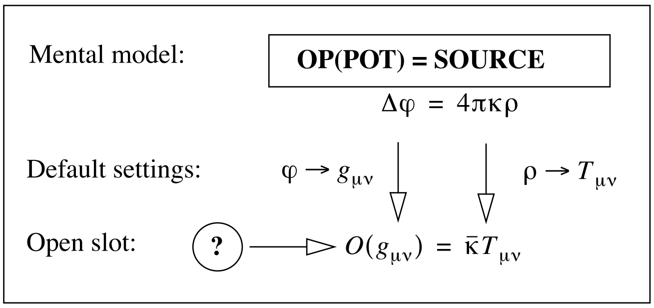


Figure 2: The Lorentz model evolved by changing the default settings for the ***POT****- and the* ***SOURCE****-slots, leaving the question of an appropriate instantiation of the* ***OP****-slot.*

4.1 The Metric as the Potential in the Gravitational Field Equation

In the middle of 1912 Einstein introduced the metric tensor as the new default setting for the **POT**- slot of the Lorentz model. This step affected both the field equation and the equation of motion. The grounds for this move had been prepared by his earlier attempts to set up a theory for the static gravitational field and his awareness that such a theory could only represent a special case within a wider framework suggested by the model. In these attempts Einstein had also learnt that Minkowski's spacetime framework for special relativity could be useful but had to be generalized for use within this larger context. This had been suggested, in particular, by Einstein's controversy with Abraham, pointing to the need for a generalization of the so-called "line element" used in the Minkowski framework, as well as by Einstein's insight into the geometrical consequences of applying special relativity to an accelerated system such as a rotating disk, pointing to the need for non-Euclidean geometry when describing gravitation.

An appropriate generalization of Minkowski's framework was found on the basis of the mathematical work of Gauss, Riemann, Christoffel, Ricci, and Levi-Civita. This led Einstein and Grossmann to the consideration of curvilinear coordinates and the introduction of a metric tensor $g_{\mu\nu}$ for a four-dimensional generalization of Gauss' theory of curved surfaces. Curvilinear coordinates are given by four functions x^{μ} with $\mu = 1, \ldots, 4$, mapping a point of spacetime to four numbers representing its coordinates similar to the use of coordinates in Gaussian surface theory. The generalized line element ds, giving the distance between two neighboring points in spacetime separated by coordinate differentials dx^{μ}, expresses a generalization of the Pythagorean theorem:

$$ds^2 = \sum_{\mu\nu} g_{\mu\nu} dx^{\mu} dx^{\nu}. \tag{21}$$

In the usual representation of Minkowski spacetime in Cartesian coordinates, this expression reduces to the four-dimensional form of the Pythagorean theorem in which the metric tensor $g_{\mu\nu}$ is given by the four-by-four matrix (c being the speed of light):

$$g_{\mu\nu} = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & c^2 \end{bmatrix}. \tag{22}$$

According to the gravito-inertial model inertial forces resulting from arbitrarily accelerated motions can be interpreted as being equivalent to effects of a "dynamic" gravitational field. In the generalized formalism, this model then suggests a natural default setting for the **POT** slot of the frame as well as a natural candidate for the equation of motion in a given gravitational field. The generalized principle of relativity finds a natural expression in terms of the admissibility of arbitrary (smooth) curvilinear coordinate systems representing accelerated reference frames. The inertial motion of a particle in such a reference frame can, on the basis of the gravito-inertial model, be interpreted as motion in a special kind of gravitational field. In the generalized Minkowski formalism such a motion can be described by a geodesic curve, in complete analogy to Gaussian surface theory where geodesic curves represent the natural generalization of straight lines in Euclidean geometry. Combining these two perspectives, it becomes plausible to assume that the motion of a particle under the influence of *any* gravitational field is represented by a geodesic line in a curved spacetime.

Mathematically, a geodesic line can be described as an extremal curve in spacetime determined by a given metric tensor:

$$\delta\left\{\int ds\right\} = 0. \tag{23}$$

From a physical perspective, this equation can be seen as Hamilton's principle (cf. eq. (18)) for the Lagrangian of a free particle of mass m:

$$L = -m\frac{ds}{dt}. \tag{24}$$

The Euler-Lagrange equations (cf. eq. (19)) then suggest to consider the metric tensor $g_{\mu\nu}$ as representing the gravitational potential, i.e. **POT =$_{\mathbf{DEFT}}$ METRIC**, as in eq. (XIII). The combination of the default setting (XIII) and the equation of motion (23) was compatible with the special case of Minkowski spacetime of special relativity where the metric tensor is given by eq. (22) and where the equation of motion of the form of eq. (23) had been developed well before Einstein had begun to work on the problem of gravitation. It was also supported by the special case a static gravitational field, as developed by Einstein in 1912, which could be integrated into the generalized Minkowski formalism in a special and seemingly natural way. Thus one could say that the **POT**-frame specializes to a $\mathbf{POT_{STAT}}$-frame in the context of Einstein's theory of static gravitation:

$$\mathbf{POT} \xrightarrow[\text{gravitation}]{\text{static}} \mathbf{POT_{STAT}} \ , \qquad \mathbf{(XV)}$$

and that the default setting for the $\mathbf{POT_{STAT}}$-frame was given by the following metric:

$$\mathbf{POT_{STAT}} =_{\mathbf{DEFT}} g_{\mu\nu} = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & c^2(x, y, z) \end{bmatrix}, \qquad (25)$$

Here $c(x, y, z)$ is the gravitational potential in Einstein's static theory. In the following we shall refer to this metric as the "canonical metric for a static field." It represents Einstein's default setting for representing a static gravitational field, This default setting mediated between a generic field and the Newtonian case, and was crucial to the heuristics of the correspondence principle.

4.2 The Source-Term in the Gravitational Field Equation

In classical and special-relativistic physics, the relation between force and acceleration (cf. eq. (III)) is not the only way to characterize the effect of a force on a physical system. The effect can also be described in terms of a change in the momentum and in the energy of the system. In classical physics, the force is equal to the rate of change in *time* of the momentum, which can be symbolically expressed as (cf. eq. (XII)):

FORCE = DIFF(MOMENTUM). (XVI)

But the force is also equal to the rate of change in *space* of the energy, which can be symbolically expressed as (cf. eq. (V)):

FORCE = – GRAD(ENERGY). (XVII)

These relations also express that whenever a system gains or loses momentum and energy this must be due to the action of an external force. Note that we have here again introduced new frames **GRAD** and **ENERGY**. As with the previous examples, one could discuss different instantiations of these frames in the context of, say, classical point mechanics, special-relativistic point mechanics, or Maxwellian electrodynamics. But from this point on, we will introduce and make use of our symbolic notation in a more roundabout and indirect way, relying on an intuitive understanding that we hope is conveyed by our choice of names for our symbolic notation (**ENERGY** for energy, **ENEMO** for energy-momentum, etc.). At crucial junctures, however, we will explicitly discuss the concrete instantiations of these frames in Einstein's research and thus provide a general argument for the impact of heuristic reasoning on a qualitative level for Einstein's concrete explorations of pathways out of classical physics.

In special relativity, the concepts of energy and momentum are integrated into a single new concept, the 10-component energy-momentum or stress-energy tensor, which we symbolically represent by the frame **ENEMO** so that the relation between force, energy, and momentum can now be written as:[46]

$$\textbf{FORCE} = -\,\textbf{DIV(ENEMO)}. \qquad \textbf{(XVIII)}$$

In his search for a relativistic gravitational field equation Einstein quickly realized that the source-term, i.e. the instantiation of **SOURCE** in the Lorentz model, had to be the energy-momentum tensor **ENEMO**. In our symbolic notation (cf. eq. (XIV)):

$$\textbf{SOURCE} =_{\textbf{DEFT}} \textbf{ENEMO}.$$

Two lines of arguments, a mathematical and a physical one, made this default setting almost inescapable. From a mathematical point of view—or alternatively, from the point of view of filling the slots of the Lorentz model—something more complex than the scalar mass density was required for **SOURCE** because the gravitational potential is represented by a tensorial object. The slots on both sides of the field equation have to be filled by analogous mathematical objects. While it was in principle conceivable to construct a scalar object out of the metric tensor, e.g. by forming its determinant, and hence to have a scalar field equation, it was more plausible to Einstein that the 10-components of the metric tensor enter into some many-component field equation, just as with the many-component object representing the electromagnetic field.[47]

From a physical point of view—or alternatively from the point of view of the default settings of the Lorentz model based on prior research experience—the energy-momentum tensor had turned out to be the appropriate generalization of the concept of mass in a four-dimensional spacetime setting, i.e.:

$$\textbf{MASS} =_{\textbf{DEFT}} \textbf{ENEMO}. \qquad \textbf{(XIX)}$$

The elaboration of four-dimensional relativistic electrodynamics and hydrodynamics had shown that the introduction of this tensor was necessary in order to adequately describe the energetic and inertial behavior of an extended physical system.[48] In view of Einstein's expectation that, in his relativistic theory of gravitation, energy and

---

46 For the sign compare (CPAE 4, Doc. 1, 92) and "Einstein's Zurich Notebook" 05R (in this volume). The significance of the sign becomes clear when considering the energy-momentum gained or lost by a physical system, for instance in the case of a system with electromagnetic interactions. The divergence of the energy-momentum tensor of the electromagnetic field at a point describes the increase of the energy-momentum of the field at that point, which corresponds to the flow of energy-momentum from the charges to the field. This in turn equals the negative flow of energy-momentum from the field to the charges, which is given by the negative of the Lorentz force.

47 For attempts to build a scalar theory, see John Norton's discussion of Nordström's theory and Einstein's objections in "Einstein, Nordström, and the Early Demise of Scalar, Lorentz Covariant Theories of Gravitation" (in vol. 3 of this series).

mass, as well as gravitational and inertial mass, would essentially be equivalent (cf. eq. (IX)), the energy-momentum tensor was the natural candidate for the right-hand side of the field equation. To sum up, the default setting for the source slot of the Lorentz model of a field equation, eq. (XIV), was inherited from the default setting for the mass slot, eq. (XIX), resulting from the special-relativistic generalization of the mass concept of classical physics.

In order to make the choice of the default setting for **SOURCE** acceptable from the broader point of view provided by eq. (XIX), it was necessary to check whether **ENEMO** also satisfies further properties of **MASS** in classical and special-relativistic physics. The field-theoretical model suggested using the same instantiation of **MASS** both in the field equation and in the equation of motion.

The structure of an equation of motion in a gravitational field involving the **ENEMO**-frame was suggested by the special-relativistic relation between force and energy-momentum represented by eq. (XVIII). Combining this equation with the relation between force and potential, eq. (V), and the appropriate default setting for **CHARGE**, (see eqs. (VII) and (IX) one obtains:

$$\mathbf{GRAD(POT)\ x\ ENEMO = DIV(ENEMO).} \qquad \mathbf{(XX)}$$

Initially, this structural relation provided merely a heuristic hint of what a general equation of motion involving the energy-momentum tensor would look like. To validate this hint, Einstein used a default-setting for the energy-momentum tensor which allowed him to establish a connection between the proper realm of the stress-energy-momentum tensor, i.e. continuum mechanics, and the mechanics of point particles, for which an equation of motion was well established (eq. (23)).[49] In this way, he built a bridge between the knowledge embodied in eq. (XX) and the knowledge that the trajectory of a particle in a gravitational field is a geodesic.

The instantiation **ENEMO** that Einstein used to build this bridge and which, in fact, became its default setting, was the energy-momentum tensor for a swarm of independent particles ("dust"). In our symbolic notation:

$$\mathbf{ENEMO =_{DEFT} DUST,} \qquad \mathbf{(XXI)}$$

where the energy-momentum tensor for **DUST** is mathematically represented by (cf. eq. (4)):

---

48 See the discussion in "Einstein, Nordström, and the Early Demise of Scalar, Lorentz Covariant Theories of Gravitation" (in vol. 3 of this series). At some point in 1913 Einstein even convinced himself that this consideration would altogether rule out a scalar theory of gravitation, which he believed to be incompatible with the conservation laws, but then had to acknowledge that his choice of a tensorial theory with the energy-momentum tensor as a source term was reasonable but not unavoidable. His own subsequent exploration of a relativistic scalar theory of gravitation made it clear, however, that such a theory was based on *a priori* assumptions about the geometry of spacetime, which Einstein was not willing to accept. Hence even this apparently far-fetched consideration, based on Mach's critique of Newton's concept of space, contributed to stabilizing Einstein's choice of the energy-momentum tensor as the default-setting for **SOURCE**.

49 "Einstein's Zurich Notebook" 05R, p. 43R (in this volume).

$$T^{\mu\nu} = \rho\frac{dx^{\mu}}{ds}\frac{dx^{\nu}}{ds}. \tag{26}$$

In the Lagrangian formalism for a point particle in a gravitational potential given by the metric tensor (cf. eq. (24)), Einstein obtained expressions for the momentum and energy of a particle. He then applied these expressions to the energy-momentum tensor for dust and interpreted the resulting terms. In this way he arrived at an equation corresponding to the structural relation eq. (XX):[50]

$$\frac{1}{2}\sqrt{-g}\frac{\partial g_{\mu\nu}}{\partial x^{\sigma}}T^{\mu\nu} = \frac{\partial}{\partial x^{\nu}}(\sqrt{-g}g_{\sigma\mu}T^{\mu\nu}). \tag{27}$$

Introducing mixed tensor densities $\mathfrak{T}_{\sigma}^{\nu} = \sqrt{-g}g_{\sigma\mu}T^{\mu\nu}$, we can write this equation more compactly:

$$\frac{1}{2}g^{\alpha\mu}\frac{\partial g_{\mu\nu}}{\partial x^{\sigma}}\mathfrak{T}_{\alpha}^{\nu} = \frac{\partial\mathfrak{T}_{\sigma}^{\nu}}{\partial x^{\nu}}. \tag{28}$$

The agreement between this concrete result for the default setting **DUST** with the general relation eq. (XX) supported the underlying physical heuristics and suggested that this equation holds for an arbitrary symmetric energy-momentum tensor.[51]

Equation (27) turned out to be very important for Einstein's further research. First, it supported the choice of the energy-momentum tensor as the source of a gravitational field equation and stabilized this instantiation of both **SOURCE** and **MASS**. Second, it provided Einstein with one of the fundamental components for the Lorentz

50 "Einstein's Zurich Notebook" 05R, (in this volume). In this form eq. (27) is valid only for a symmetric tensor $T^{\mu\nu}$. See also (CPAE 6, Doc. 9, 95).

51 The terms in eq. (27) are interpreted in (Einstein and Grossmann 1913) in the following way: "We ascribe to equation (10) [i.e. our (27)] a validity range that goes far beyond the special case of the flow of incoherent masses. The equation represents in general the energy balance between the gravitational field and a arbitrary material process; one has only to substitute for [$T^{\mu\nu}$] the stress-energy tensor corresponding to the material system under consideration. The first sum in the equation contains the space derivatives of the stresses or of the density of the energy flow, and the time derivatives of the momentum density or of the energy density; the second sum is an expression for the effects exerted by the gravitational field on the material process." ("Der Gleichung [(27)] schreiben wir einen Gültigkeitsbereich zu, der über den speziellen Fall der Strömung inkohärenter Massen weit hinausgeht. Die Gleichung stellt allgemein die Energiebilanz zwischen dem Gravitationsfelde und einem beliebigen materiellen Vorgang dar; nur ist für [$T^{\mu\nu}$] der dem jeweilen betrachteten materiellen System entsprechende Spannungs-Energietensor einzusetzen. Die erste Summe in der Gleichung enthält die örtlichen Ableitungen der Impuls- bzw. Energiedichte; die zweite Summe ist ein Ausdruck für die Wirkungen, welche vom Schwerefelde auf den materiellen Vorgang übertragen werden." p.11). Indeed, except for the original derivation of eq. (27) in (Einstein and Grossmann 1913), in all later publications up to 1916, this relation appears in a form where the two conceptually distinct terms of the left-hand side are set equal, rather than in the form of eq. (27) which asserts the vanishing of a generally-covariant object that only happens to be represented by the algebraic difference of two terms as in eq. (27).

model, a general equation of motion which describes how material processes are affected by the gravitational field. Third, this equation became, as we shall see, the starting point for the formulation of the requirement of energy-momentum conservation that had to be satisfied by any candidate for the left-hand side of the gravitational field equation. Fourth, its left-hand side suggested, in connection with the relation between **FIELD** and **POT** in eq.(II), an instantiation of **FIELD**:

$$\textbf{FIELD} = -\,\textbf{GRAD(POT)} =_{\textbf{DEFT}} -\frac{1}{2}g^{\alpha\mu}\frac{\partial g_{\mu\nu}}{\partial x^{\sigma}} \equiv \tilde{\Gamma}^{\alpha}_{\sigma\nu}. \qquad \textbf{(XXII)}$$

This choice was plausible but not without alternatives. Eq. (28) can also be written as:

$$-\Gamma^{\alpha}_{\sigma\nu}\mathfrak{T}^{\nu}_{\alpha} = \frac{\partial\mathfrak{T}^{\nu}_{\sigma}}{\partial x^{\nu}}, \qquad (29)$$

with $\Gamma^{\alpha}_{\sigma\nu}$ —in the following symbolically represented as **CHRIST**—defined as minus the so-called Christoffel symbols (of the second kind):

$$\Gamma^{\alpha}_{\sigma\nu} = -\left\{\begin{matrix}\alpha\\ \sigma\nu\end{matrix}\right\} = -\frac{1}{2}g^{\alpha\mu}\left(\frac{\partial g_{\mu\nu}}{\partial x_{\sigma}} + \frac{\partial g_{\mu\sigma}}{\partial x_{\nu}} - \frac{\partial g_{\sigma\nu}}{\partial x_{\mu}}\right). \qquad (30)$$

As a consequence, one obtains an alternative instantiation of **FIELD**:

$$\textbf{FIELD} =_{\textbf{DEFT}} -\,\textbf{CHRIST} =_{\textbf{DEFT}} \Gamma^{\alpha}_{\sigma\nu}. \qquad \textbf{(XXIII)}$$

The familiar form of the relation between field and potential in classical field theory made eq. (XXII) the natural first choice, and eq. (XXIII) only came into play when this first choice turned out to lead to difficulties.

Equation (27) had one final important implication. Written in the form:

$$\frac{\partial}{\partial x^{\nu}}(\sqrt{-g}g_{\sigma\mu}T^{\mu\nu}) - \frac{1}{2}\sqrt{-g}\frac{\partial g_{\mu\nu}}{\partial x^{\sigma}}T^{\mu\nu} = 0, \qquad (31)$$

its left-hand side could be conceived as a generic, generally-covariant differential operator known as “covariant divergence,” here symbolically represented as **DIV** $_{\textbf{COV}}$**(.)**, so that eq. (XX) can also be written as:[52]

$$\textbf{DIV}_{\textbf{COV}}\textbf{(ENEMO)} = \textbf{DIV(ENEMO)} - \textbf{GRAD(POT) x ENEMO} = \textbf{0}. \qquad \textbf{(XXIV)}$$

Although Einstein interpreted eq. (31) primarily from a physical point of view i.e. as a representation of the structure (XX), as we have seen, he knew, probably even before he became acquainted with the absolute differential calculus, that this equa-

52 Note that the embodiment eq. (31) of the symbolic eq. (XXIV) holds only for symmetric tensors.

tion involves a generic tensor operation which is generally covariant.[53] He had thus recognized the covariant divergence as a mathematical ingredient of his new theory that was meaningful in its own right and could in principle be used for other purposes. The formulation of eq. (31) is a prime example of how Einstein's physical strategy produced a result that turned out to be independent of the specifics of its derivation, such as the choice of **DUST** for **ENEMO**. Einstein even attempted to use the covariant divergence as a constituent of a candidate for the left-hand side of the gravitational field equation but failed because it vanishes when applied to the metric tensor.[54] The fact that the equation of motion expressed in terms of **ENEMO** turned out to be generally covariant must, in any case, have been an important confirmation of his program to establish a generally-relativistic theory of gravitation, suggesting that the other major constituent of the Lorentz model, the field equation, should also have this property.

4.3 The Differential Operator in the Gravitational Field Equation

For the differential operator acting as **OP** in eq. (I), Einstein did not have an immediately satisfactory candidate or even a heuristic shortcut for finding one. Substituting the metric tensor for the scalar gravitational potential quickly drove him out of any familiar mathematical terrain. He had to find a second-order differential operator acting on the metric tensor by relying either on attempts to directly construct such an operator or on the mathematical literature in order to find suitable starting points.

One of Einstein's earliest attempts[55] to construct a differential operator **OP** was to mimic the way in which the classical Laplace operator was formed, that is, by compounding the differential operations divergence and gradient familiar from three-dimensional vector calculus. In this way he obtained a first, natural instantiation for the differential operator on the left-hand side of the gravitational field equation:[56]

$$\mathbf{OP} =_{\mathbf{DEFT}} \mathbf{LAP} = \mathbf{DIV\ (GRAD)} \qquad \mathbf{(XXV)}$$

Applying **LAP** to the default setting for **POT**, we obtain what we will call the *core operator*:[57]

---

53 Einstein's remark "I have now found the most general equations." ("Ich habe nun die allgemeinsten Gleichungen gefunden.") in a letter to Ludwig Hopf, dated 16 August 1912 (CPAE 5, Doc. 416) in all probability refers to this insight, cf. the editorial note "Einstein on Gravitation and Relativity: The Collaboration with Marcel Grossmann" (CPAE 4, 294–301). The covariance of this equation was demonstrated in terms of the absolute differential calculus of Ricci and Levi-Civita by showing that it represents the covariant divergence of the (symmetric) contravariant stress-energy-tensor in Grossmann's "mathematical part" of (Einstein and Grossmann 1913, 32).

54 See p. 05R of "Einstein's Zurich Notebook" (in this volume).

55 The following discussion relies heavily on the analysis of Einstein's research notes contained in the Zurich Notebook. Since the actual historical path will be discussed in chapter 6, we will here only refer to the relevant pages of this notebook, without any further comments.

56 Cf. pp. 07R and 08L of "Einstein's Zurich Notebook" (in this volume).

57 Cf., e.g., p.07L of "Einstein's Zurich Notebook" (in this volume).

$$\mathbf{LAP(POT)} =_{\mathbf{DEFT}} \sum_{\alpha\beta} \frac{\partial}{\partial x^{\alpha}}\left(g^{\alpha\beta}\frac{\partial g^{\mu\nu}}{\partial x^{\beta}}\right). \qquad \mathbf{(XXVI)}$$

It has to be noted that this equation is only one of several possible instantiations of the frame **LAP**. Alternative instantiations typically involve additional factors of the determinant of the metric which typically affect the transformation behavior of the particular version of the core operator under consideration.

As a symbolic equation, the resulting tentative field equation reads:

$$\mathbf{LAP(POT) = ENEMO.} \qquad \mathbf{(XXVII)}$$

This instantiation of **OP** was supported by several arguments based, in particular, on the correspondence principle. The straightforward generalization of the Laplace operator was also plausible against the background of the field equation Einstein had developed for static gravitational fields. This field equation resulted from a simple instantiation of the Lorentz model obtained essentially by replacing the Newtonian potential in the classical Poisson equation by the variable speed of light, a move suggested by the equivalence principle:[58]

$$\Delta c = kc\rho. \qquad (32)$$

The constant k is related to the gravitational constant κ of the Poisson equation through $k = (4\pi\kappa/c^2)$.[59]

In the following, we discuss the implications of Einstein's heuristic framework for choosing and modifying the instantiations for the gravitational differential operator in his field equation: We examine the implications coming from the correspondence principle, the conservation principle, the generalized principle of relativity, and examine the Lagrangian formalism, respectively. This discussion is not meant as a substitute for a detailed account of Einstein's pathway, but as preparation for such an account by identifying the constraints under which it was pursued. These constraints

---

58 The explicit justification for this equation was follows. After noting that the variable velocity of light fulfills the Laplace equation for the matter-free case, Einstein continues: "It is easy to establish the presumably valid equation that corresponds to Poisson's equation. For it follows immediately from the meaning of c that c is determined only up to a constant factor that depends on the constitution of the clock with which one measures [the time] t at the origin of [the accelerated coordinate system] K. Hence the equation corresponding to Poisson's equation must be homogeneous in c. The simplest equation of this kind is the linear equation [eq. (32)] where k denotes the (universal) gravitational constant, and ρ the matter density." ("Es ist leicht diejenige vermutliche Gleichung aufzustellen, welche derjenigen von Poisson entspricht. Es folgt nämlich aus der Bedeutung von c unmittelbar, daß c nur bis auf einen konstanten Faktor bestimmt ist, der davon abhängt, mit einer wie beschaffenen Uhr man t im Anfangspunkte von K mißt. Die der Poissonschen Gleichung entsprechende muß also in c homogen sein. Die einfachste Gleichung dieser Art ist die lineare Gleichung [eq. (32)], wenn unter k die (universelle) Gravitationskonstante, unter ρ die Dichte der Materie verstanden wird.") (Einstein 1912b, 360)

59 The relation is obtained by identifying $c^2/2$ with the Newtonian potential φ and neglecting terms of order $(\partial_i c)^2$, cf. (Einstein 1912a, 362).

were rooted in the knowledge of classical physics, which provided the default settings for the frames with which Einstein operated. That these default settings often led to conflicting results necessitating their modification or replacement lies in the nature of Einstein's search, whose outcome could not be anticipated.

4.4 Implications of the Correspondence Principle

A gravitational field equation based on the core operator as given by eq. (XXVI) is in accordance with the correspondence principle, thus strengthening the role of this operator as an instantiation for the left-hand side of the field equation. For weak fields this differential operator reduces to the d'Alembertian operator, the default-setting for **OP** in the weak-field limit. The transition to this limiting case can be represented symbolically as:[60]

LIM(OP(.)) =$_{\mathbf{DEFT}}$ LIM(LAP(.)) =$_{\mathbf{DEFT}}$

$$\Box(\;) = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right). \qquad \textbf{(XXVIII)}$$

The weak-field equation thus takes on the canonical form:

$$\Box g_{\mu\nu} = \kappa T_{\mu\nu}. \qquad (33)$$

This equation can also be written as:

$$\Box h_{\mu\nu} = \kappa T_{\mu\nu}, \qquad (34)$$

where $h_{\mu\nu}$ is defined by

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \qquad (35)$$

with $|h_{\mu\nu}| \ll 1$ denoting small deviations from the Minkowski metric:

$$\eta_{\mu\nu} = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & c^2 \end{bmatrix}. \qquad (36)$$

If the source is taken to first order as a pressureless, static cloud of dust of density ρ (compare eq. (4)), one can neglect all terms of the energy-momentum tensor on the right-hand side of eq. (33) except for the T_{44} - term, which can be identified with the

60 Cf. (Einstein and Grossmann 1913, 13). Note that in contrast to the *Entwurf* operator, the core operator reduces to the Laplacian for a static metric of the form (25) for *strong* static fields as well. Einstein never seems to have considered this case.

gravitating mass density appearing in the classical Poisson equation. The neglected terms in the energy-momentum tensor involve the velocity of the gravitating matter which, in the Newtonian case, will be small compared to the velocity of light. If one now considers the case of a *static* weak field, introducing **LIM$_{\mathbf{STAT}}$** and using the static metric of the canonical form (25) on the left-hand side of the weak-field equation, one has:

$$\mathbf{LIM_{STAT}(OP(POT)) =_{DEFT} LIM(LAP(POT_{STAT}))}. \qquad \mathbf{(XXIX)}$$

This expression reduces to the Laplace operator acting on a single component of the metric. Eq. (33) thus reduces to the familiar Poisson equation:

$$\Delta g_{44} = \kappa\rho \text{ or equivalently } \Delta h_{44} = \kappa\rho. \qquad (37)$$

The equation of motion in the Newtonian limit can be obtained from eq. (31) under similar assumptions, i.e., small velocities and a weak static field. The result is:

$$\frac{d^2 x_i}{dt^2} = -\frac{1}{2}\frac{\partial g_{44}}{\partial x_i} = -\frac{1}{2}\frac{\partial h_{44}}{\partial x_i} \qquad (i = 1, 2, 3). \qquad (38)$$

This equation shows that $g_{44}/2$ resp. $h_{44}/2$ plays the role of the Newtonian gravitational potential.

The assumption that the left-hand side of the field equation has the form of eq. (XXVIII) is not independent from the assumption that the metric tensor for weak static fields has the form (25). Under appropriate circumstances, a weak-field equation of this form gives rise to solutions precisely of this canonical form.[61] In other words, the most natural assumption for the form of a weak-field equation and the most natural assumption for the metric of a static field supported each other. A further argument supporting Einstein's understanding of the correspondence principle as implying a canonical metric of the form (25) was independent from the field equation but also related to the roots of this principle in the framework of classical physics.[62] This argument is based on Galileo's principle, that is, the requirement that all bodies fall with the same acceleration in a given gravitational field, and makes use of the basic relations between force, momentum, energy, and acceleration as understood in classical physics, with some additional ingredients from special relativity such as the equivalence of mass and energy. Einstein argued that particles with different energy, and hence different inertial mass, fall with different accelerations in a static gravitational field, *unless* such a field is represented by a metric tensor of the canonical form (25). As a criterion for the validity of Galileo's principle he used the requirement that the ratio of the force acting on a particle and its energy depend neither on the particle's mass nor on its velocity.

61 See (Norton 1984, 120–121).
62 The following argument is based on a reconstruction of p. 21R of "Einstein's Zurich Notebook" (in this volume).

4.5 Implications of the Conservation Principle

The fact that the core operator was firmly anchored in knowledge about the familiar cases of static and Newtonian gravitation made it the natural starting point for Einstein's "physical strategy." The core operator, however, had to pass a number of further checks, which could result in modifications. In particular, it remained to be seen how the operator could be brought into agreement with the conservation principle and the generalized relativity principle.

An acceptable field equation (I) had to be compatible with the equation of motion and the related structural insight into energy-momentum conservation represented by eq. (XX). This compatibility could be checked by replacing **ENEMO** in eq. (XX) by the left-hand side of the field equation, i.e. by **OP**:

$$\mathbf{GRAD(POT)\ x\ OP(POT) = DIV(OP(POT))}, \tag{XXX}$$

or, in the notation of eq. (XXIV):

$$\mathbf{DIV_{COV}(OP(POT)) = 0.} \tag{XXXI}$$

It was necessary to check whether this "conservation compatibility check" could be satisfied for a given candidate field equation if need be by imposing extra conditions, in addition to the field equation.

In the course of Einstein's research documented in the Zurich Notebook it became clear that the conservation compatibility check fails for a field equation based on the core operator:

$$\mathbf{GRAD(POT)\ x\ LAP(POT) \neq DIV(LAP(POT)).} \tag{XXXII}$$

This problem may not have surprised Einstein as it was already familiar to him from his theory of static gravitational fields. There he had also encountered the difficulty that the first choice of a field equation for the static field (eq. (32)) turned out to be incompatible with momentum conservation.[63] To demonstrate this conflict, Ein-

63 Cf. Einstein's second thoughts about the paper expressed in a letter to the editor of the *Annalen der Physik*, Wilhelm Wien: "I asked you this morning to return my manuscript, and now I am asking you to keep it after all. To be sure, not everything in the paper is tenable. But I think I should let the thing stand as it is, so that those interested in the problem can see how I arrived at the formulas." ("Heute Morgen bat ich Sie, mir mein Manuskript zurückzusenden und nun bitte ich Sie es doch zu behalten. Es ist zwar nicht alles haltbar, was in der Arbeit steht. Aber ich glaube die Sache doch so lassen zu sollen, damit diejenigen, welche sich für das Problem interessieren, sehen, wie ich zu den Formeln gekommen bin.") Einstein to Wien, 11 March 1912, (CPAE 5, Doc. 371). Since (Einstein 1912b), which contains eq. (32) was received by the *Annalen* on February 26, and (Einstein 1912c) where the problem with this equation is discussed, was received four weeks later, on March 23, the problem Einstein refers to in the letter to Wien is most probably the incompatibility with the conservation principle discussed in the following. To the published discussion of the potential equation for c in (Einstein 1912b) Einstein added a footnote reading: "A soon to be published paper will show that equation (5a) and (5b) cannot yet be exactly right. However, they will be provisionally used in the present paper." ("In einer in kurzem nachfolgender Arbeit wird gezeigt werden, daß die Gleichungen [$\Delta c = 0$] und [eq. (32)] noch nicht exakt richtig sein können. In dieser Arbeit sollen sie vorläufig benutzt werden.")

stein considered an assembly of masses fixed to a rigid, massless frame and showed that this assembly of masses would set itself in motion if the field equation were assumed to be (32). Following the logic underlying eq. (XXX), he substituted the left-hand side of the field equation for the mass density ρ in the expression for the force-density (compare eq. (V)):

$$\boldsymbol{F} = -\rho \operatorname{grad} c. \tag{39}$$

The integral of this expression over space (under the assumption that c is constant at infinity) should vanish on account of momentum conservation. However, the expression

$$\boldsymbol{F} = -\frac{1}{k}\frac{\Delta c}{c} \operatorname{grad} c, \tag{40}$$

resulting from this substitution cannot be transformed into a divergence expression, and momentum conservation is violated. The rigid massless frame would start to move, in contradiction with Newton's principle *actio = reactio*.

It is easily seen that the Poisson equation of classical mechanics and electrostatics does not present this problem. In a later paper Einstein himself explained how this can be shown in a way that suggests a generalization of the argument to the case of a relativistic gravitational field theory.[64] In electrostatics the ν th component of the momentum conferred to matter per unit volume and time (or the force density, compare (V)) is:

$$-\frac{\partial \varphi}{\partial x_\nu}\rho,$$

where φ represents the potential and ρ the density of the electrical charge. It can then be demonstrated that a field equation of the form (cf. eq. (5)):

$$\sum_\nu \frac{\partial^2 \varphi}{\partial x_\nu^2} = -\rho$$

satisfies the requirement of momentum conservation. This is done by showing that the rate of change of momentum:

$$-\frac{\partial \varphi}{\partial x_\nu} \cdot \rho = \frac{\partial \varphi}{\partial x_\nu}\sum_\mu \frac{\partial^2 \varphi}{\partial x_\mu^2}$$

can be transformed into a divergence expression, i.e., an expression with the property that the integral over a closed system vanishes so that the total momentum is conserved.

The challenge resulting from the problem with Einstein's first static theory was to find an expression for the force, the momentum transferred from the gravitational

64 See (Einstein and Grossmann 1913, part 1, §5).

field to material processes that can be written as a divergence. Let us try to capture the heuristic behind his reasoning in our symbolic notation. We conceive of the **FORCE**-frame as a divergence of some **FIELDMASS**-frame in an equation of the form:

FORCE = DIV(FIELDMASS), (XXXIII)

where **FIELDMASS** represents, in the three-dimensional case, the momentum (or, alternatively, the energy) and, in the four-dimensional case, the energy-momentum of the gravitational field. Such a force expression had to be extracted from a revised field equation in which the default setting **LAP** is replaced by a modified frame, let us call it **GRAV** for **OP**:

OP(POT) $=_{\mathbf{DEFT}}$ GRAV(POT) = LAP(POT) + CORR(POT). (XXXIV)

The correction term **CORR** introduced in the new choice for **OP** in eq. (XXXIV) had to be compatible, of course, with the correspondence principle and in particular with the default setting eq. (XXVIII) so that the condition

LIM(CORR(POT)) = 0 (XXXV)

follows. The correction term has to make sure that both eq. (XXX), the conservation compatibility check, and eq. (XXXIII), the equivalent divergence condition for the gravitational force, are satisfied:

FORCE =

– GRAD(POT) X GRAV = – DIV(GRAV) = DIV(FIELDMASS). (XXXVI)

In view of the definition of **GRAV** as a sum of **LAP** and **CORR** (see eq. (XXXIV)) one thus obtains the following symbolic equation:

LAP(POT) x GRAD(POT) =

– DIV(FIELDMASS) – CORR(POT)) x GRAD(POT). (XXXVII)

The crucial result is that this relation suggests a generic operational procedure for identifying the desired correction term, regardless of specific instantiations of the frames involved. From the force expression for Einstein's first field equation for static fields, eq. (40), it follows that the term which serves as the starting point for such a procedure, corresponding to **LAP(POT) x GRAD(POT)**, is:[65]

$$\frac{1}{c}\partial_i\partial_i c\partial_k c, \tag{41}$$

By repeated application of the Leibniz rule for the differentiation of products, one obtains an equation of the form (XXXVII):

65 In the following we assume summation over repeated (spatial) indices.

$$\frac{1}{c}\partial_i\partial_i c\partial_k c = \partial_i T_{ik} + \frac{1}{2c^2}\partial_i c\partial_i c\partial_k c \tag{42}$$

with

$$T_{ik} = \frac{1}{c}\partial_i c\partial_k c - \frac{\delta_{ik}}{2c}\partial_j c\partial_j c, \tag{43}$$

so that

$$\partial_i T_{ik} = \left(\frac{1}{c}\partial_i\partial_i c - \frac{1}{2c^2}\partial_i c\partial_i c\right)\partial_k c \tag{44}$$

is a divergence term and corresponds to **– DIV(FIELDMASS)**, while

$$\frac{1}{2c^2}\partial_i c\partial_i c\partial_k c \tag{45}$$

corresponds to **– CORR(POT)) x GRAD(POT)**. Eq. (45) therefore gives the correction term necessary to satisfy the conservation principle. In other words, this principle not only served to refute the first static field equation (32), it also provided Einstein with a procedure for constructing a modified field equation complying with this principle.

Einstein thus arrived at a new field equation (Einstein 1912b), the core of his so-called "second theory:"[66]

$$\frac{\Delta c}{c} - \frac{\mathrm{grad}^2 c}{2c^2} = k\sigma. \tag{46}$$

Since this revised equation no longer represents a direct analogue of the Poisson equation, Einstein faced the challenge to find a plausible physical interpretation of it. He had to reexamine both the equivalence principle and the role of energy and momentum conservation. A remarkable feature of eq. (46) is that the first derivative of the gravitational potential enters in a non-linear way so that the left-hand side of eq. (46) may be symbolically expressed with the help of eqs. (II) and (XXXIV) as:

$$\mathbf{GRAV} =_{\mathbf{DEFT}} \mathbf{DIV(FIELD)} + \mathbf{FIELD}^2 \tag{XXXVIII}$$

The second term had not been encountered before in working with the mental model of a gravitational field theory. It also threatened one of Einstein's key heuristic assumptions, the principle of equivalence, which could only be upheld for infinitesimally small fields.[67] This restriction made it all the more pressing to provide a plausible physical justification for the correction term. Einstein found such a justification in implications of both field theory and special relativity, i.e. in the fact that a field may

66 The theory advanced in this paper is commonly referred to as Einstein's second theory of static gravitation. Its main difference pertains to the amended field equation.

carry energy and that any kind of energy, being equivalent to mass, should act as a source of the gravitational field.[68]

The physical interpretation of the modified field equation (46) is brought out more clearly by rewriting it as:

$$\Delta c = k\left(c\sigma + \frac{1}{2ck}\mathrm{grad}^2 c\right). \tag{47}$$

This form of the equation suggests that the term

$$\frac{1}{2ck}\mathrm{grad}^2 c,$$

which appears on the right-hand side on the same footing as the material source, be interpreted as the energy density of the gravitational field acting as its own source.

This physical interpretation also supported the conclusion that the general field equation would be non-linear, a conclusion which, after this experience with the special case of the static field, became a standard expectation in Einstein's further search. In terms of our symbolic equations, the revised form of the generic field equation could either be expressed with the help of the **GRAV** and **CORR**-frames (see eq. (XXXIV)) or with the help of the **FIELDMASS**-frame, representing in the general, four-dimensional case the energy-momentum of the gravitational field, as:[69]

NORM(POT) = ENEMO + FIELDMASS. (XXXIX)

NORM(POT) thus represents a new setting of the differential operator slot **OP** allowing the field equation (I) to be written in the "normal" form of eq. (XXXIX); we thus define

OP =$_{\mathbf{DEFT}}$ NORM, (XL)

with a corresponding new setting for **SOURCE**:

---

67 "Thus, it seems that the only way to avoid a contradiction with the reaction principle is to replace equations (3) and (3a) with other equations homogeneous in c for which the reaction principle is satisfied when the force postulate (4) is applied. I hesitate to take this step because by doing so I am leaving the territory of the unconditional equivalence principle. It seems that the latter can be maintained for infinitely small fields only." ("Eine Beseitigung des genannten Widerspruches gegen das Reaktionsprinzip scheint also nur dadurch möglich zu sein, daß man die Gleichungen [$\Delta c = 0$] und [eq. (32)] durch andere in c homogene Gleichungen ersetzt, für welche das Reaktionsprinzip bei Anwendung des Kraftansatzes [39] erfüllt ist. Zu diesem Schritt entschließe ich mich deshalb schwer, weil ich mit ihm den Boden des unbedingten Äquivalenzprinzips verlasse. Es scheint, daß sich letzteres nur für unendlich kleine Felder aufrechterhalten läßt.") (Einstein 1912c, 455–456)

68 The discussion referred to in the following is introduced in (Einstein 1912c) by the phrase: "The term added in equation (3b) [$c\Delta c = \frac{1}{2}(\mathrm{grad}\ c)^2 = kc^2\sigma$] in order to satisfy the reaction principle wins our confidence thanks to the following argument." ("Das in Gleichung (3b) zur Befriedigung des Reaktionsprinzipes hinzugesetzte Glied gewinnt unser Vertrauen durch die folgenden Überlegungen."), p. 456–7.

69 For the following, see "Untying the Knot ..." (in vol. 2 of this series), sec. 3.

$$\mathbf{SOURCE} =_{\mathbf{DEFT}} \mathbf{ENEMO + FIELDMASS.} \qquad \mathbf{(XLI)}$$

This form of the field equation clearly brings out the parallelism between the energy-momentum of matter and the energy-momentum of the gravitational field. Eq. (XXXIX) is the symbolic expression of what eventually became Einstein's standard or normal expectation for the form of a field equation with the property that it is compatible with the conservation principle and with the requirement that gravitational energy and momentum enter the field equation on the same footing as the energy and momentum of matter. With this normal form the conservation principle takes on a particularly simple form. From the last equality in eq. (XXXVI) and the field equation it follows that:

$$\mathbf{DIV(ENEMO) + DIV(FIELDMASS) =}$$

$$\mathbf{DIV(ENEMO + FIELDMASS) = 0.} \qquad \mathbf{(XLII)}$$

This symbolic equation expresses the expectation that the conservation laws should hold for gravitation and matter taken together. Accordingly, the conservation compatibility check for **NORM(POT)** becomes

$$\mathbf{DIV(NORM) = 0,} \qquad \mathbf{(XLIII)}$$

(cf. eq. (XXXI))

It was natural to expect that **NORM** would take on the classical form of a divergence of the field, generated both by material processes and the energy-momentum of the gravitational field itself. The field operator might be brought into such a simple form, resembling the familiar structure from electromagnetic field theory by some appropriate mathematical manipulation, involving the source-term of the field equation as well. In other words, one would have the revised settings:

$$\mathbf{OP(POT)} =_{\mathbf{DEFT}} \mathbf{NORM(POT)}_{\mathbf{CLASS}} = \mathbf{DIV(FIELD),} \qquad \mathbf{(XLIV)}$$

with a corresponding setting for **SOURCE**:

$$\mathbf{SOURCE} =_{\mathbf{DEFT}} \mathbf{ENEMO + FIELDMASS.} \qquad \mathbf{(XLV)}$$

Note, however, that the requirement expressed by eq. (XLII) may not be compatible with the requirement expressed by eq. (XXXIX) if the particular form eq. (XLIV) for the left-hand side of the field equation is imposed.[70]

In summary, Einstein's experiences with implementing the conservation principle in his theory for static gravitational fields turned out to be of crucial significance for his further research, shaping the expectation for the differential operator in the generic gravitational field equation. Reflecting on these experiences, he could conclude, in particular, that

– the field equation would probably be non-linear and contain a term representing the gravitational field acting as its own source;

70 See the discussion in "Untying the Knot ..." (in vol. 2 of this series), secs. 3, §3.

– just as with the static field equation, the field equation might have to be found in two steps with a first step involving a linear second-order differential operator and a second step involving the non-linear, first-order correction terms;
– the correction term might be identified by trying to establish an energy-momentum balance, beginning with a linear, second-order differential operator as a first step.

4.6 Implications of the Generalized Relativity Principle

When starting from an instantiation of the left-hand side of the mental model of a gravitational field equation rooted in physical knowledge such as the core operator, the most challenging problem was to identify its transformation properties and to find out whether or not they allow the implementation of a generalized principle of relativity. Alternatively, one could start from an instantiation rooted in mathematical knowledge. While the physical strategy automatically takes care of the correspondence principle, the mathematical strategy automatically takes care of the generalized relativity principle. In the latter case the main challenge was the implementation of the correspondence and conservation principles, including a check of their mutual compatibility. In the course of his research, Einstein developed a strategy for addressing this challenge. This strategy involved replacing one immediately given default setting for **OP** by a more sophisticated one, better adapted to the purpose at hand. In this respect, the strategy resembles the strategy discussed above for adapting a setting suggested by the correspondence principle to the necessities implied by the conservation principle, i.e., for the transition from **LAP(POT)** to **GRAV(POT)**.

Instantiations for **OP** suggested by the mathematical strategy typically have well-defined transformation properties (e.g. are generally covariant). As his research proceeded Einstein familiarized himself with the relevant mathematical literature, in collaboration with his mathematician friend Marcel Grossmann.[71] While the mathematical horizon enlarged it came to include more and more sophisticated mathematical objects. At the beginning, the mathematical instrumentarium was limited to that of linear vector and tensor analysis in four dimensions as developed by Minkowski, Sommerfeld, and Laue.[72] After a number of unsuccessful attempts to employ these techniques in the construction of a suitable differential operator,[73] the core operator emerged as the most satisfactory candidate which could be obtained at this level of mathematical sophistication. The core operator, however, was covariant only under linear transformations and did thus not lead to a substantial generalization of the relativity principle.

71 For a discussion of Grossmann's role in the search for and reception of pertinent mathematical literature, see (Pais 1982, chap.12c; Norton 1992b, appendix; Reich 1994, chap. 5.3; CPAE 4, 294).
72 Cf. note 30.
73 Cf. pp. 39L–40L of "Einstein's Zurich Notebook" (in this volume).

Einstein subsequently became familiar with the so-called Beltrami invariants.[74] These mathematical objects, in particular the second Beltrami invariant, could be seen as a generalization of the ordinary Laplace operator and must have looked promising. They are generally covariant and thus provide a good starting point for pursuing the mathematical strategy. It was difficult, however, to see how the second Beltrami invariant, defined only for scalar functions, could be applied to a gravitational potential represented by the metric tensor. Einstein thus had two plausible but mutually incompatible default settings, the second Beltrami invariant for the differential operator, and the metric tensor for the gravitational potential. Einstein's dilemma at this point is illustrated in Fig. 3.

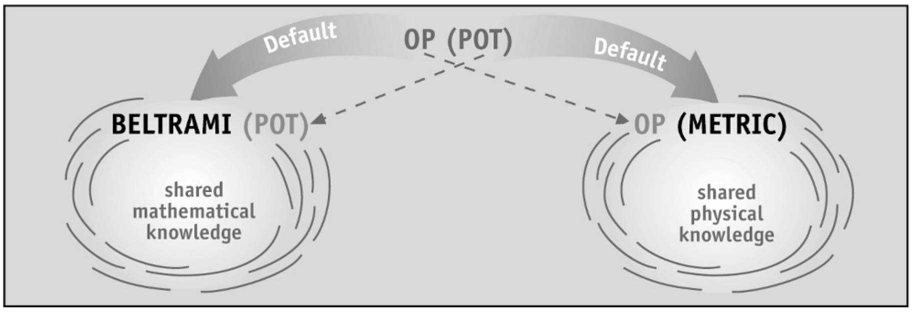


Figure 3: The incompatibility of instantiating the operator slot of Einstein's mental model of a gravitational field equation with the Beltrami invariant, and the potential slot with the metric produced posed a dilemma for Einstein.

The breakthrough for the mathematical strategy came when Einstein got acquainted with the Riemann tensor and its potential to produce suitable candidates for the differential operator in the gravitational field equation. The Riemann tensor represents a second-order differential operator on the metric and is generally covariant. Moreover, by a general theorem any generally-covariant differential operator, which consists of the metric components and its derivatives, and contains no higher than second-order derivatives and is linear in those, can be constructed from the Riemann tensor by tensor-algebraic operations.[75] It must have been clear to Einstein from the outset that the Riemann tensor itself could not play the role of an instantiation for **OP**. First, since the energy-momentum tensor appearing on the right-hand side of the field equation is a second-rank tensor with two indices, the differential operator on the left-hand side was required to have the same property. The Riemann tensor, however, is a fourth-rank tensor, with four indices. Second, a field equation with the Riemann tensor on the left-hand side would be much too restrictive. It would require that, outside the sources, the metric would be strictly Minkowskian so no

74 Cf. pp. 06L–07L of "Einstein's Zurich Notebook" (in this volume).
75 See (Einstein and Grossmann 1913, part II, §4). See also (Bianchi 1910).

non-trivial gravitational potential could exist, a conclusion, for instance, manifestly wrong for the field of a point mass.

A second-rank tensor serving as a natural candidate for **OP**, however, could be extracted from the Riemann tensor in various ways. Let us designate the frame of such a candidate by **RIEM**:

$$\mathbf{OP(POT)} =_{\mathbf{DEFT}} \mathbf{RIEM(POT).} \qquad \mathbf{(XLVI)}$$

As was pointed out above, such candidates come with assurances about their behavior under coordinate transformations (here designated as **TRAFO**). They inherit these transformation properties from their progenitor, the fourth-rank Riemann tensor. The default setting for this property is general covariance (here designated as **GCOVARIANT**):

$$\mathbf{TRAFO(RIEM)} =_{\mathbf{DEFT}} \mathbf{GCOVARIANT.} \qquad \mathbf{(XLVII)}$$

When relating the frame **RIEM** to the default settings **GRAV** or **NORM** for **OP** suggested by the correspondence and conservation principles (cf. eq. (XXXIV) and eq. (XL)), one typically finds a relation of the form:

$$\mathbf{RIEM(POT) = GRAV(POT) + DIST(POT),} \qquad \mathbf{(XLVIII)}$$

where **DIST(POT)** represents “disturbing” terms incompatible with the requirements of the correspondence and conservation principles. To obtain from **RIEM(POT)** a “reduced” candidate satisfying these principles one has to impose the revised default setting for the left-hand side of the gravitational field equation:

$$\mathbf{OP(POT)} =_{\mathbf{DEFT}} \mathbf{RIEM_{RED}(POT) = RIEM(POT) - DIST(POT),} \qquad \mathbf{(XLIX)}$$

which can be obtained from **RIEM** either by requiring that

$$\mathbf{DIST(POT) = 0.} \qquad \mathbf{(L)}$$

or by requiring that **DIST(POT)** behaves as a tensor under some group of coordinate transformations, in which case it can be subtracted leaving a reduced candidate invariant under this now restricted group of transformations:

$$\mathbf{TRAFO(RIEM_{RED}) \Rightarrow TRAFO(DIST),} \qquad \mathbf{(LI)}$$

Conditions such as (XLVII) can typically be derived from first-order conditions on the metric tensor, corresponding to a restriction of the admissible coordinate systems (here designated as **COORD(POT)**):

$$\mathbf{COORD(POT) = 0 \Rightarrow DIST(POT) = 0.} \qquad \mathbf{(LII)}$$

Such a coordinate restriction comes in turn with its own transformation behavior, but typically is at least covariant at least under linear transformations:

$$\mathbf{TRAFO(COORD)} =_{\mathbf{DEFT}} \mathbf{LINEAR.} \qquad \mathbf{(LIII)}$$

Coordinate systems selected in this way assumed for Einstein the role of privileged reference frames, similar to the distinguished role of inertial reference systems in classical physics. It is in these preferred coordinate systems that the physical laws are supposedly valid in their usual form. The condition **COORD(POT) = 0** represented for him a true limitation of the generalized relativity principle and is therefore referred to here as a "coordinate restriction."

From a modern perspective, the relation between a generally-covariant candidate for the left-hand side of the field equation and the condition expressed by eq. (LII) can be interpreted in an entirely different way: Since the Newtonian theory clearly does not hold in arbitrary coordinate systems, while generally-covariant field equations do, special coordinates have to be introduced to obtain the Newtonian limit. A *coordinate condition* in the modern sense, however, does not have the meaning of an overall restriction on the choice of admissible coordinates; it is only a tool adapted for this specific purpose. This tool in no way imposes a restriction on the covariance of the field equation, but is available precisely *because* of it. For the Einstein of the Zurich Notebook, however, it was more natural to think of eq. (LII) as a coordinate *restriction,* valid not only in the context of a special situation such as that of the Newtonian limit but necessary in general to ensure that the candidate gravitation tensor takes on the canonical form of eq. (XXXIV). The transformation properties of $\mathbf{RIEM_{RED}(POT)}$ are thus constrained by those of the coordinate restriction, a relation we can express as:

$$\mathbf{TRAFO(RIEM_{RED}) => TRAFO(COORD).} \qquad \mathbf{(LIV)}$$

An additional restriction of the generalized relativity principle typically follows from the conservation principle, given that its mathematical implementation (e.g., by eq. (XLIII)) does, in general, not lead to a generally-covariant equation:

$$\mathbf{TRAFO(DIV(NORM)) \neq GCOVARIANT.} \qquad \mathbf{(LV)}$$

Just as with the correspondence principle (cf. eq. (LII)), the condition **DIV(NORM) = 0** may be inferred from a simpler, possibly first-order condition representing the restriction to coordinate systems in which the conservation principle holds:

$$\mathbf{ENERG(POT) = 0 => DIV(NORM) = 0.} \qquad \mathbf{(LVI)}$$

For the transformation properties of the gravitational field equation we thus have similarly:

$$\mathbf{TRAFO(NORM) => TRAFO(ENERG),} \qquad \mathbf{(LVII)}$$

or, taken together with relation (LIV), replacing $\mathbf{RIEM_{RED}}$ and **NORM** by **GRAV**:

$$\mathbf{TRAFO(GRAV) => TRAFO(COORD) + TRAFO(ENERG).} \qquad \mathbf{(LVIII)}$$

This relation expresses that the transformation properties of the left-hand side of the gravitational field equation are restricted by the needs of the correspondence and the conservation principles taken together. In summary:

$$\mathbf{TRAFO(GRAV) <=>}$$

$$\mathbf{TRAFO(RIEM) + TRAFO(COORD) + TRAFO(ENERG).} \qquad \mathbf{(LIX)}$$

What this symbolic equation says is that the transformation properties of the field equation are known if those of the original default setting rooted in mathematical knowledge are given together with those of the coordinate restrictions imposed to satisfy the correspondence and the conservation principles.

The above considerations raise the more general problem of the compatibility between the mathematical implementations of the correspondence and conservation principles (the corresponding compatibility condition is designated here as **CC-COMP(GRAV)**). While this question could typically be dealt with at the level of the compatibility of the respective coordinate restrictions eq. (LII) and eq. (LVI), it was conceivable that the compatibility requirement gave rise to new conditions with implications not only for the transformation properties of the field equation but for other questions as well, including the question of whether the given default setting for **GRAV** was acceptable at all:

$$\mathbf{CC\text{–}COMP(GRAV) => (COORD = 0) + (ENERG = 0).} \tag{LX}$$

It was also conceivable that a conflict between a candidate for **GRAV** and the correspondence and conservation principles arose because the default setting for the metric of static gravitational fields eq. (25) was incompatible with one of the coordinate restrictions following from these principles:

$$\mathbf{COORD(POT_{STAT}) \neq 0,} \tag{LXI}$$

$$\mathbf{ENERG(POT_{STAT}) \neq 0.} \tag{LXII}$$

4.7 Implications of the Lagrange Formalism

At some point in his research, Einstein realized the significance of the Lagrange formalism not only for formulating the equation of motion but also for deriving the field equation.[76] Because of its earlier application in the context of classical electromagnetic field theory this formalism came with its own default-settings, which played an important role in Einstein's search for the gravitational field equation. Classical field theory suggested, in particular, to choose a Lagrangian quadratic in the field:

$$\mathbf{LAGRANGE =_{DEFT} FIELD^2.} \tag{LXIII}$$

The Lagrangian for the free Maxwell field for instance is of this form (cf. eq. (11)):

$$L = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu}. \tag{48}$$

The use of the Lagrange formalism had two immediate advantages for Einstein. First, when following the physical strategy, he could focus on a scalar object, the Lagrangian, to explore the transformation properties of his theory rather than on the

76 The first paper in which he made use of the Lagrangian formalism for this purpose is (Einstein and Grossmann 1914). This approach was fully developed in (Einstein 1914a).

more complex tensorial objects representing candidates for the left-hand side of the field equation:[77]

$$\textbf{TRAFO(GRAV)} \Leftrightarrow \textbf{TRAFO(LAGRANGE)}. \qquad \textbf{(LXIV)}$$

Second, when following the mathematical strategy, he could rely on an expression for **FIELDMASS** directly delivered by this formalism in terms of the Lagrangian to explore the validity of the conservation principle. The formalism produces a field equation which can easily be brought into a form corresponding to the default settings eq. (XLIV) and eq. (XLV):

$$\textbf{DIV(FIELD) = ENEMO + FIELDMASS}. \qquad \textbf{(LXV)}$$

It remains, of course, to be checked in each concrete case whether the resulting expression for **FIELDMASS** is compatible with the expectation for such an expression following from the conservation principle and, in particular, with eq. (XLII).[78]

The introduction of the Lagrange formalism had one further consequence for Einstein's search, which eventually turned out to be decisive for identifying the gravitational field equation of general relativity. Due to the default setting eq. (LXIII), the Lagrange formalism helped to highlight the importance of the **FIELD**-frame, pointing to the alternative between eq. (XXII) and eq. (XXIII), one leading to the non-covariant *Entwurf* theory, the other to an essentially generally-covariant theory which quickly opened up the pathway toward the field equation of general relativity.

5. TESTING THE CANDIDATES: EINSTEIN'S CHECK LIST FOR GRAVITATION TENSORS

The reservoir of candidates for the left-hand side of the field equation **OP** available to Einstein was determined by the mathematical knowledge available to him. Roughly three levels of knowledge can be distinguished, each coming with its own set of candidates as shown in Fig. 4 below. Not all candidates played the same prominent role in Einstein's research. The four most important ones were the *Entwurf* operator, the Ricci tensor, the Einstein tensor, and the November tensor (Einstein 1915a).

Einstein examined these four differential operators twice in the course of two exploratory phases of his work. He first confronted them with his heuristic requirements in the period documented by the Zurich Notebook dating from the winter 1912–1913, and then once more in the fall of 1915, as documented by publications and correspondence. He came to different conclusions in these two stages of his work. Before we discuss in detail in which way his research experience led him to

---

77 Einstein's point of view was, however, criticized by the mathematician Tullio Levi-Civita, who contested that the Euler-Lagrange equations have the same covariance group as the Lagrangian in the case of the *Entwurf* theory. See, e.g., Tullio Levi-Civita to Einstein, 28 March 1915 (CPAE 8, Doc. 67).

78 For the detailed mathematical considerations, see "Untying the Knot ..." (in vol. 2 of this series), secs. 3.1 and 3.2.

these different views, we shall systematically examine how the various candidates fare when confronted with his heuristic requirements and give an overview of the results of his checks. In this way, we shall be able to establish the potential of these candidates independently of their actual role in the dramatic history of Einstein's search for the field equation. As a consequence, the twists and turns of this search will become understandable as reactions to the epistemic constraints and potentials inherent in the knowledge resources available to Einstein. These constraints are largely embodied in the mental models and frames guiding his research, as well as in their default settings and the instantiations of their open slots. But the conflicting implications of these default settings and instantiations were only revealed in the course of Einstein's elaboration of his theory on the level of concrete mathematical

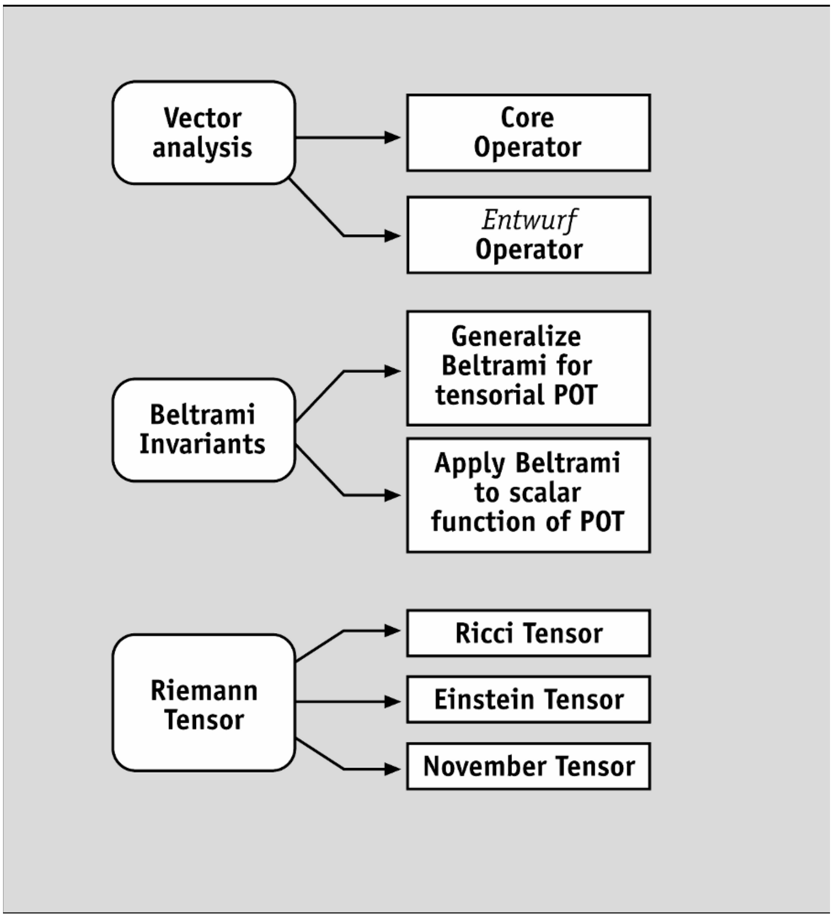


Figure 4: A list of Einstein's most important candidates for differential operators to fill the operator slot of his mental model of a gravitational field equation.

representation. As a matter of fact, the fate of a candidate not only depended on the structural constraints of his heuristics but also on the *order* in which these structures were implemented, on the *exploration depth* with which they were treated, and on the *perspective* under which Einstein examined the answers to his questions. As we shall show in more detail in the next section, in all three of these *performative dimensions* of his evaluation of candidates, the situation of the winter of 1912–1913 was very different from that of October and November 1915 when he once more examined these candidates.

5.1 The Entwurf Operator and the Correspondence Principle in the Winter of 1912–1913

The *Entwurf* operator, first written down in the Zurich Notebook and then published by Einstein and Grossmann in the spring of 1913, gives rise to the field equations:

$$\mathbf{ENTWURF} =_{\mathbf{DEFT}} \frac{1}{\sqrt{-g}}\frac{\partial}{\partial x^{\alpha}}\left(g^{\alpha\beta}\sqrt{-g}\frac{\partial g^{\mu\nu}}{\partial x^{\beta}}\right) - g^{\alpha\beta}g_{\tau\rho}\frac{\partial g^{\mu\tau}}{\partial x^{\alpha}}\frac{\partial g^{\nu\rho}}{\partial x^{\beta}} = \kappa(T^{\mu\nu} + t^{\mu\nu}) \quad (49)$$

with the following expression for the gravitational energy-momentum:

$$\mathbf{FIELDMASS} =_{\mathbf{DEFT}} -\kappa t^{\mu\nu} = \frac{1}{2}g^{\mu\alpha}g^{\nu\beta}\frac{\partial g^{\tau\rho}}{\partial x^{\beta}}\frac{\partial g_{\tau\rho}}{\partial x^{\alpha}} - \frac{1}{4}g^{\mu\nu}g^{\alpha\beta}\frac{\partial g^{\tau\rho}}{\partial x^{\alpha}}\frac{\partial g_{\tau\rho}}{\partial x^{\beta}}. \quad (50)$$

The *Entwurf* field equations satisfy the correspondence principle just as the core operator does since the correction terms which distinguish the two vanish in the limiting procedure for obtaining the Newtonian theory. One has, in particular, (cf. eq. (XXIX)):

$$\mathbf{LIM_{STAT}(ENTWURF) = LIM(LAP(POT_{STAT}))}. \quad \mathbf{(LXVI)}$$

5.2 The Entwurf Operator and the Conservation Principle in the Winter of 1912–1913

By their very construction, the *Entwurf* field equations satisfy the conservation principle since the correction terms distinguishing them from the core operator are generated in such a way that an identity of type (XXXVII) holds. Evidently, the field equations (49) are of the form (XXXIX), while an equation of the form (XLII) expresses the conservation principle:

$$\frac{\partial}{\partial x^{\nu}}(\boldsymbol{T}^{\mu\nu} + \boldsymbol{t}^{\mu\nu}) = 0. \quad (51)$$

5.3 The Entwurf Operator and the Generalized Relativity Principle in the Winter of 1912–1913

The principal challenge for the *Entwurf* theory was the question of the transformation properties of the *Entwurf* field equations and hence of the extent to which the theory satisfies the generalized relativity principle. The *Entwurf* operator had *not* been obtained from a generally-covariant object along the mathematical strategy, (cf. eq. (XLIX)). By construction, the *Entwurf* operator is covariant only under linear transformations (cf. eq. (LIII)):

$$\textbf{TRAFO(ENTWURF)} =_{\textbf{DEFT}} \textbf{LINEAR.} \qquad \textbf{(LXVII)}$$

In different stages of Einstein's work during the reign of the *Entwurf* theory, i.e., between the winter of 1912–1913 and the fall of 1915, he took different positions on the question of whether or not the theory admits a wider class of coordinate transformations. These positions ranged from the acceptance that the *Entwurf* theory is covariant only under linear transformations to the belief that it fully complies with the demands of a generalized relativity principle. Einstein at first believed that the issue of the transformation properties of the *Entwurf* equations was wide open and could be settled only by an extensive mathematical investigation. In the summer of 1913, however, he came to the conclusion that a mere inspection of the form of eq. (51) was sufficient to resolve the problem in favor of the claim that the *Entwurf* theory could *only* be covariant under linear transformations.[79] He thus accepted that the conservation principle requires a severe limitation of the generalized relativity principle.

5.4 The Entwurf Operator and the Correspondence Principle in the Fall of 1915

In the course of his elaboration of the *Entwurf* theory, Einstein succeeded in deriving the field equations from a Lagrange formalism with the default setting for the field given by eq. (XXII). The field then enters the Lagrangian in the form of eq. (LXIII). After an initial attempt to select this default setting for the field with the help of a consistency argument involving the conservation principle (see below), he returned to the correspondence principle as the main argument for choosing, among several options to specify the field variable, the default setting eq. (XXII), giving rise to the familiar *Entwurf* field equation.

5.5 The Entwurf Operator and the Conservation Principle in the Fall of 1915

In the course of his elaboration of the *Entwurf* theory, Einstein succeeded in bringing its field equation into the canonical form described by eqs. (XLIV), (XLV) with the condition (XLIII), a form that was expected on the basis of classical field theory:

79 Cf. Einstein to H.A. Lorentz, 16 August 1913, (CPAE 5, Doc. 470, Norton 1984, 126).

$$\textbf{ENTWURF} =_{\textbf{DEFT}} \frac{\partial}{\partial x^\alpha}(\sqrt{-g}g^{\alpha\beta}\tilde{\Gamma}^\lambda_{\mu\beta}) = \kappa(T^\lambda_\mu + t^\lambda_\mu), \tag{52}$$

with

$$\textbf{FIELDMASS} =_{\textbf{DEFT}} \kappa t^\lambda_\mu = \sqrt{-g}\Big(g^{\lambda\rho}\tilde{\Gamma}^\alpha_{\tau\mu}\tilde{\Gamma}^\tau_{\alpha\rho} - \frac{1}{2}\delta^\lambda_\mu g^{\rho\tau}\tilde{\Gamma}^\alpha_{\beta\rho}\tilde{\Gamma}^\beta_{\alpha\tau}\Big). \tag{53}$$

Here $\tilde{\Gamma}^\alpha_{\sigma\nu}$ represents the default setting for the field as given by eq. (XXII). In 1914 Einstein erroneously believed that a compatibility requirement resulting from the conservation principle and the generalized relativity principle (cf. eq. (87) below) would uniquely fix the *Entwurf* Lagrangian. However, this requirement merely corresponds to demanding the compatibility between **FIELDMASS** in the sense of eq. (XLV) and **FIELDMASS** in the sense of eq. (XLII) and does not substantially restrict the choice of possible gravitation tensors.[80]

5.6 The Entwurf Operator and the Generalized Relativity Principle in the Fall of 1915

Einstein quickly discovered that his argument based on the form of eq. (51) was fallacious since the energy-momentum expression of the gravitational field does not represent a generally-covariant tensor.[81] But he soon found another, seemingly powerful argument in order to justify the *Entwurf* theory's lack of general covariance, the so-called hole argument.[82] To identify the covariance group of the *Entwurf* field equations compatible with this argument, Einstein again made use of eq. (51) but now in a different way which corresponds to the conservation compatibility check as represented by eq. (XLIII), i.e. he combined energy-momentum conservation with the gravitational field equation in order to derive a condition for the class of admissible coordinate systems. By exploring the transformation properties of the Lagrangian (cf. eq. (LXIV), Einstein and Grossmann (1914) claimed to have shown that this condition is both necessary and sufficient (cf. eqs. (LV) and (XLIV)):

$$\textbf{TRAFO(NORM}_{\textbf{CLASS}}\textbf{)} \Leftrightarrow \textbf{TRAFO(DIV(NORM}_{\textbf{CLASS}}\textbf{))} \tag{LXVIII}$$

with

$$\textbf{DIV(NORM}_{\textbf{CLASS}}\textbf{)} =_{\textbf{DEFT}} B_\mu = \frac{\partial^2}{\partial x^\nu \partial x^\alpha}(\sqrt{-g}g^{\alpha\beta}\tilde{\Gamma}^\nu_{\sigma\beta}) = 0. \tag{54}$$

These four third-order differential equations for the metric tensor complement the ten gravitational field equations and embody the conditions enforcing the restriction of general covariance characteristic of the theory. They were understood by Einstein and

80 See "Untying the Knot ..." (in vol. 2 of this series), sec. 3.

81 Cf. the footnote in (Einstein and Grossmann 1914, 218).

82 For historical discussion, see (Earman and Norton 1987, Stachel 1989b) and "What Did Einstein Know ..." (in vol.2 of this series) as well as further references cited therein.

Grossmann as determining the coordinate systems "adapted" to the *Entwurf* theory.[83] It was difficult to see exactly which transformations to accelerated coordinate systems are admitted by these conditions.

5.7 The Ricci Tensor and the Correspondence Principle in the Winter of 1912–1913

The generally-covariant Ricci tensor, first taken into consideration by Einstein in the Zurich Notebook (see p. 22R), can be expressed in terms of the Christoffel symbols (cf. eqs. (30) and (XLVI)) as:

$$\mathbf{RIEM} =_{\mathbf{DEFT}} \mathbf{RICCI} =_{\mathbf{DEFT}} R_{\mu\nu} = \frac{\partial\Gamma^{\alpha}_{\mu\nu}}{\partial x^{\alpha}} - \frac{\partial\Gamma^{\alpha}_{\mu\alpha}}{\partial x^{\nu}} + \Gamma^{\beta}_{\nu\beta}\Gamma^{\alpha}_{\mu\alpha} - \Gamma^{\beta}_{\mu\nu}\Gamma^{\alpha}_{\beta\alpha}. \tag{55}$$

The validity of the correspondence principle could be examined by bringing **RICCI** into the form (cf. eqs. (XXXIV) and (XLIX))

$$\mathbf{RIEM} =_{\mathbf{DEFT}} \mathbf{RICCI}_{\mathbf{RED}} = \mathbf{LAP(POT)} + \mathbf{CORR(POT)}, \tag{LXIX}$$

and by checking whether (cf. eq. (L))

$$\mathbf{DIST(POT) = 0.}$$

In the Zurich Notebook Einstein identified the relevant terms as:

$$\mathbf{DIST(POT)} =_{\mathbf{DEFT}} \frac{\partial^2 g_{\alpha\alpha}}{\partial x_{\mu}\partial x_{\nu}} - \frac{\partial^2 g_{\mu\alpha}}{\partial x_{\alpha}\partial x_{\nu}} - \frac{\partial^2 g_{\nu\alpha}}{\partial x_{\alpha}\partial x_{\mu}}. \tag{56}$$

The vanishing of these disturbing terms can be achieved by imposing a set of four *first-order* partial differential equations for the metric tensor which is given by (cf. eq. (LII)):

$$\mathbf{COORD(POT)} =_{\mathbf{DEFT}} \mathbf{COORD}_{\mathbf{HARM}}\mathbf{(POT)} =_{\mathbf{DEFT}} g^{\lambda\kappa}\Gamma^{\mu}_{\lambda\kappa} = 0. \tag{57}$$

---

83 This expression is chosen to resolve the ambiguity of German expressions which may be translated by "condition" as well as by "restriction." Cf. the formulations in (Einstein and Grossmann 1914): "understood [...], that an acceptable theory of gravitation implies necessarily a specialization of the coordinate system." ("eingesehen [...], daß eine brauchbare Gravitationstheorie notwendig einer Spezialisierung des Koordinatensystems bedarf [...]", p. 218); "restriction" ("Einschränkung", p. 218, note); "true condition" ("wirkliche Bedingung", p. 219); "conditions [...], by which we restricted the coordinate systems" ("Bedingungen [...], durch die wir die Koordinatensysteme eingeschränkt haben.", p. 225). In a letter to Michele Besso, ca. 10 March 1914, (CPAE 5, Doc. 514), Einstein comments on eq. (54): "These are 4 third-order equations for the [...] or [...], which can be conceived as the conditions for the special choice of the reference system." ("Dies sind vier Gleichungen dritter Ordnung für die $g_{\mu\nu}$ [...], welche man als die Bedingungen für die spezielle Wahl des Bezugssystems auffassen kann.")

These equations, representing the "harmonic" coordinate restriction, were interpreted by Einstein as singling out a particular class of coordinate systems that were then called "isothermal" and are now referred to as "harmonic" coordinates.

The reduced Ricci tensor $\mathbf{RICCI_{RED}}$ suffered from yet another problem related to Einstein's understanding of the correspondence principle. It consists in a conflict between the harmonic coordinate restriction and the canonical metric for a static gravitational field $\mathbf{POT_{STAT}}$ (cf. eq. (25)):

$$\mathbf{COORD_{HARM}(POT_{STAT}) \neq 0.} \qquad \mathbf{(LXX)}$$

However, as far as the available evidence from the Zurich Notebook and other contemporary sources show, this argument played no role in evaluating the reduced Ricci tensor.[84]

5.8 The Ricci Tensor and the Conservation Principle in the Winter of 1912–1913

As far as the conservation principle is concerned, the exploration depth reached in the Zurich Notebook was characterized by the fact that Einstein examined only the weak-field equation following from a gravitational field equation based on the Ricci tensor. He considered, in other words, an equation of the type of eq. (33). For such a weak-field equation in which the source is given by pressureless dust (cf. eq. (4)), Einstein succeeded in representing the force exerted by the gravitational field as a divergence expression in the sense of eq. (XXXVI)):

$$\mathbf{-\,GRAD(POT)\ X\ LIM(RICCI) = DIV(LIM(FIELDMASS)),} \qquad \mathbf{(LXXI)}$$

which in his notation reads:[85]

$$\sum_{\kappa i m} \gamma_{\kappa\kappa} \frac{\partial^2 g_{im}}{\partial x_\kappa^2} \frac{\partial g_{im}}{\partial x_\sigma} = \sum_{\kappa i m} \gamma_{\kappa\kappa} \left[\frac{\partial}{\partial x_\kappa} \left(\frac{\partial g_{im}}{\partial x_\kappa} \frac{\partial g_{im}}{\partial x_\sigma} \right) - \frac{1}{2} \frac{\partial}{\partial x_\sigma} \left(\frac{\partial^2 g_{im}{}^2}{\partial x_\kappa} \right) \right]. \qquad (58)$$

The conservation compatibility check similarly takes on a simpler form if considered for the weak field case. In first-order approximation the covariant derivative in eq. (XXXI) can be replaced by an ordinary derivative and **OP(POT)** by **LAP(POT)** with its default setting according to eq. (XXVIII) so that this condition can be written, in Einstein's notation, as:

$$\mathbf{LIM(DIV_{COV}(OP)) = DIV(LIM(LAP)= 0.} \qquad \mathbf{(LXXII)}$$

Interchanging the two differential operations,

---

84 See "Untying the Knot ..." (in vol. 2 of this series), fn. 12 for further discussion.

85 See p. 19R of "Einstein's Zurich Notebook" and sec. 5.4.2 of the "Commentary" (in vol. 2 of this series), fn. 10 for further discussion. Einstein's notation, which is somewhat sloppy, is explained in detail in the commentary; note that he used an imaginary time coordinate and that the terms $g_{\mu\nu}$ here stand for the small deviations $h_{\mu\nu}$ from the covariant Minkowski metric.

DIV(LIM(LAP(POT)) = LIM(LAP(DIV(POT))), (LXXIII)

which, in Einstein's notation amounts to:

$$\frac{\partial}{\partial x_m}(\Box g_{im}) = \Box\left(\frac{\partial g_{im}}{\partial x_m}\right) = 0, \tag{59}$$

it becomes clear that the conservation compatibility check is satisfied at the weak-field level if an appropriate set of first-order conditions hold in the sense of eq. (LVI):

LIM(ENERG) =$_{\mathbf{DEFT}}$ DIV(POT) = 0 => DIV(LIM(LAP)) = 0. (LXXIV)

More specifically, the conservation compatibility check works out in the weak field limit if the condition:

$$\textbf{LIM(ENERG)} =_{\textbf{DEFT}} \textbf{COORD}_{\textbf{HERTZ}} = \textbf{DIV(POT)} =_{\textbf{DEFT}} \frac{\partial g_{im}}{\partial x_m} = 0 \tag{60}$$

is fulfilled. This condition was mentioned by Einstein in a letter to Paul Hertz from 22 August 1915[86] and will therefore be called the "Hertz condition" or the "Hertz restriction" depending on the context. In the case at hand, it is appropriately referred to as the "Hertz restriction" since it represents a restriction of the admissible coordinates required by the conservation principle.

As it turned out, the combination of the two coordinate restrictions eq. (57) and eq. (60), resulting from the correspondence and the conservation principle, and the weak-field field equation imposed a restriction which Einstein considered to be unacceptable. According to this condition, the trace of the source term has to vanish, which can be expressed in terms of eq. (LX) as:

TRACE(SOURCE) = 0 => CC-COMP(LIM(LAP)). (LXXV)

This condition was indeed incompatible with the default setting for the source term of the gravitational field equation, pressureless dust (cf. eq. (XXI)). Combining restrictions eq. (57) and eq. (60) furthermore implies that the trace of the potential must be constant which is obviously in conflict with the default setting for the metric of a static field eq. (25).[87]

5.9 The Ricci Tensor and the Generalized Relativity Principle in the Winter of 1912–1913

Given the compatibility problem just described, the transformation properties of the reduced Ricci tensor remained unexplored.

86 For a detailed discussion of this letter, see (Howard and Norton 1993).
87 See "Commentary ..." (in vol. 2 of this series), sec. 5.4.3 for detailed discussion.

5.10 The Ricci Tensor and the Correspondence Principle in the Fall of 1915

When Einstein returned to the Ricci tensor in November 1915 both the perspective and the exploration depth of his investigation had changed. He then considered the Ricci tensor in coordinate systems with:

$$\sqrt{-g} = 1 \tag{61}$$

in which it takes on the simpler form:

$$\textbf{RICCI} =_{\textbf{DEFT}} \frac{\partial \Gamma^{\alpha}_{\mu\nu}}{\partial x^{\alpha}} + \Gamma^{\beta}_{\mu\alpha}\Gamma^{\alpha}_{\nu\beta}. \tag{62}$$

Within this framework, the Ricci tensor could be brought into the appropriate weak-field form eq. (XXXIV) by assuming the Hertz condition:

$$\textbf{COORD}_{\textbf{HERTZ}} = \textbf{DIV(POT)} =_{\textbf{DEFT}} \frac{\partial g_{im}}{\partial x_m} = 0. \tag{63}$$

In November 1915, Einstein was aware of the fact that it was sufficient for satisfying the correspondence principle to use such an equation (cf. eq. (LII)) in the modern sense of a coordinate *condition* that simply makes use of the freedom within a generally-covariant framework to pick appropriate coordinate frames—without imposing an overall restriction. In this sense, Einstein's understanding of the correspondence principle had been substantially enhanced by a greater exploration depth of his formalism.[88]

In contrast to eq. (LXX) we now have:

$$\textbf{COORD}_{\textbf{HERTZ}}(\textbf{POT}_{\textbf{STAT}}) = \textbf{0}, \tag{LXXVI}$$

so that the conflict between the coordinate condition and the canonical metric for a static field is apparently removed. This is, in any case, what Einstein at first must have believed when he published, in November 1915, a gravitational field equation based on the Ricci tensor. What he seems to have overlooked, however, was the fact that his canonical metric given by eq. (25) was incompatible with the condition eq. (61) on which his entire framework, including the coordinate condition eq. (63), crucially depended. In other words, the available evidence suggests that Einstein had first published his theory based on the Ricci tensor although it actually violates the correspondence principle as he then conceived it.

He only realized the challenge represented by the choice of the Ricci tensor for his understanding of the correspondence principle when he examined the implication of this choice for the explanation of Mercury's perihelion motion, an examination that gave him a nearly perfect match with the observational data.[89] Einstein at first

88 See "Untying the Knot ..." (in vol. 2 of this series), secs. 1.5 and 6 for detailed discussion.

interpreted this agreement as evidence in favor of his hypothesis of an electromagnetic theory of matter which had made the proposal of a field equation with the Ricci tensor as its left-hand side acceptable to him (see below).

5.11 The Ricci Tensor and the Conservation Principle in the Fall of 1915

By the fall of 1915, the exploration depth of Einstein's investigation had been increased, in particular, by the development of a technique allowing him to derive a gravitational energy-momentum expression **FIELDMASS** for the full field equation from the Lagrange formalism, if the coordinate condition eq. (61) is assumed and the default setting for the field is given by eq. (XXIII). He was thus able to bring a field equation based on the Ricci tensor into a form corresponding to eq. (XXXIX) with the conservation equation (XLII):

$$\textbf{NORM(POT)} =_{\textbf{DEFT}} (g^{\nu\lambda}\Gamma^{\alpha}_{\mu\nu})_{,\alpha} - \frac{1}{2}\delta^{\lambda}_{\mu}g^{\rho\sigma}\Gamma^{\alpha}_{\beta\rho}\Gamma^{\beta}_{\alpha\sigma} = -\kappa(T^{\lambda}_{\mu} + t^{\lambda}_{\mu}), \tag{64}$$

with

$$\textbf{FIELDMASS} =_{\textbf{DEFT}} \kappa t^{\lambda}_{\sigma} = \frac{1}{2}\delta^{\lambda}_{\sigma}g^{\mu\nu}\Gamma^{\alpha}_{\beta\mu}\Gamma^{\beta}_{\alpha\nu} - g^{\mu\nu}\Gamma^{\alpha}_{\mu\sigma}(\Gamma_{\alpha\nu},)^{\lambda} \tag{65}$$

and

$$\textbf{DIV(ENEMO + FIELDMASS)} =_{\textbf{DEFT}} (T^{\lambda}_{\mu} + t^{\lambda}_{\mu})_{,\lambda} = 0. \tag{66}$$

Einstein, however, did not manage to comply with the requirement expressed by the default setting eq. (XLIV). Bringing the left-hand side of the field equation into the form of eq. (XLIV) would result in a formulation in which the right-hand side no longer satisfies the default setting eq. (XLI) for **SOURCE**:

$$\textbf{NORM(POT)}_{\textbf{CLASS}} = \textbf{DIV(FIELD)} =_{\textbf{DEFT}} (g^{\nu\lambda}\Gamma^{\alpha}_{\mu\nu})_{,\alpha} = -\kappa\left(T^{\lambda}_{\mu} + t^{\lambda}_{\mu} - \frac{1}{2}\delta^{\lambda}_{\mu}t\right) \tag{67}$$

with

$$\textbf{SOURCE} =_{\textbf{DEFT}} \textbf{ENEMO + FIELDMASS} \neq_{\textbf{DEFT}} -\kappa\left(T^{\lambda}_{\mu} + t^{\lambda}_{\mu} - \frac{1}{2}\delta^{\lambda}_{\mu}t\right). \tag{68}$$

Equation (66) could be used to perform the conservation compatibility check in a straightforward manner. Einstein succeeded in showing that this check turned out successful if the trace of the energy-momentum tensor vanishes (cf. eq. (LXXV))—without imposing any further conditions on the admissible coordinate systems:

TRACE(SOURCE) = 0 => DIV(ENEMO + FIELDMASS) = 0. (LXXVII)

The odd assumption of a vanishing trace, violating the default assumption eq. (XXI), was now acceptable to Einstein since both the exploration depth of his investi-

89 See (Einstein 1915b) and for historical discussion, (Earman and Janssen 1993).

gation and his perspective had changed. He now reexamined the Ricci tensor from the perspective of an electromagnetic theory of matter in which this condition was fulfilled from the outset, given that the trace of the electromagnetic energy-momentum tensor vanishes.[90]

5.12 The Ricci Tensor and the Generalized Relativity Principle in the Fall of 1915

The field equations based on the Ricci tensor as formulated by Einstein in the fall of 1915 represents, according to his heuristic criteria, a complete implementation of the generalized principle of relativity. The conservation compatibility check for these field equations had given Einstein, as we have seen, merely a condition on the trace of the energy-momentum tensor which does not imply any restriction on the choice of coordinate systems. As a consequence, there no longer was any conflict between the conservation and the correspondence principles as he had encountered it in the winter of 1912–1913. It was thus possible to impose either the harmonic coordinate condition eq. (57) or the combination of eqs. (61) and (63) in order to reduce the Ricci tensor to the canonical weak field form eq. (LXIX) from which the Newtonian limit could be obtained—at least if the objection resulting from eq. (LXX) could be solved or circumvented.

5.13 The Einstein Tensor and the Correspondence Principle in the Winter of 1912–1913

The generally-covariant Einstein tensor, first taken into consideration, albeit only in the weak-field approximation, in the Zurich Notebook, can be expressed in terms of the Ricci tensor $R_{\mu\nu}$ and its trace R (cf. eq. (55)) as:

$$\textbf{RIEM} =_{\textbf{DEFT}} \textbf{Einstein} =_{\textbf{DEFT}} E_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R. \tag{69}$$

A field equation based on the Einstein tensor may also be written by shifting the trace term to the right-hand side by a simple mathematical argument. The equation then reads

$$R_{\mu\nu} = -\kappa\left(T_{\mu\nu} - \frac{1}{2} g_{\mu\nu} T\right), \tag{70}$$

where T is the trace of the energy-momentum tensor $T_{\mu\nu}$. Here the Ricci tensor again appears on the left-hand side as the differential operator acting on the metric tensor.

---

90 See (Einstein 1915d) where the consequence is called "introducing an admittedly bold additional hypothesis on the structure of matter." ("Einführung einer allerdings kühnen zusätzlichen Hypothese über die Struktur der Materie", p. 799).

The exploration level of the first examination of the Einstein tensor in the winter of 1912–1913 was, just as that of Einstein's analysis of the Ricci tensor, characterized by a focus on the weak-field equations and the assumption that the correspondence principle could only be satisfied by a coordinate restriction. Given that the Einstein tensor results from a modification of the Ricci tensor according to eq. (69), it was natural to presuppose the harmonic coordinate restriction **COORD$_{\mathbf{HARM}}$(POT) = 0** (cf. eq. (57)). As a matter of fact, in the winter of 1912–1913 the Einstein tensor was obtained directly by an ad hoc modification of the weak-field form of the gravitational field equation eq. (33), resulting in:[91]

$$\Box\left(g_{i\kappa}-\frac{1}{2}\delta_{i\kappa}U\right) = T_{i\kappa} \tag{71}$$

with the trace term:

$$U = \sum g_{\kappa\kappa}, \tag{72}$$

or alternatively as:

$$\Box g_{ij} = T_{ij}-\frac{1}{2}\delta_{ij}\left(\sum T_{\kappa\kappa}\right). \tag{73}$$

It surely would have been possible for Einstein to carry out the corresponding modification on the level of the original Ricci tensor, turning it into what we now call the Einstein tensor, by the subtraction of a trace term.

On closer inspection, however, the harmonic coordinate restriction does not achieve the desired reduction of the field equation to the required standard form in the sense of eq. (LII). Indeed, if the left-hand side is brought into the canonical form eq. (XXVIII) so that eq. (73) is obtained, the right-hand side does obviously not represent the default setting for **SOURCE** as given by eq. (XIV):

$$\mathbf{SOURCE} \neq_{\mathbf{DEFT}} T_{ij}-\frac{1}{2}\delta_{ij}\left(\sum T_{\kappa\kappa}\right). \tag{LXXVIII}$$

Instead an additional trace term appears which in general is not constant. If one examines, in particular, a static mass distribution as the default setting for **SOURCE** so that the 44 component is the only non-vanishing term of the energy-momentum tensor, it follows from the weak-field equation (73) that *all* diagonal components of the metric tensor will be variable so that one has in general:

$$g_{ii} \neq \text{const for } i = 1, \ldots, 4. \tag{74}$$

As a consequence, the weak-field equation (73) no longer admits the canonical metric **POT$_{\mathbf{STAT}}$** defined by eq. (25) as a solution. At that point in time, Einstein saw no way to avoid this default setting for the potential, and he rejected the Einstein tensor—lin-

91 See "Commentary" (in vol. 2 of this series), sec. 5.4.3.

earized and reduced by the harmonic coordinate restriction—as a candidate for the left-hand side of the gravitational field equation.[92]

5.14 The Einstein Tensor and the Conservation Principle in the Winter of 1912–1913

At the weak-field level, the results of Einstein's check of the conservation principle turned out to be promising. In spite of the additional trace term it was possible to write the gravitational force density in the required form of a divergence of the gravitational energy-momentum density, which in Einstein's notation reads (cf. eqs. (XXXVI) and (LXXI)):

$$\sum_{i\kappa\nu}\frac{\partial}{\partial x_\nu}\left(\frac{\partial g_{i\kappa}}{\partial x_\nu}\frac{\partial g_{i\kappa}}{\partial x_\sigma}\right)-\frac{1}{2}\sum_{i\kappa\nu}\frac{\partial}{\partial x_\sigma}\left(\left(\frac{\partial g_{i\kappa}}{\partial x_\nu}\right)^2\right)$$
$$-\frac{1}{2}\sum_\nu\frac{\partial}{\partial x_\nu}\left(\frac{\partial U}{\partial x_\nu}\frac{\partial U}{\partial x_\sigma}\right)+\frac{1}{4}\sum\frac{\partial}{\partial x_\sigma}\left(\left(\frac{\partial U}{\partial x_\nu}\right)^2\right). \tag{75}$$

At the level of the weak-field equation it was also immediately clear that the conservation compatibility check is no longer in conflict with the correspondence principle, in contrast to what he had found before for the Ricci tensor. In analogy with eq. (59) one now obtains:

$$\frac{\partial}{\partial x_\kappa}\left(\Box\left(g_{i\kappa}-\frac{1}{2}\delta_{i\kappa}U\right)\right) = \Box\left(\frac{\partial}{\partial x_\kappa}\left(g_{i\kappa}-\frac{1}{2}\delta_{i\kappa}U\right)\right) = 0, \tag{76}$$

which, in symbolic notation, corresponds to (cf. eqs. (LXXII), (LXXIV)):

$$\textbf{DIV(LIM(EINSTEIN)) = LIM(LAP(COORD}_{\textbf{HARM}}\textbf{)) = 0.} \tag{LXXIX}$$

In other words, the conservation compatibility check is, in the weak-field limit, satisfied *because* of the harmonic coordinate restriction eq. (57) required by the correspondence principle—without imposing any restriction on the trace of the energy-momentum tensor.

5.15 The Einstein Tensor and the Generalized Relativity Principle in the Winter of 1912–1913

Whether or not the generalized relativity principle was satisfied would, according to Einstein's understanding in the winter of 1912–1913, depend on whether the coordinate restrictions necessary to fulfill his other heuristic criteria would leave him enough covariance. In view of the clash between the Einstein tensor and the corre-

92 See "Commentary" (in vol. 2 of this series), sec. 5.4.6.

spondence principle (see eq. (LXXVIII)), this issue remained unexplored at this point in time.

5.16 The Einstein Tensor and the Correspondence Principle in the Fall of 1915

When Einstein returned to the Einstein tensor in November 1915 he focused on coordinate systems with:

$$\sqrt{-g} = 1 . \tag{77}$$

The field equations based on the Einstein tensor then take on the form (cf. eqs (70) and (62)):

$$\Gamma^{\alpha}_{\mu\nu,\alpha} + \Gamma^{\alpha}_{\beta\mu}\Gamma^{\beta}_{\alpha\nu} = -\kappa\left(T_{\mu\nu} - \frac{1}{2}g_{\mu\nu}T\right). \tag{78}$$

In view of eqs. (69) and (70), one could use, just as in the case of the Ricci tensor, the harmonic condition eq. (57)[93] to bring the left-hand side of the field equation into the required form. This condition was now understood as a coordinate condition in the modern sense. Proceeding in this way, one reduces the Einstein field equation for weak fields to an equation of the form (73). The usual transition to the Newtonian theory could now proceed by taking the energy-momentum tensor of dustlike matter as the source and neglecting all terms except the T_{44}-term, which can be identified with the gravitating mass density ρ appearing in the classical Poisson equation (cf. eq. (37)).

What remained to be shown was that the canonical metric for a static field was compatible with the non-standard form of the right-hand side of the weak-field equations (73). Even in 1915 this conflict remained, in a sense, unresolved. The additional trace term on the right-hand side made it impossible to accept the canonical metric for static fields as a solution of the weak field equations since the 11 ... 33 components of the source term had to be retained in the transition to the Newtonian case (cf. eq. (74)). Therefore the correction term in the Einstein tensor made the transition to the Newtonian case *a fortiori* impossible following the procedure suggested by the correspondence principle. All this had been known to Einstein in 1912[94] and remained, of course, true also in 1915, when he took up the Einstein tensor a second time (Einstein 1915d).

But now Einstein was able to circumvent this problem. Even though the field equation failed to satisfy the correspondence principle as hitherto understood, this did not affect the equation of motion. In the weak-field limit of the equation of motion, the non-standard character of the weak-field Einstein equation plays no role. For weak static gravitational fields and for velocities negligible in comparison with that of light, the general equation of motion (31) reduces, as we have seen, to eq. (38).

93 For a comment on the role of the Hertz condition in this context, see Albert Einstein to Karl Schwarzschild, Berlin, 19 February 1916, (CPAE 8, Doc. 194).

94 Cf. pp. 20L–21R of "Einstein's Zurich Notebook" (in this volume).

This equation now implies that, under the conditions assumed, a gravitational field equation based on the Einstein tensor is actually compatible with the experimental data on gravitation that are adequately described by Newton's theory if $g_{44}/2$ is, as usual, identified with the Newtonian potential, while the other components of the metric tensor play no role at this level of the weak-field limit of the Einstein equation.

5.17 The Einstein Tensor and the Conservation Principle in the Fall of 1915

When Einstein returned to the Einstein tensor in late 1915, the greater exploration depth of his investigation made it possible to establish an energy-momentum expression for the gravitational field of the required form. Also the question of the conservation compatibility check could now be addressed in a straightforward manner. He succeeded in bringing the field equation into a form corresponding to eq. (XXXIX) with a conservation equation of the form of eq. (XLII):

$$\textbf{NORM(POT)} =_{\textbf{DEFT}} (g^{\nu\lambda}\Gamma^{\alpha}_{\mu\nu})_{,\alpha} - \frac{1}{2}\delta^{\lambda}_{\mu}\kappa(t+T) = -\kappa(T^{\lambda}_{\mu} + t^{\lambda}_{\mu}). \tag{79}$$

The satisfaction of the conservation compatibility check (cf. eq. (XLIII)) now no longer imposes any additional conditions interfering with the field equation as was the case for the tensor where this check implied that the trace of both sides of the field equation has to vanish.

In the field equation based on the Einstein tensor, the trace terms of the energy-momentum of matter and of the gravitational field enter, in contrast to what happens for the Ricci tensor (cf. eq. (67)), in complete parallel to each other. As a matter of fact, the introduction of these trace terms corresponds to changing the default setting eq. (XLI) for **SOURCE** into:

$$\textbf{SOURCE} =_{\textbf{DEFT}}$$

$$\textbf{(ENEMO – 1/2 TRACE(ENEMO)) +}$$

$$\textbf{(FIELDMASS – 1/2 TRACE(FIELDMASS)).} \tag{LXXX}$$

With this new instantiation for the source term, Einstein now also managed to comply with the expectation for the left-hand side of the field equation expressed by the default setting eq. (XLIV):

$$\textbf{NORM(POT)}_{\textbf{CLASS}} = \textbf{DIV(FIELD)} =_{\textbf{DEFT}} (g^{\nu\lambda}\Gamma^{\alpha}_{\mu\nu})_{,\alpha} \tag{80}$$

with

$$\textbf{SOURCE} =_{\textbf{DEFT}} -\kappa\left(\left(T^{\lambda}_{\mu} - \frac{1}{2}\delta^{\lambda}_{\mu}T\right) + \left(t^{\lambda}_{\mu} - \frac{1}{2}\delta^{\lambda}_{\mu}t\right)\right). \tag{81}$$

Note, however, that with this redefinition of the source-term the field equation no longer corresponds with the canonical expectation for its right-hand side expressed by the default setting eq. (XLII) suggested by the conservation principle. Even for a field equation based on the Einstein tensor it is simply impossible to satisfy all expectations raised by the experience of classical field theory!

5.18 The Einstein Tensor and the Generalized Relativity Principle in the Fall of 1915

Since neither the correspondence nor the conservation principle imposed any further restrictions, field equations based on the Einstein tensor fully implement the generalized relativity principle.

5.19 The November Tensor and the Correspondence Principle in the Winter of 1912–1913

The November tensor, first considered in the Zurich Notebook, can be obtained from the Ricci tensor (cf. eqs. (55) and (62)) by restricting the covariance group to unimodular transformations and then splitting off a term:

$$\mathbf{RIEM} =_{\mathbf{DEFT}} \mathbf{NOVEMBER} =_{\mathbf{DEFT}} N_{\mu\nu} = \frac{\partial\Gamma^{\alpha}_{\mu\nu}}{\partial x^{\alpha}} + \Gamma^{\beta}_{\mu\alpha}\Gamma^{\alpha}_{\nu\beta}. \tag{82}$$

The exploration level of the November tensor in the winter of 1912–1913 was, in general, characterized by a limitation to the weak-field equation and the expectation that the implementation of the correspondence and conservation principles requires a coordinate restriction. The correspondence principle, in particular, can be satisfied if the Hertz restriction eq. (60) is imposed, bringing **NOVEMBER** into the form (cf. eqs. (XXXIV) and (XLIX)):

$$\mathbf{RIEM} =_{\mathbf{DEFT}} \mathbf{NOVEMBER_{RED}} = \mathbf{LAP(POT) + CORR(POT)}, \qquad \mathbf{(LXXXI)}$$

so that also the weak-field equation takes on the canonical form of eq. (33) which can be solved by the canonical metric for a static field given by eq. (25). One now also has (cf. eqs. (LXX) and (LXXVI)):

$$\mathbf{COORD_{HERTZ}(POT_{STAT}) = 0.} \qquad \mathbf{(LXXXII)}$$

5.20 The November Tensor and the Conservation Principle in the Winter of 1912–1913

The weak-field equation for the November tensor has the same form as that obtained from a field equation based on the Ricci tensor since (cf. eq. (XXVIII)):

$$\mathbf{LIM(RICCI) = LIM(NOVEMBER) = LIM(LAP).} \qquad \mathbf{(LXXXIII)}$$

It is clear therefore that the conservation principle holds, at least in the weak-field limit. It is possible to form a divergence expression such as that given by eq. (LXXI) and to satisfy the conservation compatibility check as represented by eq. (LXXII) if the Hertz restriction eq. (60) is imposed. But contrary to the case of the Ricci tensor, the coordinate restrictions required by the correspondence and the conservation principles, respectively, now coincide (cf. eq. (LX)):

$$\mathbf{LIM(ENERG) = COORD_{HERTZ} = 0.} \qquad \mathbf{(LXXXIV)}$$

5.21 The November Tensor and the Generalized Relativity Principle in the Winter of 1912–1913

The check of the generalized relativity principle was eased by the fact that the transformation behavior of the reduced November tensor (cf. eq. (XXXIV)) could be inferred from the transformation properties of the restriction distinguishing it from the original November tensor, the Hertz restriction. Indeed, if the Hertz restriction remains covariant under a given unimodular coordinate transformation so must the reduced November tensor (cf. eq. (LIX)). In the winter of 1912–1913, Einstein examined this transformation behavior for the two cases central to the heuristics governed by the equivalence principle, the case of uniform acceleration ("the elevator") and the case of rotation ("the bucket"). To simplify matters, he considered the case of infinitesimal transformations and found that, while the Hertz restriction is satisfied by infinitesimal rotations, it is not by infinitesimal transformations to a uniformly accelerated system.[95] At least as far as the exploration level of his calculations (limited to the weak-field case) allowed, Einstein could conclude that the reduced November tensor clashes with the equivalence principle, even in the case of infinitesimal transformations. He may well have found that transformations to finite rotations are incompatible with the Hertz restriction as well.

5.22 The November Tensor and the Correspondence Principle in the Fall of 1915

When Einstein returned to the November tensor in 1915, he could make use of the results he had established earlier, in particular with regard to the correspondence principle and how to satisfy that principle by imposing the Hertz restriction (cf. eq. (LXXXI)). His reexamination was, on the other hand, characterized by an increased exploration depth, which allowed him to treat this restriction as a coordinate condition in the modern sense. As we shall see, the conservation principle again leads to a coordinate restriction following from **DIV(NORM) = 0** (cf. eq. (XLIII)) which made it necessary to recheck the compatibility of this condition with the correspondence principle. As it turned out, the conservation compatibility check only gives rise to a weak scalar condition in this case, which in the weak-field limit, can easily be satisfied if the Hertz condition is fulfilled (cf. eq. (LVI)):

95 See "Commentary" (in vol. 2 of this series), sec. 5.5.3 and secs. 4.5.2–4.5.3.

$$\mathbf{LIM(ENERG(POT))} =_{\mathbf{DEFT}} \mathbf{COORD_{HERTZ} = 0 \Rightarrow}$$

$$\mathbf{LIM(DIV(NORM)) = 0.} \qquad \mathbf{(LXXXV)}$$

Under these circumstances, the Hertz condition can thus be considered as a strengthening of the restriction **DIV(NORM) = 0** following from the conservation principle. But as this sharpening turned out to be necessary only for the purpose of implementing the correspondence principle by choosing a class of coordinate systems suitable for this purpose, the Hertz condition could now indeed be interpreted, for the first time, as a *coordinate condition* in the modern sense.

5.23 The November Tensor and the Conservation Principle in the Fall of 1915

In the fall of 1915, Einstein succeeded in deriving a gravitational energy-momentum expression **FIELDMASS** for the full field equation based on the November tensor from a Lagrange formalism in which the default setting for the field is given by eq. (XXIII). He brought the field equation into a form corresponding to eq. (XXXIX), thus obtaining eqs. (64), (65), and (66), familiar from our discussion of the Ricci tensor.

What remained was the check of compatibility with the conservation principle and the question of which coordinate transformations it allowed. This question could now be addressed not just on the weak-field level—where the transformation properties of the Hertz restriction had led to a disappointing answer—but on the level of the full field equation. Einstein succeeded in expressing the conservation compatibility check in terms of an equation of the form of eq. (XLIII) which now, however, has the remarkable property that it represents not four equations but rather follows from a single scalar condition (cf. eq. (LVI)):[96]

$$\mathbf{ENERG(POT)} =_{\mathbf{DEFT}} \mathbf{DIV(SCALAR(POT)) = 0 \Rightarrow DIV(NORM) = 0,(LXXXVI)}$$

where:

$$\mathbf{SCALAR(POT)} =_{\mathbf{DEFT}} \sum_{\alpha\beta} \frac{\partial^2 g^{\alpha\beta}}{\partial x^\alpha \partial x^\beta} - \sum_{\sigma\tau\alpha\beta} g^{\sigma\tau} \Gamma^\alpha_{\sigma\beta} \Gamma^\beta_{\tau\alpha} = 0. \qquad (83)$$

This condition clearly is much less restrictive than the Hertz restriction. As mentioned above, the Hertz restriction could therefore be reinterpreted as a coordinate condition, obtained by strengthening the weak-field version of this scalar condition.

5.24 The November Tensor and the Generalized Relativity Principle in the Fall of 1915

The November tensor was obtained from the generally-covariant Ricci tensor by imposing a restriction to unimodular coordinate transformations. The conservation compatibility check (cf. eq. (83)) gave rise to a further restriction of the choice of

96 See "Untying the Knot ..." (in vol. 2 of this series), sec. 6, eqs. (75)–(78), for detailed discussion.

admissible coordinate systems, the "November restriction," as it might be called. Combining the trace of the full field equation with eq. (83), the following scalar equation results:[97]

$$\frac{\partial}{\partial x^{\alpha}}\left(g^{\alpha\beta}\frac{\partial}{\partial x^{\beta}}\log\sqrt{-g}\right) = -\kappa T. \tag{84}$$

This additional coordinate restriction requires, in particular, that the coordinate system cannot be chosen in such a way that $\sqrt{-g} = 1$ since this would imply the physically implausible consequence that the trace of the energy-momentum tensor vanish.[98]

Since the November restriction was much weaker than the Hertz restriction, it offered a way to overcome the latter's fatal implications for the equivalence principle. In particular, transformations of a given coordinate system to a rotating system or a system whose origin moves in any given way were now allowed so that the generalized principle of relativity is amply, but not fully satisfied. Indeed if a given coordinate system, for instance the usual representation of Minkowski space in Cartesian coordinates, satisfies this coordinate restriction, any other system resulting from the given one by a unimodular transformation must also fulfill this restriction, which is covariant under unimodular transformations.

6. CHANGING HORSES: EINSTEIN'S CHOICE OF GRAVITATION TENSORS FROM 1912–1913

The checklist for candidates for the left-hand side of the field equations that we used in the preceding section was based on the heuristic criteria that Einstein had essentially established by the end of 1912. The decision as to which candidate fares best given these heuristic criteria depends on the state of elaboration of the various mathematical and physical consequences associated with that candidate. The relative arbitrariness of elaborating the consequences of a physical theory along various conceivable pathways, which from the outset can never be overlooked in their totality, therefore entails an element of historical contingency. As the comparison between the *Entwurf* theory, maintained by Einstein essentially for three years, and his final theory of general relativity shows, this contingency may take the form of different physical theories with different empirical consequences, which, at the time, were open to debate.

---

97 See (Einstein 1915a, p.785). Cf. "Untying the Knot ..." (in vol. 2 of this series), sec. 6, eqs. (79)–(82).

98 "In writing the previous paper, I was not yet aware that the hypothesis $\sum T^{\mu}_{\mu} = 0$ is, in principle, admissible." ("Bei Niederschrift der früheren Mitteilung [Einstein 1915a] war mir die prinzipielle Zulässigkeit der Hypothese $\sum T^{\mu}_{\mu} = 0$ noch nicht zu Bewußtsein gekommen.") (Einstein 1915b, 800).

Furthermore, even if the more developed state of elaboration reached by Einstein by the fall of 1915 is taken into account, it is as we have seen in the previous section the November tensor rather than the Einstein tensor which fits Einstein's original heuristic criteria best. The November tensor had passed all tests of Einstein's checklist with only a minor adjustment of the generalized relativity principle while the Einstein tensor had failed the test of the correspondence principle as originally conceived by Einstein. This was all the worse for the Einstein tensor since the generalized relativity principle was an ambitious and idiosyncratic goal which was not shared by many of Einstein's contemporaries, while the correspondence principle had all the support of classical physics and special relativity. That it was the Einstein tensor that in the end won the race can only be understood by taking into account another aspect of the historical process, which we have so far neglected, changes in the heuristic criteria themselves as well as in their relative importance. We therefore need to take a closer look at the actual development of Einstein's thinking.

Why exactly did he turn from one candidate to the other? How did his judgement of candidates evolve? What made him come back eventually to previously discarded candidates after spending almost three years working out a more or less satisfactory relativistic theory of gravitation based on one of them? These questions are the focus of this and the next chapter dealing with what one might describe as Einstein's discovery process or better, as his "investigative pathway."[99] As we have argued, the eventual success of Einstein's research was based on applying shared knowledge resources to the problem of gravitation. The actual mechanism of these applications has so far been considered only from a single perspective, that of assimilating physical and mathematical resources to the basic model of a field equation. In the following, we shall argue that focusing on the exploitation of these resources not only allows us to understand the basic pattern of Einstein's search, the alternation between physical and mathematical strategy. It also allows us to reconstruct, to a surprising extent, the actual course of his search, if we take into account an additional cognitive process as well. While the assimilation of physical and mathematical knowledge to the Lorentz model of a gravitational field equation is basically a top-down process that is guided by the relatively stable high-level cognitive structures at the core of Einstein's heuristic criteria, a reflection on the experiences resulting from such an assimilation, including its failures, could trigger a corresponding bottom-up process of accommodating these high-level structures, including the very mental model itself, to the outcome of these experiences. These two complementary processes were mediated by the external representation of the mental model in terms of mathematical language. The combination of these processes produced conclusions that evolved with the elaboration of the formalism and with the accumulation of Einstein's experience. In order to substantiate this schematic account, we shall, in the following, review his pathways, first in the period documented by the Zurich Notebook and then—in the next chapter—in the period between 1913 and 1915. Relying heavily on the joint

99 See (Renn, Damerow and Rieger 2001; Holmes, Renn and Rheinberger 2003).

work presented in this volume,[100] we shall interpret these pathways as being governed by an interplay between assimilation and accommodation, mediated by the mathematical formalism.

6.1 The Tinkering Phase in the Zurich Notebook

The earliest notes on gravitation in the Zurich Notebook represent a stage of Einstein's search for the field equation in which he had few sophisticated mathematical tools at hand that would allow him to construct candidates fitting the framework provided by the Lorentz model. Even his knowledge of the metric tensor and its properties was still rudimentary. Only gradually did he find ways of exploiting his knowledge of vector analysis for his search. Eventually he familiarized himself with the scalar Beltrami invariants as another instrument that allowed him to investigate the few building blocks at his disposal, that is, the metric as a representation of the gravitational potential, the four-dimensional Minkowski formalism, and his theory of the static gravitational field. In spite of the lack of mathematical sophistication characterizing this early tinkering phase, not to mention the failure to produce promising candidate field equations, it is in this period that Einstein acquired essential insights shaping his research in subsequent phases of work.

These insights consisted, first of all, in a number of concrete results that later turned out to be useful, such as the identification of the core operator (cf. eq. (XXVI)), the establishment of a repertoire of techniques for dealing with coordinate transformations, results on the transformation properties of the Hertz restriction, and, most importantly, the successful implementation of the Lorentz model of an equation of motion in a generally relativistic framework (cf. eq. (XX)). The most far-reaching insights of this period were, however, of a different nature. They consisted in more general ideas resulting from a reflection on the experiences in the tinkering phase, ideas that were largely independent from the concrete mathematical material to which they were applied. Here we encounter a second function of reflection in this context, beyond that of modifying one or the other of Einstein's heuristic principles: reflection could also result in higher-level structures operating on a strategic level, that is, guiding the implementation of these heuristic principles. The most important example is certainly the idea to first impose a coordinate restriction on an object of broad covariance in order to satisfy the correspondence principle and then to explore the transformation properties of this coordinate restriction in order to check the extent to which the generalized principle of relativity is satisfied as well (cf. eq. (LIV)). Even the alternation between more physically and more mathematically motivated approaches emerged as a distinct pattern in this period, again with far-reaching implications for Einstein's subsequent research. The reflection on the experiences of this tinkering phase thus led to what one might describe as a "chunking" of Einstein's

---

100 See, in particular, "Commentary" for this section and "Untying the Knot ..." for the next section (both in vol. 2 of this series).

heuristic principles in terms of procedures that interconnected them in such a way as to ease their implementation as a whole.[101] Such procedures could involve the subsequent translation of these principles into well-defined mathematical requirements on a gravitational field theory (e.g., the choice of a generally-covariant candidate, followed by the stipulation of a coordinate restriction) or they could consist in alternating between physical and mathematical default settings.

6.2 Assimilating Knowledge about the Static Gravitational Field to a Metric Formalism (39L–39R)

When Einstein began systematically to explore a metric theory of gravitation, he was confronted with the problem that the knowledge resources available to him for constructing such a theory presented themselves as more or less isolated building blocks that could not easily be fitted together. On the page of the Zurich Notebook which documents the point of departure of his exploration (p. 39L), he therefore started his investigation simply by listing three such building blocks, the line element in terms of the metric tensor representing the gravitational potential, the four-dimensional Minkowski formalism, and his theory of the static gravitational field. How could they be brought into relation to each other?

The principal challenge was to assimilate the knowledge about the special case of a scalar, static gravitational potential to a tensorial formalism. If such an assimilation were successful, the mental model of a field equation for a tensorial gravitational potential would acquire a physically meaningful instantiation. Einstein's first consideration of the problem of gravitation that is recorded in the Zurich Notebook is precisely such an attempt to assimilate the static case to a metric formalism, concentrating on two of the slots of the Lorentz model for a field equation, that for the differential operator and that for the gravitational potential. For reasons that we have discussed earlier, the default setting for the latter slot was given by the canonical metric for a static field (cf. eq. (25)). Brought into proper mathematical form, Einstein's scalar field equation for the static gravitational field could therefore be conceived as a second-order partial differential equation for the one variable component of this special metric tensor, expressed in a special coordinate frame in which the metric takes on its canonical form. Exploiting mathematical knowledge about the behavior of a tensorial field equation under coordinate transformations, one should then be able to generalize this equation for one component to a field equation for the full metric tensor.

By transforming the equation for the static field into a more general coordinate system, Einstein made an observation that suggested a new pathway to him. He found that, under linear coordinate transformations, the metric tensor behaves exactly the same way as the second-order partial derivatives of a scalar function. This observation opened up a new possibility for drawing on hitherto unexploited mathematical

101 For the concept of "chunking" in cognitive science, see (Minski 1987).

resources and thus for identifying a suitable differential operator acting on the metric tensor.

6.3 Assimilating Knowledge about Scalar Differential Invariants to a Metric Formalism (40L–41L)

Einstein's key problem was that the default settings for two of the slots of the mental model for a gravitational field equation, suggested by his earlier experiences with implementations of this model, could not be matched to each other (see Fig. 3, p. 173). While the default setting for the gravitational potential was represented by the canonical metric, the default setting for the differential operator was, at this point, an object like the Laplace operator, applicable only to scalar functions and covariant only under linear transformations. Was there a way of bridging this gap between a scalar differential operator and a tensorial potential? Einstein's insight into the analogy between the transformational properties of the metric tensor and those of the second-order partial derivatives of a scalar function offered such a bridge, allowing him to bring to bear on this problem mathematical knowledge about scalar differential operators. It suggested the possibility of building some higher-order differential operator acting on a scalar function, which could then be translated into a differential operator acting on the metric tensor. All that was needed for such a translation was the replacement of a second-order partial derivative term by the corresponding components of the metric tensor; the remaining partial derivatives could then be considered as a differential operator acting on the metric.

What could be gained by such a roundabout procedure? If the scalar differential operators involved are just linearly covariant, like the Laplace operator, relatively little. If, however, scalar differential invariants are taken as the building blocks of such a construction, it could lead to the formulation of a generally-covariant differential operator for the metric tensor. There is some indication in the Zurich Notebook that this may have been Einstein's hope. In any case, he systematically checked whether various higher-order scalar differential operators would yield, after translation, a suitable candidate for the left-hand side of the gravitational field equations. But apparently he was unable to single out a candidate promising to fulfill his other heuristic criteria as well, and did not pursue this investigation for the time being. As is clear from later pages of the notebook, however, Einstein did not consider the potential of scalar differential invariants for his project to be exhausted. The purpose of a somewhat obscure calculation on the immediately following pages (pp. 40R-41L), dealing with linear transformations of an algebraic quadratic form, might well have been to learn more about such invariants and their properties.

6.4 Implementing the Lorentz Model of the Equation of Motion (05R)

At some later point, Einstein made a new beginning in his research on a theory of gravitation. He now turned to the other element of the field-theoretical model, the equation of motion. He probably had realized by this time that the equation of motion for a point particle in a gravitational field corresponds to the equation for a geodesic curve in a four-dimensional curved spacetime (cf. eq. (23)).[102] But he probably also had realized that an equation of motion in this sense was not quite the match of the gravitational field equation for which he was looking. The default-setting for the source-slot of the Lorentz model for the field equation was not a point particle but the energy-momentum tensor (cf. eq. (XIV)). The mental model of a field equation together with special relativistic continuum theory now suggested what such an equation should look like in terms of the energy-momentum tensor (cf. eq. (XX)).[103] Such an equation would provide, at the same time, an expression for energy-momentum balance in the presence of a gravitational field.

When Einstein studied the equation of motion problem, he was confronted with the challenge of how to link his general expectations concerning the structure of such an equation with his concrete knowledge about the motion of point particles in a gravitational field. To bridge this gap he made use, as we have discussed before, of a particular model of matter, which allowed him to link point mechanics and continuum mechanics, i.e. the model of "dust" (cf. eq. (XXI)). At a mathematical level, the bridge was built with the help of the Lagrangian formalism (cf. eq. (19)). The dust model allowed Einstein to generalize the equation of motion derived within the Lagrange formalism into a relation between components of the energy-momentum tensor. This relation suggested, in turn, what the full tensorial equation of motion in gravitational field should look like, if it was supplemented by both mathematical and physical default-assumptions provided by the corresponding special relativistic equation.[104] As discussed above, eq. (XX) expresses the energy-momentum balance in a gravitational field, i.e. the generalization of the special relativistic relation between force, energy, and momentum (cf. eq. (XVIII)). Einstein also realized that, from a mathematical point of view, it corresponds to the covariant divergence of the energy-momentum tensor (cf. eq. (XXIV)). This remarkable convergence of physical and mathematical perspectives must have confirmed the expectation that his result also applies to other kinds of sources and turned Einstein's equation into the default-setting for the equation of motion in the Lorentz model and for the energy-momentum balance in a gravitational field.

---

102 He reproduced the proof that the trajectory of a force-free motion constrained to a two-dimensional surface is a geodesic on a page of the notebook immediately following the consideration of quadratic invariants mentioned in the previous subsection (see p. 41R).

103 Einstein emphasized the central role of the energy-momentum tensors and the importance of special-relativistic continuum mechanics in an article he wrote in 1912 but never published, see (CPAE 4, Doc. 1, 63).

104 See the discussion in (Norton 2000, Appendix C).

6.5 A Mathematical Toy Model as a New Starting Point (6L–7L)

The mismatch between the instantiations for two of the slots of the mental model of a field equation, that for the differential operator and that for the gravitational potential, left Einstein with two principal options as to how to proceed. He could continue trying to build an appropriate differential operator applicable to the metric tensor or he could tentatively explore substitutions of the default-setting for the gravitational potential, thus creating "toy-models" in the sense of obviously unrealistic instantiations of the model. Even if that meant temporarily suspending the insight that the gravitational potential is represented by the metric tensor, it might still be possible to gain knowledge from exploring such toy-models that could be helpful in constructing a more realistic candidate field equation.

When Einstein became familiar with the generally-covariant Beltrami invariants as a generalization of scalar differential operators, they must have appealed to him as a promising starting point for his search for a relativistic gravitational field equation. A field equation based on those invariants would automatically satisfy the heuristic requirement of the generalized principle of relativity. A first attempt to construct a differential operator for the metric out of operators acting on a scalar function had, as we have seen, turned out to be too speculative. It was hence worth trying to explore a generalization of the scalar Poisson equation in a generally-covariant setting by using—instead of the Laplace operator—the second Beltrami invariant applied to a scalar function. While such a generally-covariant scalar field equation was only a toy model, it confronted Einstein with a serious problem, viz. that of reconciling a mathematically satisfactory candidate with the physical knowledge of his theory of static gravitational fields, (see Fig. 3, p. 173). In a sense, a scalar field equation formulated in terms of the second Beltrami invariant represents the counterpart of the scalar field equation of Einstein's static theory: while the latter constitutes an initial, physically plausible instantiation for the field-theoretical model, the former represents an equally plausible initial instantiation rooted in mathematical knowledge. In both cases, the resulting field equations were merely starting points for further investigations that had to make contact with knowledge not yet embodied in these first default-settings.

It therefore comes as no surprise that Einstein tried to find out under which conditions a generally-covariant scalar field equation formulated in terms of the second Beltrami invariant reduces to the ordinary Poisson equation. Such a reduction must be possible if the candidate (or rather toy) field equation is to comply with the correspondence principle. It turned out that the implementation of this heuristic principle in this concrete case requires an additional constraint on the choice of the coordinates, supplementing the field equation. Essentially by inspection, Einstein could identify the harmonic coordinate restriction (cf. eq. (57)) as a condition that would make sure that the Beltrami field equation reduces to the ordinary Poisson equation for weak gravitational fields. In other words, the exploration of a toy field equation taught Einstein that a candidate field equation obtained from a mathematical default-setting may require an additional coordinate restriction to be viable from a physical point of view as well; it also familiarized him with a specific example of such restriction, which later turned out to be useful when studying the Ricci tensor.

How could the toy field equation be turned into a real candidate field equation? If the Beltrami field equation is considered as a mathematically reasonable structure to which physical knowledge should now be assimilated, such as the insight that the gravitational potential is actually represented by the metric tensor, it made sense to try to bring this knowledge into an appropriate mathematical form. If unimodular coordinate transformations are assumed, the determinant of the metric transforms as a scalar and can be used to fill the potential-slot of a scalar field equation. The next question was whether the resulting field equation, for the special case of a static field, could be related to the familiar static field equation. Einstein tried to extend this approach by taking into account different versions of a Beltrami-type field equation but failed to integrate the mathematical and the physical knowledge in this way.

6.6 A Physical Toy Model as a New Starting Point (7L–8R)

Einstein's first exploration of the Beltrami invariant had not answered the question as to how to get from a mathematically plausible scalar differential equation to a tensorial field equation that is both mathematically and physically plausible. Reflecting on this gap, Einstein may well have considered the possibility of dividing this transition into two steps. The first would be to construct a tensorial field equation that, even if its mathematical properties were unclear at the outset, made good sense physically. The second step would take him, relying on mathematical tools, from such a physically-plausible toy field equation to the final equation.

In any case, instead of taking a simplified instantiation for the potential-slot of Lorentz model for a field equation to explore a mathematical toy model, Einstein now chose a simplified instantiation for the differential operator slot, while keeping the realistic setting for the potential slot, i.e. the metric tensor. His experience with vector calculus and its use in physics allowed him to write down a straightforward translation of the ordinary Laplacian operator into a differential operator acting on the metric tensor, the core operator. Einstein's experience with the Beltrami invariants must have made it clear that the core operator could hardly represent a generally-covariant object. From the way in which it was constructed, however, it was equally clear that a field equation based on the core operator satisfies the correspondence principle. For this reason, the core operator (cf. eq. (XXVI)) became the default-setting for all of Einstein's subsequent attempts to implement this principle.

This candidate now had to be checked against the other heuristic requirements and, in particular, its behavior under coordinate transformations needed to be explored. This could be done in two distinct ways: either by directly checking the transformational behavior of the core operator, or by considering it—in the sense indicated above—an intermediate step towards the final field equation. Einstein began with the first option. To get beyond linear transformations, however, he used a special kind of coordinate transformations, which explicitly depend on the metric tensor, and which he later called "non-autonomous transformations".[105] The behavior of the core operator under such transformations is determined by differential equations for the transformation matrices involving the metric tensor and its derivatives. Einstein succeeded in writing down, at least for infinitesimal transformations, the essential term in such a differential equation. But probably in view of the complexity that this condition would take on for finite transformations and, more generally, in view of the unfamiliar character of these non-autonomous transformations, he abandoned this approach and turned instead to the more familiar territory of ordinary coordinate transformations.

In that case the only way to go beyond linear transformations was to generalize the core operator. Einstein developed an ingenious method for doing so. First of all, he considered the two differential operators constituting the core operator separately, the divergence and the gradient (or exterior derivative, cf. eq. (XXV)). He then took the familiar form of these operators applied to some second-rank tensor in Minkowski spacetime with (pseudo-)Cartesian coordinates as his starting point. Einstein now made the assumption that these operators actually transform as tensors under arbitrary coordinate transformations. Under this assumption, a coordinate transformation carrying these operators from their special form in Cartesian coordinates to arbitrary coordinates should reveal their generic form. The idea was similar to that of obtaining a generalization of the line element of Minkowski spacetime to that of a generic curved spacetime by passing from pseudo-Cartesian to arbitrary coordinates in Minkowski spacetime. In both cases one simply had to assume that an equation obtained for the Minkowski metric in arbitrary coordinates is actually valid for the metric of a generic spacetime. Although Einstein did not see his calculations through to the end, he essentially succeeded in finding covariant generalizations of the constituents of the core operator. Eventually he must have realized, however, that this success amounted to no more than a Pyrrhic victory since these generally-covariant differential operators give zero when applied to the metric tensor. In other words, a generalized core operator built from these covariant differential operators is not suitable as a candidate for the left-hand side of the gravitational field equation. Eventually, this failure forced Einstein to take the peculiar non-autonomous coordinate transformations of his first approach much more seriously than he had probably intended when he first encountered them.

6.7 Identifying the Core Operator as the Target of the Mathematical Strategy (8R–9R)

In his next attempt Einstein, reflecting on his earlier failures and insights, combined his prior experiences to develop a procedure for constructing candidate field equations that he would repeatedly use in the notebook (cf. eq. (XLVIII)). The genesis of

---

105 See Einstein to H. A. Lorentz, 14 August 1913, (CPAE 5, Doc. 467). For discussion, see "Commentary ..." (in vol. 2 of this series), sec. 4.3.

this procedure as the result of an oscillation between a more mathematically and a more physically motivated attempt illustrates Einstein's learning experience in the course of his search, which therefore cannot be seen simply as the successive elimination of unsatisfactory alternative candidates.

The attempt to conceive of the physically plausible core operator as the representation of a more general covariant object in specific coordinates had failed because of the degeneracy of the corresponding differential operations when applied to the metric. It made therefore sense to return to the earlier direct exploration of the transformation properties of the core operator. This pathway had not definitively failed yet but turned out to be too rough. Considered from a higher level of reflection, the core operator could not just serve as a physically plausible starting point but also as the possible target of a strategy starting from a mathematically well-defined object. At this point, the only such mathematically well-defined objects that Einstein had at his disposal were the Beltrami invariants. It therefore was natural to deal with them once again, but now not with the theory of the static gravitational field but with the core operator as the more promising physically meaningful target. This approach came with a new challenge, the task to extract a tensorial object, the core operator, from a scalar invariant. This challenge turned out to be manageable.

In short, the idea was to once more start from a mathematically motivated instantiation, the second Beltrami invariant, trying to exploit its familiar mathematical properties in order to determine the transformational behavior of the physically plausible core operator. The necessary bridges between tensorial and scalar objects were readily at hand. From his earlier experience, Einstein knew that he could use the determinant of the metric tensor in the second Beltrami invariant if he considered only unimodular transformations. Now he realized that he could, in turn, try to extract a tensor from a scalar by conceiving the latter as a contraction between two tensors, in this case of the metric tensor and the core operator.

The concrete implementation of this approach confronted Einstein with a number of problems, minor and major. There was, first of all, the need for a restriction to unimodular coordinate transformations. More importantly, when trying to extract the core operator from the second Beltrami invariant applied to the determinant of the metric tensor, he encountered an additional first-order term that required further consideration. Einstein's understanding of the conservation principle, and in particular his experience with his second theory of the static gravitational field, must have immediately suggested to him that this first-order term might be related to an expression for gravitational energy-momentum (cf. eq. (XXXIV)).

In the end, however, Einstein did not succeed in establishing a convincing bridge between core operator and Beltrami invariants. As a consequence, he failed to clarify the transformational properties of the core operator or of a suitably amended candidate gravitation tensor constructed from it. At this point, he took up once more the direct exploration of the transformational properties of the core operator earlier abandoned because of the intricacy of the non-autonomous transformations involved in it. This resort to an earlier approach was, however, no return to square one. Einstein could benefit from the insights he had made in the meantime, in particular from the breaking down of the original problem into simpler ones that the introduction of the Beltrami invariants had made possible. The new first-order term posed a problem analogous to the one Einstein had first encountered when comparing a mathematical toy model based on the second Beltrami invariant with the ordinary Laplace operator and thus suggested the remedy of introducing a coordinate restriction as an additional hypothesis under which a mathematically acceptable expression reduces to a physically plausible one. One could limit then the direct exploration of transformational properties to the remainder term distinguishing the second Beltrami invariant from the contraction of the core operator with the metric tensor. This task was simpler than the original one given the structure of the remainder term. If the class of non-autonomous transformations leaving this term invariant could be determined, one would thereby have found the class of transformations leaving the principal term, i.e., the contraction of the core operator with the metric, invariant as well (cf. eq. (LI)). In this way, a bridge would have been built between the transformational behavior of the mathematically well-defined second Beltrami invariant and that of the physically plausible core operator. In spite of this simplification with respect to Einstein's original attempt to determine the transformational behavior of the core operator, even this reduced task still turned out to be too cumbersome to carry out.

Although this entire episode was fraught with frustrations of reasonable hope, it gave Einstein strategic insights well beyond the concrete mathematical material at hand. There was, first of all, the recognition of the canonical form for the left-hand side of a gravitational field equation, which would have to consist of a core operator plus first-order correction terms somehow related to gravitational energy-momentum (cf. eq. (XXXIV)). Second, the experiences of this episode brought the mathematical strategy into a form that was to dominate much of the subsequent work documented in the notebook. The general idea now was to start from an object with well-defined mathematical properties, in particular with a broad enough covariance group to meet the demands of the generalized principle of relativity (cf. eq. (XLVI)). The next step was to extract from it a candidate gravitation tensor with well-defined physical behavior, more specifically the core operator possibly with correction terms not invalidating the correspondence principle (cf. eq. (XLVIII)). The extent to which the generalized principle of relativity was actually fulfilled could be determined by checking the transformational behavior of the term distinguishing the candidate gravitation tensor from the mathematical starting point (cf. eq. (LI)). This term could be eliminated by imposing the appropriate coordinate restriction (cf. eqs. (L), (LII)). It is remarkable that this strategy, crucial to Einstein's exploration of the Riemann tensor, was in place before he had even seen a single realistic candidate gravitation tensor.

6.8 Subjecting the Core Operator to a Piecemeal Approach (10L–12R, 41L–R)

While in the last episode Einstein had developed an overall strategy for solving his problem and failed, he now took a more piecemeal approach. He focused on a mathematically much simpler object to avoid the complexity of the non-autonomous transformations he had considered so far. On the whole, this phase of his work was characterized by the attempt to break down his main problem, the identification of appropriate field equations, into smaller, more manageable pieces in the hope of identifying reliable building blocks that could then be used to put the puzzle together.

This line of pursuit was largely shaped by the options and constraints that had emerged in the course of Einstein's preceding experience. In particular, while he had just established a paradigm for what was to become his mathematical strategy, for the time being this strategy was powerless for want of mathematical objects other than the Beltrami invariants that could serve as input. The core operator, on the other hand, inspired confidence as a solid achievement that would be a physically meaningful starting point were it not for the difficulties of determining its transformation properties. Yet in view of the absence of other mathematical resources, the use of non-autonomous transformations may have seemed unavoidable. And if hope was to remain of connecting possible results to the Beltrami invariants, the only advanced mathematical objects at Einstein's disposal, the transformations should be unimodular as well.

It is against this backdrop that the emergence of the main idea guiding Einstein's work in this episode becomes understandable. This work may have sprung from the idea to consider simpler mathematical objects that would make a direct approach to the examination of their transformation properties feasible. In any case, at some point it must have become clear to Einstein that the full problem, the determination of the transformation properties of the core operator, could actually be broken down into the study of such simpler objects, if possible vectors or even scalars (cf. eq. (XXV)). The preceding experience with the attempt to extract the core operator from the scalar Beltrami invariant may have triggered this idea. It therefore made sense to carefully check the covariance of these simpler objects in particular under the transformations relevant to the implementation of the elevator and the bucket models, i.e. transformations in Minkowski spacetime to uniformly accelerated or to rotating systems.

The realization of the idea just described brought Einstein to study the transformation properties of the Hertz restriction (cf. eq. (60)). If this restriction was imposed, the core operator reduces to a simpler object whose transformation properties can then be determined separately. But in spite of the greater simplicity of these objects, it turned out to be necessary to introduce a further simplification and to limit the analysis to infinitesimal transformations. With these presuppositions in place, Einstein was able to obtain some specific results even if these were not all that encouraging. In particular, when attempting to implement the elevator model and satisfy the equivalence principle, he found that compatibility with the covariance properties of the Hertz restriction required a modification of the transformation to a uniformly accelerated system, which turned out to be unacceptable for physical reasons. In short, Einstein found it difficult to establish a match between the transformational properties of the objects under study and his physical expectations. How much of the room opened up by his main idea, that of splitting the core operator into two simpler pieces, did he actually investigate in the course of his calculations? It is clear from his notes that he was aware of variants of this operator and hence of alternative splits (involving e.g. the harmonic restriction instead of the Hertz restriction), but he left this option unexplored.

In the end, Einstein once again assembled a number of isolated results that later became useful. In addition, he gained strategic insights governing the subsequent course of his research. As far as his specific results are concerned, he established, for instance, that the Hertz restriction is covariant under infinitesimal non-autonomous transformations to a rotating system in Minkowski spacetime but, as mentioned above, not under a transformation to a uniformly accelerated system. He also found that the core operator is covariant under antisymmetric non-autonomous transformations in Minkowski spacetime, without however being able to associate physical meaning with this result. In the course of his work, he gradually shifted the emphasis of his quest from the transformation properties of the constituents of the core operator to a careful reexamination of the physically relevant transformations themselves. He thereby again accumulated some useful findings such as the derivation of the metric for Minkowski spacetime in rotating coordinates from the Lagrangian formalism. Eventually he made a fresh start, taking unimodular transformations as a starting point for implementing the generalized principle of relativity. He once more tried to match them with transformations to a uniformly accelerating system, again without success. He then abandoned this attempt to incorporate the equivalence principle, alongside with his entire endeavor to deal with the covariance properties of the core operator on the basis of his piecemeal strategy. What remained, apart from specific achievements, was the experience that non-autonomous transformations could be handled after all, at least when applied to sufficiently simple objects. But this was an insight that Einstein would be able to put to good use only much later, when he explored the transformation properties of the finished *Entwurf* theory, in particular in its Lagrangian formulation (cf. eqs. (LXIII), (LXIV)). Of more immediate impact was his realization, fostered by the disappointments produced even by his piecemeal strategy, that it might be prudent to put the pursuit of an audacious interpretation of the generalized principle of relativity on hold, turning instead to the physical requirements embodied in the conservation principle.

6.9 Using the Core Operator as the Starting Point for the Physical Strategy (13L–13R)

Einstein's next move was to look at his problem from a different angle, bracketing the intricate problems raised by the generalized principle of relativity and making sure that what he had achieved so far was at least sound in other respects. And even after the disappointing yield of his piecemeal approach, the identification of the core operator as a candidate compatible with the correspondence principle remained such a sound result. Einstein now tried to address the seemingly intractable aspects of the exploration of the core operator by reducing the ambitious goals imposed by the generalized principle of relativity. For this purpose, he set up a more manageable framework for dealing with the other central, but as yet unexamined aspect of the validity of the core operator as a candidate gravitation tensor, its compatibility with the conservation principle.

His previous research had already suggested that perhaps the core operator needed to be supplemented by additional first-order terms representing the energy-momentum of the gravitational field in the gravitational field equation. But at that point the problem of energy-momentum conservation had occurred only as a marginal aspect of the relativity problem at the center of Einstein's attention. The impasse of his work on this problem provided a natural occasion to return to the issue.

Einstein created a manageable framework by restricting all considerations to unimodular, linear transformations. The requirement of linearity would secure the tensorial character of the core operator, while the requirement of unimodularity kept the door open for establishing contact with the Beltrami invariants later. By setting up a systematic framework for generating vectors and tensors involving the metric Einstein could hope, first of all, to reduce the ambiguities of his approach and second, to gain solid ground for examining the relation between core operator and conservation principle without the interference of the relativity problem. Such an examination might, in particular, help to find the correction terms that he had earlier tried to obtain from the first Beltrami invariant.

Einstein constructed a framework for generating tensorial objects involving the metric with well-defined transformation properties, beginning with the Hertz expression (cf. eq. (60)). He set up a survey of the first-order objects and then stopped, either because these were the objects in which he was mainly interested with a view to the correction terms needed for the core operator or because even this limited overview dashed any hopes he might have had for a reduction of the space of possible candidates. Whatever the case may be, it was at this point that he once more took up the core operator directly, checking its compatibility with energy-momentum conservation. As it turned out, his network of results had become dense enough to allow for such a check which, even if it failed, would probably still provide hints about what was still needed to enforce compatibility. Einstein combined the left-hand side of a field equation based on the core operator with the expression for the energy-momentum balance he had established earlier (cf. eq. (XXX)). He thus produced an expression corresponding to (cf. eq. (XXXII)):

GRAD(POT) X LAP – DIV(LAP) **(LXXXVII)**

which *a priori* could be expected to be of third differential order. But if one now assumes that a field equation of the form (cf. eq. (XXXIX)):

LAP = ENEMO + FIELDMASS **(LXXXVIII)**

holds, then compatibility with the conservation principle requires that the above expression reduces to the second-order expression (cf. eqs. (XXXI), (XXIV)):

GRAD(POT) x FIELDMASS – DIV(FIELDMASS) **(LXXXIX)**

with an appropriate explicit form of **FIELDMASS**. Einstein found indeed that imposing the Hertz condition implies that no third-order terms appear. But, unfortunately, he failed to arrange the terms in the resulting expression in a way that would have allowed him the extraction of an explicit form of **FIELDMASS**.

In summary, Einstein succeeded neither in identifying the conditions under which the core operator is compatible with the conservation principle nor in finding the correction terms that could possibly help establishing such compatibility. Even for such a simple object as the core operator the differential equation resulting from his compatibility check appeared to be too complicated. His calculations and the reflections stimulated by them had nonetheless laid the groundwork for the global approach we have called his "physical strategy," all elements of which were now assembled. The core operator provided him with the starting point for this approach and the calculations just considered constituted the conservation compatibility check for this candidate (cf. eq. (XXXI)). The restriction to linear transformations made it possible to postpone a check of the extent to which the generalized relativity principle was satisfied. What was still lacking was a procedure for guessing or generating suitable correction terms to be added to the core operator to turn it into a viable candidate. Einstein had arrived at a dead end, but not with empty hands. Not only had he accumulated a reservoir of insights and tools that would be useful for his further search, but he had developed two overall strategies, each capable of guiding this search. For the time being, however, both strategies were doomed to be abeyant as long as certain elements that could trigger their application were missing. But as soon as an appropriate incentive was provided, either of them could be activated. For the physical strategy to become productive, all that was needed was a way to generate plausible correction terms to the core operator. For the mathematical strategy to become productive, all that was required were tensors with second-order derivatives of the metric and a well-defined transformational behavior. As it turned out, the latter option was realized first.

6.10 The Systematic Search Phase in the Zurich Notebook

The raw material needed to set the mathematical strategy in motion was evidently delivered by Marcel Grossmann whose name appears next to the first occurrence of the Riemann tensor in the notebook. With this entry, the first phase of Einstein's research was over and a phase of systematic searching for suitable gravitational field equations began. The Riemann tensor represented something like a raw diamond for Einstein to which he could now apply the various extraction schemes that he had elaborated earlier as well additional stratagems he developed in the course of his search. Among these schemes was the contraction of the fourth-rank Riemann tensor to yield a second-rank candidate gravitation tensor, the extraction of such a candidate from a scalar object, the stipulation of coordinate restrictions, and the possibility of modifying candidate field equations by adding or subtracting terms. The products to which these extraction processes gave rise had an impact on Einstein's search procedure going well beyond their immediate evaluation as being either refinements or debris, as is particularly evident from the identification and subsequent rejection of the Einstein tensor.

Einstein's procedure was guided throughout by the Lorentz model, which suggested that candidates for the left-hand side of the field equations have the form of a core operator plus correction terms (cf. eq. (XXXIV)). His prior experience with extracting such candidates from the second Beltrami invariant furthermore gave him guidance on how to handle those terms not fitting his expectations, i.e., how to eliminate them with the help of a coordinate restriction. Since he started from objects with well-defined transformation properties, the main heuristic criteria to be checked were the correspondence and conservation principles. From the point of view of the mathematical strategy, both criteria posed similar challenges and thus seemed to call for similar responses, viz. coordinate restrictions to be imposed in addition to the field equations. This parallel strengthened Einstein's expectation that the stipulation of these heuristic principles required a restriction of the covariance of the object used as the starting point of the mathematical strategy. On the weak-field level, the two restrictions, one resulting from the correspondence, the other from the conservation principle could easily be compared with each other; their compatibility or rather the lack thereof was an important driving force in the search for field equations (cf. eq. (LX)).

The severe restriction on the generalized relativity principle that seemed to be the almost unavoidable consequence of Einstein's procedure made it all the more urgent to check whether or not at least the most essential requirements associated with this principle were satisfied and, in particular, whether the important special case of rotation was included. Not surprisingly, Einstein more than once reexamined this special case during his search.

The difficulties Einstein encountered in the course of his attempts to enforce his heuristic criteria within the formalism he was weaving around the Riemann tensor naturally provoked a reflection on the validity, the physical meaning, and the mathematical implementation of these criteria. After all, they may just have been prejudiced. Does the conservation principle really require the covariant divergence of the stress-energy tensor to vanish (cf. eq. (XXIV))? Does the correspondence principle really demand a static gravitational field to be represented by a spatially flat metric (cf. eq. (25))? Can the generalized relativity principle perhaps be satisfied for rotation by metric tensors other than the one obtained from a coordinate transformation of the standard Minkowski metric? Such questioning of his original heuristic criteria and the default settings suggested by them would eventually pave the way for the breakthrough of 1915. But in the winter of 1912–1913, the answers that Einstein found to these questions confirmed his original conceptions and solidified them by extending the network of inferences in which they were embedded. Ironically, it was precisely the lack of a candidate field equation complying with his heuristic criteria and worthy of further elaboration that also prevented, for the time being, the construction of an even wider network of inferences that would allow these criteria to be overcome.

While Einstein's search would eventually turn up just such a candidate, the *Entwurf* field equation, this candidate was no longer the result of an extraction from the Riemann tensor.

What was overturned in the course of the research documented by the notebook was not Einstein's reliance on his heuristic criteria but the way in which he tried to meet them following his mathematical strategy. Even when he managed to find a candidate for which the coordinate restrictions implied by the correspondence and conservation principles, respectively, could be matched, at least on the weak-field level, the ensuing restriction of the generalized relativity principle and the question of how to satisfy the conservation principle for the full equation made the entire attempt look futile. The appeal of the generalized principle of relativity thus gradually faded away, and the conservation principle gradually emerged as the major stumbling block of the mathematical strategy and, at the same time, as the key stone for a new approach corresponding to a successful implementation of the physical strategy.

Einstein's checks of the conservation principle in the context of the mathematical strategy were limited to weak-field equations. Accordingly, all of his results concerning candidate gravitation tensors—positive as well as negative—were provisional only. For the time being, the limited exploration level of the conservation principle could not be overcome in the context of the mathematical strategy. First, Einstein had no systematic mathematical technique at his disposal for implementing this principle beyond the weak-field level. He would acquire such a technique only much later when developing a variational formalism for the *Entwurf* theory in 1914. Second, the ad-hoc strategies he used to implement the conservation principle beyond the weak-field level necessitated substantial modifications of the candidate field equations serving as the starting point of the mathematical strategy, modifications that made the transformation properties of the proposed field equations intractable despite their origin in the generally-covariant Riemann tensor.

Einstein's experiences with extracting candidate gravitation tensors from the Riemann tensor thus displayed a remarkable parallelism to his prior experiences with the Beltrami invariants. In both cases, the advantages gained by starting from an invariant or generally-covariant object had to be gradually given up in favor of satisfying the other heuristic requirements rooted in the knowledge of classical physics until finally nothing was left of the covariance properties that recommended these objects in the first place. But it was not only this twofold experience of failure that ultimately triggered a switch from the mathematical to the physical strategy. It was precisely the main weakness of Einstein's attempts to come to terms with the conservation principle, i.e., the limitation to the weak-field level, that indicated a way out of the impasse. If the implementation of the conservation principle at the weak-field level could not be the final word, it made sense to take energy-momentum conservation for the weak-field equations as a starting point for identifying those additional terms that were needed to turn the core operator into a viable candidate complying with this heuristic requirement, irrespective of the generalized principle of relativity. Einstein thus found a way of solving the problem that had blocked the pursuit of the physical strategy before, viz. the lack of a procedure for generating plausible correction terms to the core operator. His difficulties with the mathematical strategy suggested a procedure, whose first elements were found before they turned into a systematic mechanism. Once more the essential pattern governing the next step of Einstein's research, the derivation of the *Entwurf* field equations, had been prepared by reflecting on the blocked pathways encountered in the previous episode.

The failure of Einstein's pursuit of the mathematical strategy in the Zurich Notebook resulted in the derivation of the *Entwurf* field equations along the physical strategy. The establishment of these field equations, compatible with both the correspondence and the conservation principles, ended, for the time being, his systematic search for gravitational field equations. What remained from his efforts in the winter of 1912–1913, however, was more than yet another and, as it eventually turned out, unsatisfactory candidate that would eventually be discarded. There was the November tensor, which did not immediately fall victim to any knock-out argument derived from Einstein's heuristic checklist, and which was dropped, not in favor of a better candidate, but in favor of a seemingly better strategy. There were the Ricci tensor and the linearized Einstein tensor, which had been explored only at the weak-field level. From this perspective, their later revival is not surprising. But apart from candidates that he would consider again in late 1915, Einstein's search for field equations in the winter of 1912–1913 also left its mark at the strategic level, both in his subsequent attempts to consolidate the *Entwurf* theory and in his renewed search for field equations at the end of 1915. In fact, even when he focused exclusively on the *Entwurf* theory, he never abandoned the expectation, grounded in the experience documented by the Zurich Notebook, that it should be possible to arrive at the same field equations using either the physical or the mathematical strategy. It was this persistence, perhaps more than the potential of any not yet fully explored candidate, that prevented Einstein from ceasing his quest before he had reached his goal of a generally-relativistic theory of gravitation in late 1915.

6.11 Fitting the Riemann Tensor to the Lorentz Model (14L–18R)

When Marcel Grossmann introduced Einstein to the Riemann tensor, this new mathematical resource fell on ground that was well-prepared by Einstein's previous investigations. The expectations with which he approached the new object, however, sent him in a direction very different to where our modern expectations would take us, viz. the derivation of the Einstein field equation from the Riemann tensor. For Einstein, the Lorentz model essentially prescribed the steps to take to evaluate the new candidate. His prior attempts to implement this model had led him, in particular, to expect a field equation with a left-hand side of the form (XXXIV), i.e., a left-hand side of the form 'core operator plus correction terms,' which is incompatible with what we now take to be the correct field equations.

The central role of the Riemann tensor within the absolute differential calculus as the wellspring of all other "differential tensors" and "differential invariants"—a role of which Grossmann was certainly aware (Einstein and Grossmann 1913, 35)—and its unexplored status in Einstein's investigations must initially have nourished high hopes for the project of extracting from it a suitable left-hand side of the field equations. Einstein may even have expected that the direct pathway from the Riemann tensor to an object fitting the Lorentz model would produce the desired result, without any of the moves and tricks that had been necessary in the earlier attempts based on more pedestrian mathematics. If needed, however, by now such auxiliary schemes were available to Einstein should difficulties arise. In any case, the fourth-rank Riemann tensor had to be turned into a second-rank tensor that could serve as the left-hand side of gravitational field equations whose right-hand side was the second-rank stress-energy tensor (cf. eq. (XIV)). This was a straightforward mathematical operation, which Einstein carried out as soon as he had been handed the Riemann tensor. Unfortunately, the result of this operation, the second-rank Ricci tensor (cf. eq. (55)) did not fit to instantiate the open operator slot of the Lorentz model for the field equation but contained additional, unwanted second-order terms invalidating the correspondence principle (cf. eq. (56)). Einstein's first attempt to assimilate the Riemann tensor to his mental model thus resulted in the condition that these disturbing terms would have to vanish. The appearance of such an additional condition is reminiscent of similar hindrances he had encountered exploring the Beltrami invariants.

Now that the direct approach had failed, Einstein was forced to exploit the tricks and tools he had assembled before. The most obvious way to connect his new predicament with his earlier experiences was the construction of a scalar object from the Riemann tensor, the Ricci curvature scalar. This scalar object could be subjected to exactly the same procedure as the scalar Beltrami invariant. Einstein thus attempted to extract a tensorial object from it in analogy to his earlier treatment of the second Beltrami invariant, i.e., by conceiving the scalar as the contraction of this new, contravariant tensorial object and the covariant metric tensor. Also in analogy with his earlier work on the Beltrami invariants, he set the determinant of the metric equal to unity to simplify his calculations, thereby imposing a restriction to unimodular coordinate transformations.

The hope was that the new second-rank tensor extracted in this way from the curvature scalar would represent a suitable candidate for the left-hand side of the field equations, meeting the requirements of the Lorentz model. Unfortunately, the considerable calculational effort required to pursue this option failed to produce more acceptable results than the direct approach. Einstein even briefly considered introducing an additional condition on the metric tensor—a weaker form of the Hertz restriction—but apparently gave up this idea because it did not seem to promise an easy way out either. He then tried to make some progress by comparing the two unsatisfactory candidates he had extracted from the Riemann tensor in both their contravariant and covariant forms. This procedure also followed the example set by his experiments with the Beltrami invariants and may similarly have been driven by a concern for uniqueness and the hope to learn from combining different pathways. While the procedure was given up without reaching a definite conclusion, it gave an insight that quickly proved to be important. Using techniques familiar from his Beltrami experiments, Einstein found that the constancy of the determinant of the metric could be used to replace one of three disturbing second-order terms occurring in the Ricci tensor by a first-order expression.

At the same time, it must have been clear to him that disturbing second-order terms of some sort were there to stay, and, consequently, that at least some aspects of what might initially have appeared to be mere stop gaps were there to remain as well. Precisely because of the original promise of the Riemann tensor, it was clear that the problem could no longer be the lack of mathematical resources and that no amount of calculational sophistication would suffice to turn the Riemann tensor into an acceptable candidate gravitation tensor without introducing further hypotheses, in all likelihood with serious physical repercussions. The need for further hypotheses was also suggested by the fact that the conservation principle had not played any role in the analysis of the Riemann tensor so far. It was to be expected that this heuristic requirement would exact its price as soon as the physical consequences of a gravitation theory based on the Riemann tensor were pursued any further.

6.12 Establishing a Contradiction between the Correspondence and the Conservation Principles (19L–19R)

When Einstein had tried to match the second Beltrami invariant to the correspondence principle, he had hit upon the harmonic coordinate restriction as a suitable auxiliary hypothesis. Trying to match the Ricci tensor to the correspondence principle, he found that the same hypothesis could be used to eliminate *all* disturbing second-order terms. This first-order condition on the metric tensor was suggested by the condition following from the restriction to unimodular coordinates. This immediately gave it a similar status, i.e., that of a global coordinate restriction. Against the background of his earlier experience with the Beltrami invariants, the introduction of such an auxiliary hypothesis was clearly an application of what we have called the mathematical strategy. The natural next step would thus have been to explore the transformational properties of this additional restriction (cf. eq. (LIV)).

However, Einstein's earlier experience had also involved wrestling with the conservation principle. He had come to realize that this principle might entail further restrictions, affecting the covariance properties of the theory. Knowing that exploring the transformational properties of such extra conditions could become quite involved, he first tackled the issue of conservation. It made sense to collect all necessary restrictions first, and establish the transformational properties of overall restriction later (cf. eq. (LIX)). To explore the emerging network of conditions, Einstein simplified his framework, focusing on a first-order, weak-field approximation. He thereby effectively introduced another toy model, now with the goal to explore the entanglement of correspondence and conservation principles.

In weak-field approximation, the harmonic coordinate restriction coming from the correspondence principle could easily be related to the restriction coming from the requirement of compatibility between field equation and energy-momentum conservation. It was immediately clear that the field equations in first-order approximation satisfy the divergence condition (LXXI). The conservation compatibility check (LXXIV) gave rise to an additional restriction which could also be brought into a first-order form, for comparison with the harmonic restriction. The combination of the resulting Hertz restriction with the harmonic restriction implies that the trace of the metric tensor must be constant. This implication was inacceptable to Einstein on physical grounds. It was incompatible not only with the default-setting for the metric tensor of a weak static gravitational field (25) but also, via the field equations (cf. eq. (LXXV)) with the default-setting for the stress-energy of matter as given by eq. (XXI). In view of this discrepancy between the mathematical consequences of his heuristic principles and his physical expectations, it is not surprising that Einstein at this point reexamined the legitimacy of the conservation compatibility check which had evidently triggered this conflict. A crucial implication with physical significance was the vanishing of the covariant divergence of the energy-momentum tensor (cf. eq. (XXIV)). Within his weak-field approximation, Einstein therefore rederived this relation from first principles, i.e., from the continuity equation and the equation of motion. In this way, he not only extended his network of arguments to include the latter results but, more importantly, he firmly established the existence of a contradiction within this network, with no simple escape by adjusting his heuristic principles.

In summary, Einstein's exploration of the Ricci tensor as a candidate for the left-hand side of the gravitational field equations had ended in an impasse. At the same time, this exploration had helped him to further extend his strategic resources. They now included, in particular, the consideration of a weak-field equation. Furthermore, the mathematical strategy was amplified by adding as a routine a compatibility check of the restrictions resulting from the correspondence and the conservation principles, respectively. As a result, the notion of coordinate restrictions as a virtually unavoidable consequence of combining a generalized relativity principle with other physical requirements was solidified. Perhaps the most important result of Einstein's exploration of the Ricci tensor was, however, the establishment of a sharp contradiction in the argumentative framework. The identification of this contradiction offered a range of fairly clear options of how to avoid it. Among the alternative pathways to explore was the option of changing the physical default settings entering his argument, in particular those for the metric tensor of a static field and for the stress-energy or energy-momentum tensor. Another option was to reconsider the implementation of the correspondence principle with the help of the harmonic coordinate restriction, e.g., by extracting a new candidate from the Riemann tensor with the help of a different coordinate restriction. Probing a different implementation of the correspondence principle probably looked like the more sensible option given that Einstein's reconsideration of the conservation principle had strongly confirmed *its* implications. In the course of his research, Einstein eventually pursued all of these options. The option he chose to explore first came courtesy of the new toy model he had introduced, the weak-field equation. Why should it not be possible to tinker with the field equations themselves, within the weak-field framework, in order to find out whether there really was no way to satisfy all requirements on the table, including the harmonic coordinate restriction?

6.13 Matching the Riemann Tensor and the Correspondence Principle: the Failure of the Linearized Einstein Tensor (20L–21R)

The preceding considerations had shown Einstein that the contradiction between the coordinate restrictions implied by the correspondence and conservation principles, respectively, had to be taken seriously enough to entertain even a modification of the form of the field equations. His starting point had been a weak-field equation obtained from the Ricci tensor by imposing the harmonic coordinate restriction to satisfy the correspondence principle. The most obvious conflict was that between the implication of the conservation principle that the trace of the stress-energy tensor of matter must vanish, on the one hand, and the default-setting for this tensor (XXI), on the other hand. Einstein's earlier experience with the adjustment of his original theory of the static gravitational field to the requirements of the conservation principle helped to make a modification of the field equation acceptable as a possible way out of this dilemma. In addition, the weak-field equations made the exploration of possible modifications easier by making it possible to study the interplay between the various constraints in a mathematically simplified form. While the overall logic of this exploration was dominated by the mathematical strategy, the challenges it produced for the various physical default-settings of Einstein's search for the gravitational field made it necessary to reflect on his heuristic presuppositions as well and to go back once more to the physical principles guiding his search such as the equivalence principle and even to the more secure part of his theory in the making, the equation of motion.

Since neither the conservation principle nor the default-setting for the stress-energy tensor of matter could be given up easily, the conflict between them first turned Einstein's attention to the source slot of the field equation, or rather on its default-setting, the energy-momentum tensor of matter according to eq. (XIV). If this default setting could be changed, the default setting **DUST** (XXI) for the stress-energy might well be retained without leading to a conflict with the conservation principle. By replacing the default-setting eq. (XIV) with a traceless quantity, Einstein was indeed able to avoid the conclusion that the trace of the stress-energy tensor has to vanish if the trace of the field equation vanishes, as it would have to as a result of combining harmonic and Hertz restrictions, as we have seen.

This remarkable achievement did not provide an entirely satisfactory solution to the compatibility problem of the correspondence and conservation principles. Einstein had resolved the conflict between the combined coordinate restrictions following from these principles and the default-setting for the energy-momentum tensor (XXI), the discrepancy between the combined coordinate restrictions and the default-setting for the metric tensor of a weak static gravitational field (25) still existed. The preceding experience had taught Einstein how modifying the field equation could help in dealing with disturbing coordinate restrictions. If that method had worked to get rid of the unwanted trace condition, why not try to use it again to get rid of the Hertz restriction altogether rather than to make it compatible with the harmonic coordinate restriction?

Once again, an unsuccessful line of thought had thus paved the way for an important strategic insight, which, in this case, gave Einstein the harmonically reduced and linearized Einstein tensor as a candidate for the left-hand side of the field equations. Instead of giving up and replacing the default-setting for the right-hand side of the field equations, changing the way in which the source-term enters the equation, he modified the way in which the gravitational potential enters the left-hand side of the equations, i.e. the default-setting for the weak-field version of **LAP** (XXVIII), the d'Alembert operator. In a sense, this may have appeared to Einstein as the more conservative approach because it interfered less with the canonical form of the field equation. More specifically, Einstein changed the left-hand side of the field equations by adding a trace term in such a way that the object on which **LAP** operates becomes equal to the left-hand side of the harmonic coordinate restriction if the divergence of this left-hand side is taken (cf. eq. (LXXIX)). In this way, the vanishing divergence of the energy-momentum tensor, which is required by the conservation principle and which originally resulted in the Herz restriction, is now implied by the harmonic coordinate restriction alone—without imposing an additional constraint (cf. eq. (76)).

Now that the correspondence principle and the conservation compatibility check in its weak-field form had been taken care of, the next step was to make sure that the conservation principle was satisfied in all of its facets. For the modified field equations, Einstein needed to check, in particular, whether the gravitational force could be represented as the divergence of a gravitational stress-energy expression. The weak-field equations passed this test without any problem, in spite of the additional trace term they involve (cf. eq. (75)). It was less obvious how this success could be extended to the full version of the equations. A half-hearted attempt to solve this problem was, apparently, enough for Einstein to see that this extension represented a major challenge and that it might even bring back additional coordinate restrictions.

After this preliminary exploration of the conservation issue beyond the weak-field case, Einstein returned to the correspondence principle and discovered that another conflict between the modified field equations and the default-settings of his search was still unresolved. The canonical metric for a static field (25) is no longer a solution of the modified field equations. Since the weak-field equation with the added trace term had otherwise fared fairly well in comparison to earlier candidates, it made sense to carefully reexamine the legitimacy of the one obstacle that remained, the default-setting for the weak static gravitational field. He tested its justification by physical knowledge in the same way in which he had earlier checked the legitimacy of the Hertz restriction when it proved to be an obstacle. He turned to the more solid ground provided by the equation of motion. Since the entire theory of the static field had, in a sense, originated from the equation of motion with the help of the equivalence principle, a check of the default assumption about the static gravitational field with the help of this principle was the most natural option to pursue. From this perspective, the crucial question was whether the default-setting for the weak static gravitational field was actually inescapable given the equivalence principle. What does Galileo's principle of equal acceleration in a gravitation field, on which the equivalence principle hinges, imply about the form of a metric tensor for a weak static gravitational field?

All Einstein had to do to address this question was to formulate Galileo's principle in terms of his metric formalism. The conceptual framework within which his question was formulated suggested to do so by trying to identify the elements of Newton's equation (cf. eq. (III)) within this formalism, in particular the force term and the mass (or energy) term. His earlier work on the equation of motion and his experience with the Lagrange formalism gave him the tools for writing down the required quantities. Since their interpretation was governed by the conceptual framework of Newtonian physics, Einstein could draw the conclusion that, if the force was to vary as the energy, so as to ensure the validity of Galileo's principle, the metric for a static field must take on its canonical form. As a consequence of this inference, based on combining physical with mathematical elements, in a way that in hindsight can be recognized as problematic, the default-setting for the weak static gravitational field became even more firmly rooted in Einstein's heuristic framework, making its clash with the linearized Einstein tensor so much the worse for the latter.

In summary, Einstein's attempt to match the Riemann tensor first with the correspondence principle and then with the conservation principle by setting up a field equation for which only the harmonic coordinate restriction was needed as a supplementary condition had left him in the end without a viable candidate to pursue. The promising candidate he had found in the process had to be rejected because it seemed to be irreconcilable with the equivalence principle. Thus, in this dramatic episode of the search for gravitational field equations, the Einstein tensor of general relativity was, albeit only in a weak-field approximation and for harmonic coordinates, identified and discarded. Clearly, the criteria that led to its rejection had to be changed before it could be accepted. In particular, the default-setting for the metric of a weak static field had to be given up, in spite of its support by the canonical form of the weak-field equation and the—in hindsight—spurious argument based on the equivalence principle. The rejection of the equations in the winter of 1912–1913 was a matter of heuristic criteria that were still rooted in classical physics and that were incompatible with general relativity as we know it today. It was also a matter of a network of arguments that were still too loosely woven to produce a contradiction between any candidate field equation and these classical criteria, that could seriously challenge the latter rather than leading only to the rejection of the former.

Again, the failure to establish an acceptable candidate field equation in this preceding episode strengthened Einstein's vision and generated new strategic insights. The mathematical strategy was now fully operative, from the extraction of a candidate from the Riemann tensor, via the introduction of a weak-field approximation, to the matching of the coordinate restrictions following from the correspondence and the conservation principle, respectively. Among the new insights may have been an appreciation of the difficulty in passing from a weak-field to a full-fledged implementation of the conservation principle. But before this problem could even be addressed another candidate equation was needed. The options for avoiding the original conflict between the coordinate restrictions resulting from the correspondence and the conservation principle, respectively, suggested a different implementation of the correspondence principle. The pathway toward such a different implementation and thus to a new candidate was, in a sense, suggested by the original conflict itself. So far, Einstein had tried to get rid of the Hertz restriction in order to follow the path indicated by the harmonic restriction. Since this path looked like a dead end, it made sense to abandon the harmonic restriction, retaining the Hertz restriction instead.

6.14 Matching the Riemann Tensor and the Conservation Principle: the Failure of the November Tensor (22L–25R)

The Ricci tensor and the Einstein tensor do not exhaust the potential represented by the Riemann tensor and the mathematical strategy for producing candidates for the gravitational field equations. As mentioned above, the direction in which to proceed was indicated by the as yet unresolved conflict between the correspondence and the conservation principles, which was embodied in the clash between two coordinate restrictions, the harmonic restriction and the Hertz restriction, respectively. Since the constraints imposed by the conservation principle appeared to be unavoidable, and since Einstein's earlier attempt to suppress the need for the Hertz condition had failed, he now explored the possibility of realizing the correspondence principle in a new way, without the help of the harmonic restriction.

It was once again Marcel Grossmann who prepared the ground for pursuing this other possibility. At the price of a restriction to unimodular transformations, the Ricci tensor could be split into two parts, each part individually transforming as a tensor under unimodular transformations. One of those two parts was a promising new candidate for the left-hand side of the field equations, the "November tensor" (cf. eq. (82)).

The November tensor has a surprisingly elegant form: the divergence of a Christoffel symbol plus a quadratic expression in the Christoffel symbols. If the Christoffel symbols were taken to represent the gravitational field (cf. eq. (XXIII)), the candidate would have the canonical form of eq. (XXXVIII). Such an interpretation, however, was in conflict with Einstein's heuristics at this stage, which demanded the implementation of the correspondence principle first by imposing an appropriate coordinate restriction; furthermore the default setting for the gravitational fields was given by eq. (XXII). At this point, Einstein only looked for an interpretation of a candidate in terms of field components once he had found the *reduced* field equations, i.e., once he had imposed a coordinate restriction to meet the demands of the correspondence principle (cf. eq. (LXXXI)).

As he had done before, Einstein expanded the Christoffel symbols in terms of derivatives of the metric to identify the disturbing second-order terms preventing the implementation of the correspondence principle. These disturbing second-order terms, it turned out, could be eliminated with the help of the Hertz restriction so that the harmonic restriction was no longer needed. Since the Hertz restriction also guarantees the vanishing of the divergence of the linearized stress-energy tensor, the conflict between correspondence and conservation principle was thus resolved, at least at the weak-field level.

Now that this major conflict was settled, new problems arose, among them the question of the transformations allowed by the reduced field equations and the question of the implementation of the conservation principle for the full field equations. Einstein first addressed the issue of covariance which, given the known transformational behavior of the November tensor and following a strategy first established in the context of the Beltrami invariants (cf. eq. (LIX)), could be addressed by exploring the transformation properties of the Hertz restriction. Einstein could also build on an earlier analysis of the Hertz restriction which seemed to indicate that transformations in Minkowski space to a linearly accelerated frame presented a problem, but that transformations to rotating frames did not.

A complete clarification of the transformation properties of the Hertz restriction could be obtained by a larger effort dealing with non-autonomous transformations. Before undertaking such an effort, Einstein preferred, it seems, to turn once more to the conservation issue. How could he extend his results concerning the conservation principle from the weak-field level to the full field equations? He must have been aware of the crucial role of the first-order correction terms to the core operator. Considering the reduced November tensor, i.e. the terms left of the November tensor after imposing the Hertz restriction, Einstein was confronted with a number of such first-order terms, destroying the simple structure which the new candidate displayed when written in terms of the Christoffel symbols. Einstein tried to reintroduce the Christoffel symbols. While this allowed him to group certain terms more effectively, the resulting expression became even more opaque, mixing as it did first-order derivatives of the metric and Christoffel symbols. It was difficult to see, on the basis of this expression, how the conservation principle for the full field equation could be satisfied.

At this point Einstein had an idea that may seem ingenious but whose grounds were prepared by the contrast between the simple and elegant original structure of the November tensor expressed in terms of the Christoffel symbols and its confusing complexity when written in terms of derivatives of the metric. What was needed was a preservation of the original structure in terms of what, in Einstein's understanding, would be the true representation of the gravitational field, viz. the first-order derivatives of the metric (cf. eq. (XXII)). The resulting candidate gravitation tensor would then be of the canonical form (XXXVIII) and, in all likelihood, comply with both the correspondence and the conservation principle. The idea was to impose a new coordinate restriction that would effectively allow Einstein to replace the Christoffel symbols by first-order derivatives of the metric. This was somewhat more difficult than the above sketch would suggest, as it required in particular the introduction of an indirectly defined coordinate restriction amounting to the stipulation that an object we have designated as the "theta expression" behaves as a tensor. It nevertheless proved to be fairly successful. Not only did Einstein manage to obtain a "theta-reduced November tensor" of the desired canonical form but, along the way he also found out that he no longer needed the Hertz restriction as an additional condition to recover the Newtonian theory.

The next challenge was to determine the covariance properties of this theta-reduced November tensor implied by the somewhat strange new coordinate restriction. Einstein first derived a general condition for the infinitesimal non-autonomous transformations leaving the theta expression invariant, which, however, was just as formidable as the conditions of this kind that he had encountered earlier. He then turned to a special case and tried to identify the class of transformations in Minkowski space that preserve the theta condition. In doing so, he found a puzzling result: among the metric tensors satisfying the theta coordinate restriction was a metric corresponding to Minkowski space in rotating coordinates but with interchanged covariant and contravariant components, an object we shall call the "theta rotation metric," or simply the "theta metric." This curious result continued to concern Einstein almost until the end of the research period covered by the Zurich Notebook.

What did this strange finding actually mean? Was rotation covered by the theta restriction or was it not? To answer this question, Einstein had to find a physical interpretation of the curious theta rotation metric, exploring whether or not it was possible to connect it to the dynamics of rotation. He did so in various ways. First he rederived the equations of motion with the help of the Lagrange formalism in order to identify Coriolis and centrifugal forces. He abandoned this approach because it became too involved. Then he switched covariant and contravariant components in the theta condition, since this reformulated condition would obviously admit the ordinary rotation metric as a solution. Finally, in yet another attempt to come to terms with the physical interpretation of the theta condition, Einstein took recourse to the law of energy-momentum conservation, reformulating it in terms of the covariant rather than the contravariant stress-energy tensor and trying to extract from the reformulated law and from the theta metric the correct expression for the centrifugal force. Due to an error, Einstein at first convinced himself that this was actually possible but then appears to have developed doubts.

While the physical meaning of the theta rotation metric remained obscure, its exploration, nonetheless, had two consequences for Einstein's subsequent work: First, it proved increasingly difficult to reach a comprehensive implementation of the generalized relativity principle, and rotation increasingly became something of a litmus test, the one case of accelerated motion that Einstein expected his theory to cover to comply with his original heuristic mission. Second, checking the theta metric with respect to the dynamics of rotation may well have directed Einstein's attention once again to the significance of the force expression as a clue to viable field equations.

Let us try to reconstruct such a clue by means of our symbolic expressions. It was indeed possible to derive a force expression from the linearized field equation, expressing it as the divergence of the gravitational stress-energy density (cf. eqs. (XXXIII), (XXVIII)):

$$\textbf{LIM(FORCE) = DIV(FIELDMASS).} \tag{XC}$$

If such an expression offers a physically meaningful starting point, for instance because it vanishes for rotation, it might serve as a criterion for picking a suitable gravitation tensor instead of just serving as an indirect consistency check by way of a particular solution such as the theta metric. In that case, the force expression could perhaps be reinterpreted as representing an exact quantity even though it was obtained in linear approximation. The force expression could thus become, in a way similar to the transition from Einstein's first to his second theory of the static gravitational field, the starting point for extracting a suitably corrected full gravitation tensor from it (cf. eq. (XXXVI)):

$$\textbf{FORCE = GRAV x FIELD} \tag{XCI}$$

A gravitation tensor **GRAV** constructed in this way would automatically satisfy the conservation principle and looked promising with respect to the generalized relativity principle, at least as far as rotation was concerned. After all, it fulfilled a necessary condition for being compatible with the relativity of rotation, the vanishing of the corresponding force expression in the case of rotation.

If Einstein were in fact trying to implement such ideas, he ran into a number of difficulties, caused in part by calculational errors. First of all, Einstein did at first not systematically construct a candidate gravitation tensor **GRAV** but seems to have merely guessed it. Second, the gravitation tensor he extracted from the force equation does not vanish for rotation as he had hoped, but then he found that this extraction itself involved errors whose elimination might well yield the desired result after all. Third, he must have realized that by postulating a physically meaningful force expression as his new starting point he effectively abandoned the link with the November tensor with its well-defined transformation properties. It therefore made sense to interpret the candidate field equation extracted from the force expression not as the definitive result of a physical strategy but rather as the new preliminary target, itself subject to further corrections, of the mathematical strategy starting from the November tensor. In this way, the advantage of well-defined transformational properties might be combined with that of a physically meaningful force expression ensuring the satisfaction of the conservation principle and perhaps even covering the generalized relativity principle for the case of rotation.

In a sense, Einstein may have reached once again reached the constellation he had reached earlier when establishing the core operator as the physically meaningful target (modulo correction terms) of a mathematical strategy taking the second Beltrami invariant as its starting point. Now the place of the Beltrami invariant was taken by the November tensor and that of the core operator by a candidate gravitational field equation that received its physical meaning not just from the correspondence principle but from the conservation principle as well. In the end, however, Einstein once again was unable to build a convincing bridge between his mathematical starting point and his physically meaningful target.

Unable to build a bridge between the two, Einstein put the November tensor to one side for the time being and explored the field equation suggested by his force expression. However, at this bifurcation point of his research, he does not seem to have had, perhaps for the first time, any promising idea about how to proceed. Neither his overall heuristics nor his remarkable ability to draw strategic lessons from failure suggested a plausible next step. The entries in the notebook at this point do not seem to follow any coherent and well-defined strategy. As noted above, the field equation suggested by the force expression vanishing for rotation does itself not vanish for the rotation metric. Does it perhaps vanish for the curious theta rotation metric obtained by interchanging covariant and contravariant components? This question may sound absurd but was nevertheless pursued by Einstein. He even convinced himself—arbitrarily adjusting a coefficient—that his candidate field equation does indeed vanish for the theta metric, a conclusion that is in fact erroneous. This specious result encouraged him to resort to an earlier trick: if the candidate field equation vanishes for the theta metric, a new candidate field equation could be constructed by interchanging contravariant and covariant components that would vanish for the ordinary rotation metric. The new candidate resulting from this crude operation was mathematically ill-defined. Nevertheless, Einstein explored it. It covered, or so he may have believed, the case of rotation, was compatible with the correspondence principle, and looked promising as far as the conservation principle was concerned.

This last issue called for a closer examination and brought Einstein back to the starting point of this phase of his search, the expression for the force density. The new candidate would be compatible with the conservation principle if it gave rise to a force density that can be represented as the divergence of a stress-energy expression for the gravitational field. Now that he had found an apparently viable candidate complying with the rotation criterion by merely formal manipulations, it made sense to repeat the procedure that originally brought him to the expression for the force density and that had been the point of departure of this whole line of reasoning. Following this procedure, Einstein began to write down the force density for the linearized version of the new candidate field equation, which he then tried to make exact. If everything worked out as expected, his procedure should correspond to the transition from eq. (XC) to eq. (XCI) so that he should be able to reconstruct his full candidate in this way, that is, essentially from inserting the right candidate for **LAP** into the expression for the force expression **LIM(FORCE)** and then generating the correction terms yielding **GRAV**.

Unfortunately, things did not work out in the end. Einstein managed to extract terms from **LIM(FORCE)** that had the required form of a divergence or that could be put on the left-hand side of the field equation as correction terms, but he also encountered a term that could not be treated in either of these two ways. But he got close as only one disturbing term remained. Remarkably, the terms that were put on the left-hand side of the field equation not only induced correction terms of the form **CORR(POT) x FIELD** but also a term of the form **LAP(POT) x FIELD**. This suggested that Einstein's expansion of **LIM(FORCE)** might actually produce an identity if only the right expression for **LAP(POT)** was taken as the starting point. Eventually, Einstein nonetheless abandoned the entire calculation, probably not only because he failed to establish the compatibility of his candidate with the conservation principle but also because he may have realized at some point that this candidate did not make good sense mathematically in the first place.

In summary, Einstein had extracted yet another candidate from the Riemann tensor in addition to the Ricci, the harmonically reduced Ricci, and the harmonically reduced Einstein tensors: the November tensor. In the end, this candidate was judged to be just as unsatisfactory as its predecessors. Its original appeal gradually waned because of the problems Einstein ran into when he tried to turn the November tensor into a viable candidate by adding appropriate coordinate restrictions. The discouraging result was that it hardly made any difference whether a candidate resulted from reducing a covariant object by additional coordinate restrictions or whether it was merely constructed *ad hoc*. Either way, the main challenges, the compatibility with rotation and the satisfaction of the conservation principle, had to be addressed directly, by explicit construction. As a consequence, Einstein's mathematical strategy lost its appeal and gave way to another tinkering phase.

In this tinkering phase Einstein focused on the expression for the gravitational force which had the advantage of having a clear physical interpretation. Such an expression had already played a key role in the transition from his first to his second theory of the static gravitational field. If the force can be written as a divergence, the conservation principle is satisfied automatically. And if the expression for the force happened to vanish for rotation, there was at least a chance of meeting some of the demands of the generalized relativity principle as well. Einstein's problem was that he had yet to find a way of systematically extracting a candidate gravitation tensor from such a force expression. His attempts to construct or guess candidate gravitation tensors along this line tended to destroy the promise of his initial ansatz. Given a candidate consisting of some version of the core operator plus correction terms suggested by a force expression, it still had to be checked against the conservation principle. This in turn meant forming a force expression from the linearized field equation which then was to be rewritten as a divergence, possibly with the help of introducing new correction terms to the original ansatz. Having tried this procedure once if not twice without being able to reproduce his original ansatz, Einstein noticed that he could actually begin simply with the core operator and *use the conservation principle as a means for producing correction terms* to it. The realization of this possibility was the birth of the *Entwurf* strategy and, as far as the research documented in the Zurich Notebook is concerned, the end of the mathematical strategy.

6.15 Matching Correspondence and Conservation Principles: The Emergence of the Entwurf Equations (25R–26R)

As the November tensor gradually dropped out of sight, the mathematical strategy launched with the introduction of the Riemann tensor into Einstein's research fell victim to the attrition associated with the efforts to realize Einstein's physically motivated heuristics when starting from a generally-covariant object. His principal instrument for implementing the correspondence and the conservation principles was to impose coordinate restrictions, which, first of all, had to be brought into agreement with one another and then tended to consume the original benefit provided by a generally-covariant starting point. Meanwhile, he had developed more concise ideas about what a viable candidate satisfying both the correspondence and the conservation principles should look like. In the course of his research he had even encountered candidates that seemed close to satisfying these heuristic criteria. Einstein, however, never succeeded in building a bridge between a mathematically viable starting point such as the November tensor and a candidate that looked promising from a physical point of view. While the November tensor, in particular, was never quite refuted, it just became more and more uninteresting.

On a heuristic level, Einstein's difficulties in implementing simultaneously the correspondence and the conservation principles counteracted the potential advantage of starting from a candidate satisfying the generalized relativity principle. Even when the battle was won at the weak-field level, the mathematical strategy failed to provide any hint for winning it at the level of the full equations. Instead, such a hint came from mere formal manipulations of the force expression that had already guided Einstein's pathway from his first to his second theory of the static field. Against the background of his prior experience with a physical strategy and the inadequacy of the mathematical strategy to cope with the conservation principle, this hint prepared the ground for the derivation of the *Entwurf* field equation.

The gist of this derivation consists in starting from the force expression for the core operator which is then transformed into a divergence expression plus terms which are identified as correction terms (cf. eq. (XXXVII)). The resulting identity then yields both the correction terms and the gravitational stress-energy expression whose divergence corresponds to the force expression for the definitive gravitation tensor. The gravitation tensor produced in this fashion even happened to involve the gravitational stress-energy expression in such a way that the field equation could be written in the canonical form of eq. (XXXIX), with the energy-momentum of the gravitational field entering the field equations on the same footing as the energy-momentum tensor of matter (cf. eqs. (49) and (50)). This strategy was the result of Einstein's reflection on his earlier attempts to generalize the representation of the force as a divergence expression from the weak-field to the general case. Rather than guessing the right correction terms, he had now found a systematic construction procedure, which seemed to uniquely identify a candidate gravitation tensor compatible with both the correspondence and the conservation principle.

The match between these two heuristic principles was achieved at the expense of the generalized principle of relativity. All that could be known in that respect about a candidate gravitation tensor produced in this way was its covariance under linear transformations (cf. eq. (LXVII)). Einstein was ready to accept this consequence. He had already turned his attention to a simplified approach encompassing only linear transformations once before so that he could get a better handle on the conservation principle. This was when his earlier attempts to determine the covariance properties of the core operator with or without the help of the Beltrami invariants had run into similar difficulties as his efforts involving to the Riemann tensor. While, at that point, the restriction to linear transformations was merely a presupposition for formulating the conservation problem, Einstein may now have felt that this was the price to pay for its solution.

In the course of Einstein's pursuit of the mathematical strategy, the conservation principle had emerged as the major challenge for his search for the gravitational field equation. This challenge triggered the switch to a physical strategy, judiciously incorporating results found while pursuing the mathematical strategy, from the form of the field equation to the role of coordinate restrictions. With the establishment of the *Entwurf* field equations with the help of this physical strategy Einstein had succeeded, for the first time in the course of his research documented by the notebook, to satisfy the conservation principle without restriction to the weak-field level. The major challenge for Einstein's research now was the generalized principle of relativity. How far could the covariance properties of the *Entwurf* field equation be extended or was their covariance really restricted to linear transformations only? Why were the physically satisfying *Entwurf* field equations not generally-covariant to begin with? These were the questions delineating Einstein's research program for the further exploration of the *Entwurf* equation. Its compliance with the heuristic principles rooted in classical physics, the correspondence and the conservation principle, made it possible to consider these questions not as incentives for continuing the search for gravitational field equations but as remaining puzzles within an established conceptual framework, that of the *Entwurf* theory. For the time being, Einstein's search for the gravitational field equations was over—even if this meant turning his back on a reservoir of possible further candidates.

7. PROGRESS IN A LOOP: EINSTEIN'S GENERAL RELATIVITY AS A TRIUMPH OF THE "*ENTWURF*" THEORY IN THE PERIOD FROM 1913 TO 1915

7.1 Consolidation, Elaboration, and Reflection

This chapter focuses on what has traditionally been seen as the most uneventful period of Einstein's search for a generalized theory of relativity, the time between spring of 1913 and fall of 1915, in which he clung to the erroneous *Entwurf* theory, which he published together with Marcel Grossmann before the end of June 1913. According to the dramatic narratives of the emergence of general relativity, this period was one of stagnation, it was the calm interval between two major thunderstorms, Einstein's tragic struggle with and eventual rejection of generally-covariant field equations in the winter of 1912–1913 in favor of a theory with only limited covariance properties and the sudden revelation of errors in the *Entwurf* theory, which led immediately to its demise and then to a triumphant, if gradual, return to generally-covariant field equations in the fall of 1915. The assumption of a "pitfall" in 1912–1913 and of a "breakthrough" in late 1915 constitutes the traditional explanation for the most peculiar feature of the genesis of general relativity, Einstein's double discovery of generally-covariant gravitational field equations, first formulated in 1912 and then rediscovered in 1915. How else, if not by the introduction and later elimination of errors, can this closed loop be explained?[106]

From the perspective of an historical epistemology, the supposed period of stagnation between 1913 and 1915 can be considered a period in which new knowledge was assimilated to a conceptual structure still rooted in classical physics. As a result of this assimilation of knowledge, this conceptual structure became richer, both in terms of an ever more extended network of conclusions that it made possible, and in terms of new opportunities for ambiguities and internal conflicts within this network. It was this gradual process of enrichment that eventually created the preconditions for a reflection on the accumulated knowledge which, in turn, induced a reorganization of the original knowledge structure. The enrichment of a given conceptual structure by the assimilation of new knowledge and the subsequent reflective reorganization of the enriched structure are the two fundamental cognitive processes which explain the apparent paradox that the preconditions for the formulation of general relativity matured under the guidance of a theory that is actually incompatible with it.

As we will argue in the following, the results achieved on the basis of the *Entwurf* theory should not be understood as so many steps in the wrong direction, whereupon it appears that their only function was to make the deviation from the truth evident, but rather as instruments, first for accumulating knowledge and then for rearranging it in a new order. Both these processes are essential to the development of scientific knowledge. The second half of this chapter covers Einstein's papers of November 1915, with the intention to demonstrate the role in these papers of insights and techniques developed in the period before.

When elaborating the *Entwurf* Theory, Einstein still pursued the same heuristics that had shaped his search for a gravitational field equation in the winter of 1912–1913 as documented by the Zurich Notebook, although the heuristics were now governed by the perspective of consolidation rather than by that of exploration of alternatives. In particular, the unresolved tensions between Einstein's heuristic principles guided his attempts to consolidate the *Entwurf* theory. These attempts were characterized by two complementary approaches. Following a defensive approach, he attempted to justify the restricted covariance of the *Entwurf* field equations by arguments based on the knowledge of classical physics. Following a bold approach, he attempted to look for a generalization of the relativity principle even in the framework of the *Entwurf* theory. The outcome of these efforts was, as we shall see, that he eventually succeeded in both, the defensive and the bold approach.

In the first phase of the consolidation period of the *Entwurf* theory, lasting roughly from spring to summer 1913, Einstein formulated two problematic argu-

106 This section relies heavily on (Renn 2005c) and "Untying the Knot ..." (in vol. 2 of this series).

ments, an argument based on the consideration of energy-momentum conservation and the notorious hole argument, both justifying the restricted covariance properties of the theory. Although these arguments later turned out to be erroneous, they were nevertheless significant in bringing out decisive conceptual peculiarities of general relativity that distinguished it from classical theories, including the *Entwurf* theory. The further development, refutation or clarification of these arguments revealed the non-locality of energy-momentum conservation, as expressed by the nonexistence of a local energy-momentum tensor for the gravitational field, the impossibility of ascribing physical significance to single spacetime points independent of the metric, and the fundamental connection between conservation laws and symmetries of spacetime structure later explicated in the Noether theorems.

In the second phase of the consolidation of the *Entwurf* theory, roughly lasting from fall 1913 to the end of 1914, Einstein elaborated this theory, which he originally found along the physical strategy, from the point of view of the complementary mathematical strategy. Guided by this heuristic strategy, Einstein found a new derivation of the *Entwurf* field equations, which he completed by the fall of 1914. The second phase of the consolidation period of the *Entwurf* theory had, like the first phase, a paradoxical character. On the one hand, Einstein's findings stabilized the *Entwurf* theory, on the other hand they provided instruments for overcoming the objections that had earlier prevented him from accepting candidate gravitational field equations found along the lines of the mathematical strategy. In particular, the variational techniques explored in the context of the new derivation of the *Entwurf* theory made it possible for Einstein to solve one of the crucial problems associated with the candidates for a gravitational field equation that he had discarded in the winter of 1912–1913, the establishment of energy-momentum conservation. That he did not immediately draw this consequence was partly a matter of perspective since, from the point of view governing the consolidation period, there was no reason for reexamining the earlier candidates.

Having discovered flaws in the *Entwurf* theory and its derivation along a modified mathematical strategy, Einstein eventually abandoned the consolidation phase and subsequently returned to a new exploratory phase, searching once more for the correct gravitational field equation. In hindsight, he gave three reasons for his rejection of the *Entwurf* theory: It could not explain the perihelion shift of Mercury; it did not allow the interpretation of a rotating system as being equivalent to the state of rest, and hence did not satisfy his Machian expectations, and finally, the derivation of its field equations along a mathematical strategy involved an unjustified assumption. For a short time, the theory survived *all* of these problems. Even the last problem, the discovery of a flaw in the derivation of the field equations, did not lead to a refutation of the Entwurf theory but only to a successful attempt of repairing it on a technical level. But the discovery of this problem had nevertheless fargoing consequences on the level of Einstein's reflection on the results he had achieved. By its very nature this discovery had a double effect:

– It showed that the adaptation of the mathematical strategy to the *Entwurf* theory failed and forced Einstein to return to the arguments at the core of the physical strategy as the only possible justification of the *Entwurf* theory.
– It showed that the mathematical strategy adapted to the *Entwurf* theory did not single out this theory but rather opened up the possibility of examining other candidate field equations. And he *needed* new equations because of the problem of rotation.

Together with the other short-comings found earlier, the discovery of the error in the derivation of the field equations, after a period of reflection, caused Einstein to drop his attempts to consolidate the *Entwurf* theory and eventually brought him back to an exploratory phase.

The second part of this chapter deals with the short period of three weeks before Einstein presented the definitive field equations of general relativity to the Prussian Academy on 25 November 1915. This period began when Einstein started to check whether the *Entwurf* field equations are necessarily the only solution to his problem and thus returned to his 1912–1913 attempts to search for a solution by examining candidate field equations familiar from his pursuit of the mathematical strategy, the November tensor, the Ricci tensor, and the Einstein tensor. By focusing on the impact of Einstein's achievements under the reign of the *Entwurf* theory, it is possible to answer the question of why in 1915 he could accept field equations that he had rejected in 1912–1913. In a note Einstein submitted to the Prussian Academy on 4 November 1915, he proposed a new gravitation theory based on the November tensor, considered earlier in the Zurich Notebook. In contrast to the situation in 1912–1913, he was now able to demonstrate that the new theory complies with the conservation principle. Just as he had done in the *Entwurf* theory, Einstein continued to interpret the conservation principle as implying a restriction of the admissible coordinate systems which now, however, turned out to be much less restrictive than the condition he had earlier found on the basis of his examination of the weak field equation (cf. eq. (LXXXVI)). He thus reached a decoupling of the coordinate restrictions implied by the conservation and the correspondence principles, respectively. Reflecting on this decoupling, Einstein was now able, for the first time, to conceive the choice of coordinates required for implementing the correspondence principle as a coordinate condition in the modern sense.

In an addendum to the note published on 4 November, Einstein reinterpreted another already familiar candidate in a new context, the Ricci tensor. This new context was provided by a speculative electromagnetic theory of matter, probably stimulated by the contemporary work of David Hilbert on such a theory. Due to this new context, Einstein shifted the restriction on the choice of coordinates, which he had found for the theory based on the November tensor, to a restriction of the choice of a particular kind of matter acting as the source of gravitational fields. Einstein thus arrived at a generally-covariant theory based on the Ricci tensor, which he considered as being merely a reinterpretation of the theory based on the November tensor so that he could take over essential results such as those concerning energy-momentum conservation.

He thus partly resolved—on the basis of the *Entwurf* theory—and partly circumvented—on the basis of an electromagnetic theory of matter—the objections he had earlier encountered against such a theory when he first considered it in 1912–1913.

In a paper submitted on 18 November 1915 Einstein calculated the perihelion shift of Mercury, claiming to provide support for the hypothesis of an electromagnetic nature of matter on which his new theory of gravitation was based. In a sense, the Mercury problem now offered a theoretical laboratory for the Ricci Tensor. Einstein's paper is largely based on techniques he had developed jointly with Besso in 1913 in the context of the *Entwurf* theory. It also includes the crucial insight that the determination of the Newtonian limit for a gravitational field equation involves, in general, more complex considerations than originally envisioned along the physical strategy, and that were used earlier to object to the harmonically reduced Ricci tensor in the Zurich Notebook. This insight into the complex nature of the correspondence principle had already been attained in 1913 as well, in the context of the *Entwurf* theory (at least by Besso) but was then of no relevance as Einstein and Besso did their original calculations in the consolidation phase of this theory.

Einstein's more sophisticated understanding of the Newtonian limit had, in the context of the renewed exploratory phase at the end of 1915, decisive consequences: It made it possible for him, in his final paper of that period, to base a theory of gravitation on the Einstein tensor, whose harmonically reduced and linearized form had been rejected in 1912 because of its apparent incompatibility with the correspondence principle. Also the status of energy-momentum conservation changed in the new theory. The insight into its different status was a consequence and not a presupposition of the establishment of the definitive version of general relativity, which on Einstein's part was established entirely in the conceptual framework of the *Entwurf* theory. The *Entwurf* theory and general relativity were initially not separated by a conceptual gulf, but merely by technical insights on the one hand, and a change of perspective on the other. Remarkably, these were both the result of the same process, the elaboration of the *Entwurf* theory during the supposed period of stagnation. Since the technical achievements attained in this period could still have been, in principle, assimilated to the theory that had given rise to them, taken by themselves they would have induced only a linear progress, thus yielding an increasingly sophisticated and increasingly complex *Entwurf* theory. It thus becomes clear that in light of the new technical achievements of the consolidation phase of the *Entwurf* theory, Einstein's reflection on his earlier knowledge, including previously discarded candidate gravitation tensors, was the crucial process that made the establishment of general relativity the result of progress in a loop.

7.2 The First Phase of the Consolidation Period of the Entwurf Theory: The Defensive and the Bold Approach

With the formulation of the *Entwurf* field theory and its publication by Einstein and Grossmann in the spring of 1913, the search for the field equations, as documented by the Zurich Notebook, had manifestly come to an end. Einstein no longer examined different candidates by comparing them with his heuristic expectations. Instead, he used his growing mastery of the mathematical representation to develop the one most promising candidate he had found. He thus entered the consolidation period of the *Entwurf* theory. A consolidation of the *Entwurf* theory was necessary in view of the main problem left open by the *Entwurf* paper of 1913, the determination of the covariance properties of the field equations and thus of the extent to which the new theory realized the generalized principle of relativity. The covariance was more restricted than that of the candidate gravitation tensors derived from the generally-covariant Riemann tensor that had formed the points of departure of Einstein's mathematical strategy. It therefore made sense to address this problem by trying to explain and justify the restricted covariance of the *Entwurf* equations and to explore these covariance properties in the hope of generalizing the relativity principle as much as possible.

Einstein's probing of these two approaches came to a first conclusion in August 1913. All his bold efforts up to that point to identify by explicit calculations non-linear transformations under which the *Entwurf* field equations might be covariant had failed, including an attempt to show that the metric for Minkowski spacetime in rotating coordinates is a solution of these equations. At the same time, his defensive efforts had led to a first result. Not surprisingly, this result was based on the conservation principle which had earlier motivated a restriction of the generalized principle of relativity on several occasions in the Zurich Notebook. An interpretation of the expression for energy-momentum conservation in the *Entwurf* theory, following the model of classical and special-relativistic physics, was now taken by Einstein to indicate that the *Entwurf* theory is covariant only under linear transformations. Both results, the failure of his attempts to identify non-linear transformations and the conservation argument, as we shall call it, thus pointed in the same direction and encouraged Einstein to look for further arguments along the defensive.

By the end of August 1913, he found, quite possibly in discussion with Michele Besso,[107] another argument against general covariance, the so-called "hole argument," which is based on the assumption, again motivated by classical and special-relativistic physics, that points in spacetime can be identified by means of coordinate systems, independently from any physical processes. In a formulation by Besso, the argument merely seeks to express the non-uniqueness of the metric tensor in terms of two distinct sets of functions which solve the same set of differential equations with given boundary values. Einstein elaborated this argument to a construction of two distinct solutions for the metric tensor considered within one and the same coordinate system. This more sophisticated version of the hole argument makes use of the idea, in modern parlance, to drag values of the metric tensor from one spacetime point to the other and later raised the important question of which aspects of a generally-covariant theory are physically meaningful.

---

107 See "What Did Einstein Know ..." (in vol. 2 of this series).

By means of the hole argument, Einstein convinced himself that generally-covariant gravitational field equations, together with boundary values, do not determine uniquely the metric tensor representing the gravitational field. Having thus identified an apparently fundamental reason for rejecting general covariance, he interpreted his earlier argument from the conservation principle as providing the necessary specialization of the reference frames to be used within the theory. With these results, the *Entwurf* theory had come to a certain closure, ending the first phase of its consolidation period.

7.3 The Failure of the Generalized Principle of Relativity: A Conflict Between Formalism and Physical Intuition

Einstein's decision to settle for the non-generally-covariant field equations of the 1913 *Entwurf* paper was the consequence of his failure to find generally-covariant equations and not of a conviction or an insight that such equations could not exist. In the spring of 1913, he could not be sure that he had just failed to find generally-covariant equations that would fulfill his hopes for fully implementing the generalized principle of relativity. In the notebook, he had considered several candidates for generally or at least unimodularly covariant field equations and found them defective. But the fact that these candidates failed for *different* reasons must have made it difficult for Einstein to accept that generally-covariant field equations did not exist since these different reasons did not include a clear hint as to why a full implementation of the generalized relativity principle could not exist. Either the conflict with the realization of the Newtonian limit as required by the correspondence principle, or with the demonstration of energy-momentum conservation as required by the conservation principle, or both, led to the rejection of a promising candidate, but these conflicts by themselves did not provide any counter-argument against the possibility of the generalized principle of relativity. The failure to find generally-covariant field equations was, after all, merely a technical result, incompatible with the physical intuition incorporated in the generalized principle of relativity.[108] The conflict between formalism and physical understanding motivated Einstein's further elaboration of the *Entwurf* theory.

If Einstein, at the time of the notebook or of the publication of the *Entwurf* paper, had seen any reason to modify or restrict this principle, he might have done so explicitly in order to justify his failure to achieve its full implementation. In a letter to Ehrenfest from May 1913, in which he announced the forthcoming publication of the *Entwurf* paper, he asserts his firm belief in a generalized principle of relativity, but points out that he had been unable to realize this principle on the level of the theory's formalism:

---

108 For discussion of the problematic relation between the physical intuition incorporated in the generalized principle of relativity and general covariance, see (Janssen 2005).

> I slowly convinced myself that *privileged coordinate systems do not exist at all*. However, I succeeded only partly in arriving at this position also from a formal point of view.[109]

In the *Entwurf* paper itself, Einstein refers to the conflicts between the generalized relativity principle and his other heuristic principles in order to justify the new theory's lack of general covariance (Einstein and Grossmann 1913, 11). He emphasized, in particular, the difficulties he had found in realizing the correspondence principle, suggesting a second-order field equation, as a justification for his failure to achieve a generally-covariant field equation. He also admitted that his introduction of the *Entwurf* field equations was merely based on plausible assumptions and not on a strict derivation from postulates such as a generalized principle of relativity.

The failure to find generally-covariant field equations was most evident from an intrinsic asymmetry of the *Entwurf* theory, between the non-generally-covariant field equation and the generally-covariant equation for material processes in a gravitational field, i.e., the equation for energy-momentum conservation. This asymmetry is also emphasized in the *Entwurf* paper itself. Einstein attempted to interpret it as a clue for justifying the failure to establish a generally-covariant field equation, pointing at the different ways in which the metric tensor enters into the equation for energy-momentum conservation, on the one hand, and the field equation, on the other:

> This exceptional position of the gravitational equations in this respect, as compared with all of the other systems, has to do, in my opinion, with the fact that only the former can contain second derivatives of the components of the fundamental tensor.[110]

In a sense, he simply turned the description of the problem into its solution at this point. Einstein's problem remained that this assertion was merely speculative and, in the final account, based on nothing but his failure to find appropriate, generally-covariant, second-order gravitational field equations.

7.4 The Failure of Einstein's Search for Non-Linear Transformations

While in the spring of 1913 Einstein made his first attempts at justifying the lack of general covariance of the *Entwurf* field equation, he tried, at the same time, to overcome this problem. In fact, it could not be excluded that, even though all derivations in the *Entwurf* theory merely involve the assumption of linear covariance, the field equations would turn out to be covariant under a wider class of transformations. In the *Entwurf* paper, this question is singled out as the most important one left to be resolved (Einstein and Grossmann 1913, 18).

---

109 "Die Überzeugung, zu der ich mich langsam durchgerungen habe, ist die, *dass es bevorzugte Koordinatensysteme überhaupt nicht gibt*. Doch ist es mir nur te[i]lweise gelungen, auch formal bis zu diesem Standpunkt vorzudringen." Einstein to Paul Ehrenfest, 28 May 1913, (CPAE 5, Doc. 441).

110 "Die diesbezügliche Ausnahmestellung der Gravitationsgleichungen gegenüber allen anderen Systemen hängt nach meiner Meinung damit zusammen, daß nur erstere zweite Ableitungen der Komponenten des Fundamentaltensors enthalten dürften." (Einstein and Grossmann 1913, 18)

There are indications that Einstein attempted to find out by calculation whether the *Entwurf* field equation transforms also under a wider class of transformations. From his Zurich Notebook, he was familiar with techniques for exploring the transformational behavior of field equations. But the *Entwurf* field operator confronted him with a case that was far more complex than any other candidate in the notebook for which he had attempted to determine the transformational behavior by directly subjecting it to coordinate transformations.

One indication for Einstein's explorative attempts and their failure comes from a couple of pages of the so-called Einstein-Besso manuscript, pages that were probably written around June 1913 and that deal specifically with the problem of rotation.[111] Another indication comes from a letter Einstein wrote on 14 August 1913 to Hendrik A. Lorentz. This letter marks the preliminary end of Einstein's search for non-linear transformations of the *Entwurf* field equation and provides a succinct resume of the situation of the *Entwurf* theory just on the verge of the renouncement of the bold approach. The letter begins with the confession that the lack of general covariance represents a profound dilemma for the new theory of gravitation:

> And now to gravitation. I am delighted that you so warmly espouse our investigation. But, unfortunately, there are still such major snags in the thing that my confidence in the admissibility of the theory is still shaky. So far the "Entwurf" is satisfactory insofar as it concerns the effect of the gravitational field on other physical processes. For the absolute differential calculus permits the setting up of equations here that are covariant with respect to arbitrary substitutions. The gravitational field ($g_{\mu\nu}$) seems to be the skeleton, so to speak, on which everything hangs. *But, unfortunately, the gravitation equations themselves do not possess the property of general covariance*. Only their covariance with respect to *linear* transformations is certain. But all of our confidence in the theory rests on the conviction that an acceleration of the reference system is equivalent to a gravitational field. Hence, if not all of the equation systems of the theory, and thus also equations (18), permit other than linear transformations, then the theory refutes its own starting point; then it has no foundation whatsoever.[112]

The letter to Lorentz continues with a description of Einstein's unsuccessful attempts to find non-linear transformations under which the *Entwurf* field equation remains covariant, discussing two types of transformations, autonomous and non-autonomous ones.[113] In the Zurich Notebook, he had attempted on several occasions to find the transformational properties of a physically plausible candidate by deriving differential equations for the transformation coefficients involving the metric tensor. But he had never found a simple solution to the problem posed in this way. In view of the many reasons in favor of the *Entwurf* theory, he must have applied this technique to it with even greater persistence. Einstein's letter to Lorentz shows, however, that these efforts remained as unsuccessful as they had been in the Zurich Notebook. He was, in fact, ready to give up the bold approach to solve the most fundamental problem of the *Entwurf* theory and turn to the defensive approach, searching for more substantial

111 See (CPAE 4, Doc 14 [pp. 41–42]) and the discussion in (Janssen 1999).

arguments to justify the lack of covariance of the field equations than those adduced in the *Entwurf* paper.

7.5 Einstein's Reinterpretation of the Conservation Principle

For the candidates which Einstein had encountered pursuing the mathematical strategy, the conservation principle typically implied a restriction of the generalized principle of relativity. Given this experience, it must have been plausible for him to examine whether an explanation for the restricted covariance of the *Entwurf* field theory could perhaps also be found in the context of energy-momentum conservation. Only a day after Einstein sent the letter to Lorentz quoted above, on 15 August 1913, he indeed found a way to justify the limited covariance of the *Entwurf* theory on the basis of the conservation principle. A crucial heuristic ingredient of his argument was the parallelism between gravitational energy and other forms of energy, represented in the *Entwurf* theory by eq. (51). In another letter to Lorentz, written on 16 August 1913, Einstein wrote:

> Furthermore, yesterday I found out to my greatest delight that the doubts regarding the gravitation theory, which I expressed in my last letter as well as in the paper, are not appropriate. The solution of the matter seems to me to be as follows: The expression for the energy principle for matter & gravitational field taken together is an equation of the form (19), i.e., of the form $\sum \frac{\partial F_{\mu\nu}}{\partial x_\nu} = 0$; starting out from this assumption, I set up

---

112 "Nun zur Gravitation. Ich bin beglückt darüber, dass Sie mit solcher Wärme sich unserer Untersuchung annehmen. Aber leider hat die Sache doch noch so grosse Haken, dass mein Vertrauen in die Zulässigkeit der Theorie noch ein schwankendes ist. Befriedigend ist der Entwurf bis jetzt, soweit es sich um die Einwirkung des Gravitationsfeldes auf andere physikalisch[e] Vorgänge handelt. Denn der absolute Differenzialkalkül erlaubt hier die Aufstellung von Gleichungen, die beliebigen Substitutionen gegenüber kovariant sind. Das Gravitationsfeld ($g_{\mu\nu}$) erscheint sozusagen als das Gerippe an dem alles hängt. *Aber die Gravitationsgleichungen selbst haben die Eigenschaft der allgemeinen Kovarianz leider nicht*. Nur deren Kovarianz *linearen* Transformationen gegenüber ist gesichert. Nun beruht aber das ganze Vertrauen auf die Theorie auf der Überzeugung, dass Beschleunigung des Bezugssystems einem Schwerefeld äquivalent sei. Wenn also nicht alle Gleichungssysteme der Theorie, also auch Gleichungen (18) [i.e. the gravitational field equations] ausser den linearen noch andere Transformationen zulassen, so widerlegt die Theorie ihren eigenen Ausgangspunkt; sie steht dann in der Luft." Einstein to Hendrik A. Lorentz, 14 August 1913, (CPAE 5, Doc. 467). In view of the later development, in which the question of whether the manifold with its coordinate systems or the metric tensor is the "skeleton" on which all physical processes depend acquired a certain significance, it is remarkable that in the above formulation Einstein singled out the metric tensor as the crucial object. As we will see below when discussing the hole argument, in defending the lack of general covariance of his field equations Einstein for a while assumed that the points of the manifold identified by certain sets of coordinates actually have a reality and physical significance by themselves, that is, also independently from the metric tensor.

113 This fact also suggests that Einstein at this point did not assume that the *Entwurf* field equations remain covariant under rotations.

> equations (18). But a consideration of the general differential operators of the absolute differential calculus shows that an equation so constructed is never absolutely covariant. Thus, by postulating the existence of such an equation, we tacitly specialized the choice of the reference system. We restricted ourselves to the use of such reference systems with respect to which the law of momentum and energy conservation holds in this form. It turns out that if one privileges such reference systems, then only more general linear transformations remain as the only ones that are justified.[114]

With this insight, the conservation principle was no longer merely a technical impediment to the full implication of the generalized relativity principle but provided the concrete physical reason for the restriction to specific, well-defined transformations. Accordingly Einstein continued in his letter to Lorentz:

> Thus, in a word; *By postulating the conservation law, one arrives at a highly determined choice of the reference system and the admissible substitutions*. Only now, after this ugly dark spot seems to have been eliminated, does the theory give me pleasure.[115]

Einstein's argument presupposes, however, that the objects appearing in his equation for energy-momentum conservation eq. (51) behave themselves as tensors. This is true for the stress-energy tensor of matter, but not for the stress-energy expression for the gravitational field, as Einstein came to realize a few months later.[116] But if this quantity fails to behave as a tensor, the transformational properties of eq. (51) cannot be read off by inspection as Einstein claimed to be able to do in his letter to Lorentz.

Einstein immediately incorporated the argument for linear covariance found on 15 August 1913 and exposed to Lorentz a day later into his manuscript for a lecture he was invited to hold on 23 September 1913 in Vienna.[117] In § 6 of this lecture, enti-

---

114 "Ferner fand ich gestern zu meiner grossen Freude, dass die gegenüber der Gravitationstheorie in meinem letzten Briefe, sowie in der Arbeit geäusserten Bedenken nicht angezeigt sind. Die Sache scheint sich mir folgendermassen zu lösen. Ausdruck des Energieprinzips für Materie & Gravitationsfeld zusammen ist eine Gleichung von der Form (19) d.h. von der Form $\sum \frac{\partial F_{\mu\nu}}{\partial x_\nu} = 0$; von dieser Voraussetzung ausgehend stellte ich die Gleichungen (18) [i.e. the field equations] auf. Nun zeigt aber eine Betrachtung der allgemeinen Differenzialoperatoren des absoluten Differenzialkalküls, dass eine so gebaute Gleichung niemals absolut kovariant ist. Indem wir also die Existenz einer solchen Gleichung postulierten, spezialisierten wir stillschweigend die Wahl des Bezugssystems. Wir beschränkten uns auf den Gebrauch solcher Bezugssysteme, inbezug auf welche der Erhaltungssatz des Impulses und der Energie in dieser Form gilt. Es zeigt si[ch], dass bei der Bevorzugung solcher Bezugssysteme nur mehr allgemeine lineare Transformationen als allein berechtigt übrig bleiben." Einstein to Hendrik A. Lorentz, 16 August 1913, (CPAE 5, Doc. 470).

115 "Also kurz gesagt: *Durch die Postulierung des Erhaltungssatzes gelangt man zu einer in hohem Masse bestimmten Wahl des Bezugssystems und der zuzulassenden Substitutionen*. Erst jetzt macht mir die Theorie Vergnügen, nachdem dieser hässliche dunkle Fleck beseitigt zu sein scheint."

116 Einstein found the fallacy of this argument in the period between ca. 20 January 1914 and ca. 10 March 1914, see Einstein to Heinrich Zangger, ca. 20 January 1914, (CPAE 5, Doc. 507) and Einstein to Paul Ehrenfest, before 10 March 1914, (CPAE 5, Doc. 512), as well as Einstein to Heinrich Zangger, ca. 10 March 1914, (CPAE 5, Doc. 513). The last two letters mention, for the first time after a long period of intermission (to which the first letter to Zangger evidently still belongs), progress in the work on the gravitation problem related to the covariance properties of the *Entwurf* equation.

tled “Bemerkungen über die mathematische Methode,” Einstein summarized some of the mathematical properties of the *Entwurf* theory. In the previous paragraph he had dealt with equations describing the effect of the gravitational field on other physical processes, and in particular with energy-momentum conservation, in the next he was going to develop the gravitational field equation. Towards the end of § 6, he took the occasion to comment on the remarkable asymmetry between the transformational behavior of these two types of equations (Einstein 1913, 1257). He explained that the arguments in favor of a generalized principle of relativity, on the one hand, and his argument in favor of its restriction, on the other, are located on different levels, one on the general level of the spacetime structure, the other on the level of concrete physical requirements. Whereas the first level remains the relevant one for all equations dealing with the effects of the gravitational field on other physical processes, the second level becomes relevant only for the gravitational field equation itself. The introduction of a specialization of the reference system only “after the fact” may well have motivated him to continue his search for an explanation that makes the restriction of the generalized relativity principle understandable also on the level of the spacetime structure. Furthermore, the asymmetry between the transformational behavior of the gravitational field equations and that of all other equations of physics hinted at an explanation that is related to the mathematical nature of the field equations. As we shall now see, Einstein eventually found such an explanation in the hole argument.

7.6 The Construction of the Hole Argument[118]

The use of the conservation principle to justify the limited covariance of the *Entwurf* field equation was a relief but contributed, in effect, little to *understanding* the restriction of the generalized principle of relativity. In a first attempt to come to terms with this unsatisfactory situation, Einstein reformulated the Machian heuristics that had originally suggested the introduction of this principle. This attempt is visible already in his earliest statements concerning the conservation argument described above, but is most clearly expressed in a letter Einstein wrote to Ernst Mach in the second half of December 1913:

> It seems to me absurd to ascribe physical properties to “space.” The totality of masses produces the $g_{\mu\nu}$ field (gravitational field), which in turn governs the course of all processes, including the propagation of light rays and the behavior of measuring rods and clocks. First of all, everything that happens is referred to four completely arbitrary space-time variables. If the principles of momentum and energy conservation are to be satisfied, these variables must then be specialized in such a way that only (completely) linear substitutions shall lead from one justified reference system to another. The reference sys-

---

117 Einstein completed the manuscript for this lecture about a month earlier, see (CPAE 4, note 17) and “What Did Einstein Know ...” note 19 (in vol. 2 of this series).

118 See also “What Did Einstein Know ...” (in vol. 2 of this series) and further references cited there.

> tem is, so to speak, tailored to the existing world with the help of the energy principle, and loses its nebulous aprioristic existence.[119]

In this reformulated argument against absolute space, the role of the cosmic masses and their relations in constituting space is now taken over by the conservation principle which is claimed to provide the physical justification of preferred reference frames. But Einstein's revised Machian heuristics nevertheless failed to completely ban the plausibility of a generalization of the relativity principle. This is evident from the efforts by Einstein and his friend Michele Besso to find a deeper connection between the limited transformation properties of the *Entwurf* field equation and the structure of spacetime.

These efforts are documented, in particular, by a manuscript in the hand of Michele Besso.[120] In a group of pages, the first of which carries the dateline 28 August 1913, Besso listed a number of problems which he had probably encountered while working jointly with Einstein on the problem of Mercury's perihelion shift in the context of the *Entwurf* theory.[121] These problems were evidently not intended as a program for the further development of the *Entwurf* theory but rather constituted the results of reflections on what had been achieved so far. In a note found in the part of the manuscript belonging to a later period, Besso modestly characterized himself as being merely an "*orrechiante*," that is, as an amateur who had the privilege of listening to a great master. But some of Besso's observations turned out to be most consequential, introducing a possibly naive but fresh perspective. It seems in fact that he went together with Einstein through his list of problems, some of them directly formulated as questions, and that they discussed them one by one; at some later point Besso then entered Einstein's responses.[122] Three of Besso's problems are relevant for the present discussion, the first concerning the issue of rotation, the second regarding the logical status of the restriction implied by the conservation principle, and the third concerning what later was to become the hole argument, Einstein's central argument to defend the restricted covariance of the *Entwurf* theory.

---

119 "Für mich ist es absurd, dem "Raum" physikalische Eigenschaften zuzuschreiben. Die Gesamtheit der Massen erzeugt ein $g_{\mu\nu}$-Feld (Gravitationsfeld), das seinerseits den Ablauf aller Vorgänge, auch die Ausbreitung der Lichtstrahlen und das Verhalten der Massstäbe und Uhren regiert. Das Geschehen wird zunächst auf vier *ganz willkürliche* raum-zeitliche Variable bezogen. Diese müssen dann, wenn den Erhaltungssätzen des Impulses und der Energie Genüge geleistet werden soll, derart spezialisiert werden, dass nur (ganz) *lineare* Substitutionen von einem berechtigten Bezugssystem zu einem anderen führen. Das Bezugssystem ist der bestehenden Welt mit Hilfe des Energiesatzes sozusagen angemessen und verliert seine nebulose apriorische Existenz." Einstein to Ernst Mach, second half of December 1913, (CPAE 5, Doc. 495).

120 For more detailed discussion of both the contents and the dating of this document, see "What Did Einstein Know ..." (in vol. 2 of this series).

121 For a facsimile reproduction of these pages, see (Renn 2005a, 126–130).

122 This pattern is made evident also by the arrangement of problems and answers on these pages, the answers being often written in a slightly different hand than the questions and squeezed in between the lines or at the margin.

In his notes on the first problem, Besso summarized the failure of Einstein's hopes to fully implement his Machian heuristics within the *Entwurf* theory. Besso noted that it was impossible to conceive of rotation as equivalent to a state of rest in a gravitational field that is a solution of the *Entwurf* field equations.[123] This represented a challenge either for the *Entwurf* theory or the generalized principle of relativity. In the consolidation period of the *Entwurf* theory the decision was made in favor of the theory. That was bound to change only in the context of Einstein's renewed exploratory phase in the fall of 1915. In his notes Besso studied the question of whether the failure of the *Entwurf* theory to interpret rotation as a state of rest can possibly be explained by a failure of the conservation principle in a rotating system. If that were so, he would have succeeded in establishing the desired physical connection between the problem of implementing the generalized relativity principle and the restriction of the admissible coordinate systems required by the conservation principle. The remainder of the text, probably going back to Einstein's intervention, shows that this attempt of explaining the problem with rotation in terms of the conservation principle does not work.

In a second passage, Besso posed a more general question concerning the role of energy-momentum conservation for the selection of admissible coordinate systems:

> Is every system that satisfies the conservation laws a justified system?[124]

If the conservation principle is really the explanation for the restriction on the choice of coordinate systems, it should not only be a necessary but also a sufficient condition for this choice. Besso therefore wondered whether *any* coordinate system satisfying the constraints imposed by the conservation laws is compatible also with the covariance of the field equations. Since there is no note which may be traced back to a reaction by Einstein on this issue, it seems that he did, at first, not seriously consider Besso's suggestion. In fact, given Einstein's belief at that point that the conservation equation (51) is covariant under linear transformations only, the question may have held little interest for him since it offered no promise of generalizing the covariance properties of the *Entwurf* field equations beyond linearity. Eventually, however, when Einstein turned to the exploration of a mathematical strategy for the *Entwurf* theory, he did realize, as we shall see, the significance of Besso's question.

After writing down his second question, Besso sketched an idea of how the failure of realizing general covariance on the level of the gravitational field equation might be explained, namely as a problem of the uniqueness of its solutions. The text of his third problem reads:

> The requirement of [general] covariance of the gravitational equations under arbitrary transformations cannot be imposed: if all matter [is given] were contained in one part of space and for this part of space a coordinate system [is given], then outside of it the coordinate system could still [essentially] except for boundary conditions be chosen arbi-

123 See "What Did Einstein Know ..." sec. 3 (in vol. 2 of this series).

124 "Ist jedes System, welches den Erhaltungssätzen genügt, ein berechtigtes System?"

> trarily, [through which the g arbitrarily] so that a unique determinability of the g ’s cannot be obtained.
>
> It is, however, not necessary that the g themselves are determined uniquely, only the observable phenomena in the gravitation space, e.g., the motion of a material point, must be.[125]

In this passage Besso imagines a central mass to be surrounded by empty space and wonders whether the solution for the metric tensor is, in this case, determined uniquely for the empty region. His mental model appears to be the inverse of Einstein’s famous hole argument where matter may be anywhere outside an empty hole for which the problem of the ambiguity of solutions then supposedly arises. In Besso’s argument the ambiguity of solutions is conceived as being due to the arbitrary choice of the coordinate system in the empty region, giving rise to arbitrary coordinate expressions for the metric tensor (which have to satisfy, however, the boundary conditions). Apart from the inversion of hole and matter, Besso’s “proto-hole argument” thus corresponds to the primitive version of the hole argument that was traditionally ascribed to Einstein, charging him with the naivety of being unaware that different coordinate representations of the metric tensor do not correspond to different solutions of the field equations. It is therefore remarkable that even Besso immediately realized the flaw of this naive version since he added, in the second paragraph of the above text, that only observable phenomena, such as the motion of a particle, should be determined uniquely.

How did Besso’s idea emerge and how was it transformed into the hole-argument familiar from Einstein’s later publications? As to the first question, it seems plausible that Besso related, in the context of his reflections on the *Entwurf* theory, the problem of the restriction of general covariance to other problems that had arisen for this theory, in particular in the course of his joint research with Einstein. One such problem was rotation, as we have just seen. Another problem was the perihelion shift of Mercury, the central subject of a paper Besso planned to write.

In 1913 Besso had in fact encountered the problem of uniqueness when he worked on the perihelion problem in the context of the *Entwurf* theory. In fact, a note in the Einstein-Besso manuscript explicitly refers to the question of uniqueness in connection with the ansatz used for solving the *Entwurf* field equation by an approximation procedure. For the first step of that iterative procedure Einstein and Besso had used a metric with only one variable component, the same spatially flat metric (25)

125 “Die Anforderung der [allgemeinen] Covarianz der Gravitationsgleichungen für beliebige Transformationen kann nicht aufgestellt werden: wenn in einem Teile des Raumes alle Materie [gegeben ist] enthalten wäre und für diesen Teil ein Coordinatensystem, so könnte doch ausserhalb desselben das Coordinatensystem noch, [im wesentlichen] abgesehen von den Grenzbedingungen, beliebig gewählt werden, [wodurch die g beliebig eine] so dass eine eindeutige Bestimmbarkeit der g s nicht eintreten könne.
Es ist nun allerdings nicht nötig, dass die g selbst eindeutig bestimmt sind, sondern nur die im Gravitationsraum beobachtbaren Erscheinungen, z.B. die Bewegung des materiellen Punktes, müssen es sein.”

that is also crucial for obtaining the Newtonian limit of the *Entwurf* theory. Besso then wondered whether that choice of a metric was sufficiently general for recovering all possible solutions of the field equation:

> Is the static gravitational field in § 1 $g_{\mu\nu} = 1$, 1 to 3 $g_{44} = f(x, y, z)$ a particular solution? Or is it the general solution expressed in particular coordinates?[126]

It may have been this question arising in the context of the perihelion calculation that suggested to Besso that covariant field equations suffer, in general, from a problem of uniqueness. In fact, the physical model of Besso's proto-hole argument is strikingly similar to that of the perihelion problem, a central mass surrounded by empty space. And when Besso reminded himself that it was not the expression for the metric tensor that mattered but physically observable phenomena, he chose the motion of a material particle, such as that of Mercury around the sun, as an example.

How did Einstein react to Besso's consideration of the proto-hole argument and how did the definitive version of the hole argument emerge? Einstein's reaction is, it seems, preserved in a text written below Besso's note quoted earlier. It starts, just as Einstein's earlier remark concerning rotation and energy conservation, with a characteristic "Of no use" ("Nützt nichts"):

> Of no use, since with [the] a solution a motion is also fully given. If in coordinate system 1, there is a solution K_1, then this same construct is also a solution in 2, K_2; K_2, however, also a solution in 1.[127]

Remarkably, Einstein did not simply agree with Besso's conclusion that the ambiguity of the coordinate representation of the metric tensor was of no physical consequence. He apparently found Besso's idea to justify the lack of general covariance of the *Entwurf* field equations on the basis of a uniqueness argument intriguing and effectively reinterpreted it as an argument about the nature of space and time, and, in particular, about the role of coordinate systems in identifying points in space and time. In fact, a solution of the field equation in a particular coordinate system, expressed in terms of functions representing the components of the metric tensor, can be transformed to another coordinate system, producing a different set of functions representing the same solution. But if, as Einstein's argument suggests, this set of functions "this same construct" ("dieses selbe Gebilde") can somehow be related to the original coordinate system, it there represents a different metric which, however, solves the same field equation, provided that the right-hand side of these equations remains unchanged by the coordinate transformation, which is the case for an empty region where the stress-energy tensor of matter vanishes. To avoid the issue of additional boundary conditions, it turned out to be convenient for Einstein to reverse the

126 "Ist das stat Schwerefeld des § 1 $g_{\mu\nu} = 1$, 1 bis 3, $g_{44} = f(x, y, z)$ ein spezielles? oder ist es das allgemeine, auf spec. Coordinaten zurückgeführtes" (CPAE 4, Doc. 14, [p. 16]).

127 "Nützt nichts, denn durch eine Lösung ist auch eine Bewegung voll gegeben. Ist im Koordinatensystem 1 eine Lösung K_1, so ist dieses selbe Gebilde auch eine Lösung in 2, K_2; K_2 aber eine Lösung in 1." For a facsimile of this passage, see Fig. 2 on p. 300 of "What Did Einstein Know ..." (in vol. 2 of this series) and (Renn 2005a, 128).

physical model proposed by Besso and consider, instead of a void with a lump of mass, matter with a hole in it—the well-known configuration of the hole argument.[128]

Einstein's interpretation of a solution initially given in one coordinate system as referring to another coordinate system implicitly presupposes that coordinate systems have their own physical reality and allow to identify points in spacetime. The crucial but hidden point of this reinterpretation of Besso's proto-hole argument is therefore a reification of coordinate systems, which are conceived as part of the physical set-up constituting a solution and not only as a mathematical device for describing it. Only the later refutation of the hole argument made it eventually clear that it is not "motions" in the sense used here which constitute physically real events but rather spacetime coincidences for which a coordinate-independent description can be given.[129]

Our reconstruction suggests that the hole argument was, in spite of its philosophical appeal, not rooted in a metaphysical prejudice concerning the nature of space and time or the role of coordinate systems, preventing Einstein from accepting generally-covariant field equations. On the contrary, it was the necessity of justifying a non-generally-covariant field equation that led to the construction of this argument, triggering a peculiar interpretation of the physical significance of coordinate systems, an interpretation moreover that largely remained implicit in the initial formulation of the argument. The hole argument was just the kind of argument Einstein had been after in his earlier attempts to justify the failure of general covariance: a mathematical argument related to the structure of space and time. It was this peculiar perspective, shaped by the context of the consolidation period of the *Entwurf* theory, that probably led him to take Besso's naive point seriously and search for a physically significant interpretation of a mathematically trivial property, the coordinate dependence of expressions for the metric tensor. It is hardly surprising that to formulate such an interpretation, Einstein relied on the conceptual resources of classical physics, implicitly defining what a motion is in terms of the relation between a particle and a coordinate system. As a result, he found a way of relating the formalism of absolute differential calculus to a physical interpretation of coordinate systems that allowed him to justify the restricted covariance of the *Entwurf* field equations. In short, the necessity of interpreting a complex mathematical formalism under a peculiar perspective was crucial for the emergence of the hole argument. Only when Einstein eventually succeeded in formulating physically acceptable, generally-covariant field equations did he abandon this argument and revise the physical interpretation of coordinate systems as well as of space and time associated with it.[130] The deep conceptual insight into the crucial role of spacetime coincidences was thus no presuppo-

---

128 For detailed discussion, see sec. 4 of "What Did Einstein Know ..." (in vol. 2 of this series).

129 For selected references to the extensive literature on the hole argument, see "What Did Einstein Know ..." note 95 (in vol. 2 of this series).

130 For discussion of Einstein's later retraction of the hole argument, see "What Did Einstein Know ..." sec. 4 (in vol. 2 of this series) and (Janssen 2005, 73–74)

sition of general relativity but merely a consequence of its establishment—effectively implying a refutation of the hole argument.

With the advent of the hole argument, the conservation principle lost its role as the primary physical reason for the restriction of general covariance of the *Entwurf* field equations; the restriction to linear transformations now merely appeared as a concrete result in harmony with a more general insight. Einstein reinterpreted his first argument defending the restricted covariance of the *Entwurf* field equation accordingly as a specific physical complement to what he saw as a general "logical" principle.[131] That he considered the hole argument not merely as an addition but as the solution of a puzzle left unresolved by the earlier physical argument is confirmed by a letter he wrote in the beginning of November 1913 to Ehrenfest:

> The gravitation affair has been clarified to my *complete satisfaction* (namely the circumstance that the equations of the gr. field are covariant only with respect to *linear* transformations. For it can be proved that *generally covariant* equations that determine the field *completely* from the matter tensor cannot exist at all. Can there be anything more beautiful than this, that the necessary specialization follows from the conservation laws?[132]

7.7 The Second Phase of the Consolidation Period of the Entwurf Theory: A Mathematical Strategy for the Entwurf Theory

The insights Einstein had acquired pursuing the mathematical strategy in the Zurich Notebook continued to set standards for his further elaboration of the *Entwurf* theory developed along the lines of the physical strategy. In particular, the procedure at the core of the mathematical strategy by which non-generally-covariant field equations could be extracted from a generally-covariant object remained plausible. Even if generally-covariant field equations were excluded for a satisfactory relativistic theory of gravitation, the physical and the mathematical strategies should converge because only in this way was it possible to fully clarify and stabilize the relation between the physical and the mathematical knowledge expressed in the theory. Even if Einstein had good reasons for restricting the generalized principle of relativity, it still made sense for him to search for a derivation of the field equations of the *Entwurf* theory along the mathematical strategy, albeit now with the aim of confirming what had

---

131 In a later paper he formulated with regard to these two arguments: "But there are two weighty arguments that justify this step [i.e. the restriction of general covariance], one of them of logical, the other one of empirical provenance" ("Es gibt aber zwei gewichtige Argumente, welche diesen Schritt rechtfertigen, von denen das eine logischen, das andere empirischen Ursprungs ist", CPAE 4, Doc. 25, [178]).

132 "Die Gravitationsaffäre hat sich zu meiner *vollen Befriedigung* aufgeklärt (der Umstand nämlich, dass die Gleichungen des Gr. Feldes nur *linearen* Transformationen gegenüber kovariant sind. Es lässt sich nämlich beweisen, dass *allgemein kovariante* Gleichungen, die das Feld aus dem materiellen Tensor *vollständig* bestimmen, überhaupt nicht existieren können. Was kann es schöneres geben, als dies, dass jene nötige Spezialisierung aus den Erhaltungssätzen fliesst?" Einstein to Paul Ehrenfest, before 7 November 1913, (CPAE 5, Doc. 481).

already been found through the physical strategy. This approach is characteristic for what we call the "second phase" of the consolidation of the *Entwurf* theory.

It was probably in pursuing a mathematical strategy for the *Entwurf* theory that, in early 1914, Einstein discovered a flaw in his conservation argument for linear covariance. It turned out that the quantity representing the stress-energy of the gravitational field is not a tensor. Einstein thus realized that, while the conservation principle still requires a restriction on the admissible coordinate systems, this restriction was not as stringent as it had seemed before. The issue of the covariance properties of the *Entwurf* theory was therefore reopened since the hole argument only excluded general covariance but did not by itself prescribe a specific covariance group. It remained to be clarified, in particular, in which sense the transformational properties of the *Entwurf* field equation were restricted by the conservation principle, which now appeared as implying a "weak" restriction only taking full advantage perhaps of the leeway left by the hole argument.

A suggestion by the Zurich mathematician Paul Bernays made it possible for Einstein and Grossmann to return in early 1914 to the "bold" approach, once again exploring the transformational properties of the *Entwurf* field equation by direct calculation. Bernays suggested to derive the *Entwurf* field equation from a variational principle in order to be able to focus attention on a single scalar quantity, the Lagrangian, rather than on the complex tensorial objects constituting the field equation itself. Einstein and Grossmann succeeded indeed in finding a Lagrangian from which the *Entwurf* field equations could be derived. They found that this Lagrangian is invariant under transformations between coordinate systems specified solely by the requirement of energy-momentum conservation. The necessary restriction of covariance following from the conservation principle thus turned out to be also a sufficient one, just as Michele Besso had envisaged. This result seemed to be in perfect agreement also with the hole argument since the four additional equations resulting from the conservation principle were apparently just enough to remove the ambiguity in the metric field on which this argument turns.

With these results, attained by March 1914, the "defensive" and the "bold" approaches had converged and, once again, a sense of closure in the development of the *Entwurf* theory was reached, this time on a higher level than half a year earlier and more durable: it would last until October 1915. Einstein was convinced that he had obtained an optimal realization of the generalized relativity principle and that he had understood the profound reasons for the impossibility of general covariance. He even came to believe that the restriction of covariance imposed by the conservation principle in fact does not imply a restriction of the possible solutions to the field equation but merely a restriction of the possible coordinate systems in which these solutions can be expressed. Consequently, Einstein also became convinced that the *Entwurf* theory fully realized the equivalence principle and other heuristic ideas, such as a description of Minkowski spacetime in a rotating frame of reference as a special case of the gravitational field, in spite of the difficulties at the level of explicit calculations.[133]

The success of this exploration of the *Entwurf* theory along the lines of the mathematical strategy encouraged Einstein to undertake a new derivation of its field equations; he completed this derivation by the fall of 1914. By February 1914, he had abandoned his earlier conviction that the *Entwurf* field equation had no relation to the absolute differential calculus essential to the mathematical strategy. When Einstein took up the project of deriving the *Entwurf* field equation along the mathematical strategy, however, he did not start from the generally-covariant objects suggested by the original mathematical strategy as he had done in the Zurich Notebook. He rather generalized the variational derivation, developed together with Grossmann, into a mathematical formalism applicable not only to the *Entwurf* field equations but to other candidate field equations. He then searched for a mathematical reason to justify choosing the Lagrangian corresponding to the *Entwurf* equation and erroneously convinced himself that he had actually found such reasons.

7.8 A Prelude: The First Reawakening of the Mathematical Strategy

Einstein's failure to reach general covariance had been a target of criticism by his colleagues.[134] In January 1914 he wrote a paper in reply to such criticism (Einstein 1914b). Apart from presenting his arguments in favor of a restricted covariance of the *Entwurf* field equations such as the hole argument, he had to admit that the relation of this field equations to the generally-covariant objects of the absolute differential calculus was still an open problem. As a consequence, the relation between physical and mathematical strategies, which should have been just two different pathways to the same result, also remained unclear.

Defending the restricted covariance of the *Entwurf* theory, Einstein had to acknowledge that there are profound reasons why generally-covariant equations should exist which correspond to the *Entwurf* field equation (Einstein 1914b, 177–178). He argued that there must be, in modern terms, a coordinate-free representation of any meaningful mathematical relation between physical magnitudes. Ideally, the *Entwurf* equation should be derived from such a representation by a suitable specialization of the coordinate system. This would correspond to its derivation along the lines of the mathematical strategy. But as if to excuse himself for the failure to realize such a derivation, Einstein claimed that the hole argument and the argument from energy momentum conservation suggested that it would not be worthwhile to search for the generally-covariant counterpart of the *Entwurf* equation.[135]

In spite of this excuse Einstein embarked, at about the same time, on precisely such a search, albeit for another gravitation theory with restricted covariance properties serving as a toy model. On 19 February 1914 he submitted a joint paper with

---

133 See "What Did Einstein Know ..." sec. 3 (in vol. 2 of this series).

134 See, e.g., (Abraham 1914, 25).

135 (CPAE 4, Doc. 25, [179]). See "Untying the Knot ...," (in vol. 2 of this series) note 57, for the relevant passage. See (Norton 1992a) for a historical discussion.

Adriaan Fokker on a generally-covariant reformulation of Nordström's special relativistic theory of gravitation (Einstein and Fokker 1914). They demonstrated that the field equation of this theory, in its original version only Lorentz covariant, can in fact be obtained from a generally-covariant equation. Just as in the mathematical strategy employed in the Zurich Notebook, a generally-covariant expression derived from the Riemann tensor served as the starting point of their approach, from which a suitable left-hand side of the gravitational field equation was then obtained by imposing additional conditions on the metric tensor. In the case of the Nordström theory, the additional condition amounted to the requirement of the constancy of the speed of light. This additional condition in turn led to a restriction on the admissible coordinate systems, in this case to those systems which are adapted to the principle of the constancy of the velocity of light (Einstein and Fokker 1914, 326).

As a consequence of the successful reformulation of Nordström's theory in generally-covariant terms, it was only natural to search for an analogous reformulation also of the *Entwurf* theory, in spite of the skepticism which Einstein had expressed in his earlier paper. That such a search made sense was precisely the conclusion which Einstein and Fokker drew at the end of their joint paper:

> Finally, the role that the Riemann-Christoffel differential tensor plays in the present investigation suggests that this tensor may also open the way for a derivation of the Einstein-Grossmann gravitation equations that is independent of physical assumptions. The proof of the existence or nonexistence of such a connection would represent an important theoretical advance.[136]

In a footnote to the above passage, they added:

> The argument in support of the nonexistence of such a connection, presented in §4, p. 36 of "Entwurfs einer verallgemeinerten Relativitätstheorie" ["Outline of a Generalized Theory of Relativity"], did not withstand closer scrutiny.[137]

On the cited page of the *Entwurf* paper, Einstein and Grossmann had simply claimed that, in the case of field equations with restricted covariance, it was understandable that no relation to generally-covariant tensors could be established (Einstein and Grossmann 1913, 36). But in view of Einstein's realization that a connection with a generally-covariant formulation must exist for any physically meaningful theory, the failure to discover such a connection could no longer be defended in this simple way.

In their paper, Einstein and Fokker used a terminology for the relation between generally-covariant equations and field equations with restricted covariance that

136 "Endlich legt die Rolle, welche bei der vorliegenden Untersuchung der Riemann-Christoffelsche Differentialtensor spielt, den Gedanken nahe, daß er auch für eine von physikalischen Annahmen unabhängige Ableitung der Einstein-Großmannschen Gravitationsgleichungen einen Weg öffnen würde. Der Beweis der Existenz oder Nichtexistenz eines derartigen Zusammenhanges würde einen wichtigen theoretischen Fortschritt bedeuten." (Einstein and Fokker 1914, 328).

137 "Die in §4, p. 36, des "Entwurfs einer verallgemeinerten Relativitätstheorie" angegebene Begründung für die Nichtexistenz eines derartigen Zusammenhanges hält einer genaueren Überlegung nicht stand."

would soon become standard in the further analysis of the covariance properties of the *Entwurf* theory. They spoke, in particular, of "preferred" ("*bevorzugt*") coordinate systems, "adapted" ("*angepasste*") to a certain physical situation.[138] Nordström's theory and the terminology developed for its treatment helped to further pursue the questions which had to be answered for a derivation of the *Entwurf* theory along the lines of the mathematical strategy to succeed: What were the "preferred" coordinate systems of the *Entwurf* theory? And what was the physical condition to which these coordinate systems are "adapted"? Although Einstein must have believed that he had answers to these questions, given his argument in favor of a restriction to linear transformations from energy-momentum conservation, it remained open how these answers could assist him in relating the *Entwurf* theory to its unknown generally-covariant counterpart. It was the experience gathered with Nordström's theory that eventually helped him to make progress in this regard—by challenging the answers that had seemingly settled the fate of the generalized relativity principle in the *Entwurf* theory.

7.9 A First Consequence of the Return to the Mathematical Strategy

In early March Einstein wrote to his friends about a breakthrough in his work on the *Entwurf* theory.[139] By this time he had not only recognized the fallacy of his argument for restricted covariance from energy-momentum conservation but had also investigated, jointly with Marcel Grossmann, the covariance properties of the theory in a new way. This new analysis, contained in a joint paper published on 29 May 1914 (Einstein and Grossmann 1914), was based on the use of variational techniques which allowed them to pursue the bold approach of exploring covariance properties by direct calculation.

Einstein's breakthrough was prepared by his reflection on the relation between non-covariant and covariant formulations of a theory, substantiated by his analysis of Nordström's theory. In light of these considerations, the *Entwurf* theory appears as a specialization of a generally-covariant theory to coordinate systems which are adapted to a certain physical condition. In the case of the *Entwurf* theory, this physical condition was the validity of energy-momentum conservation in the sense of eq. (51).

If generally-covariant field equations are expressed in coordinates adapted to this condition, they should take on the form of the *Entwurf* field equation eq. (52).

In analogy to the treatment of the Nordström theory, eq. (51) should be considered as a condition on the metric tensor $g_{\mu\nu}$, providing the necessary coordinate restriction. But in Einstein's original version of the argument for restricted covariance from energy-momentum conservation, this equation does not so much *provide* a con-

138 See (Einstein and Fokker 1914, 326).

139 See Einstein to Paul Ehrenfest, before 10 March 1914, (CPAE 5, Doc. 512) and Einstein to Heinrich Zangger, ca. 10 March 1914, (CPAE 5, Doc. 513).

dition on the metric tensor, but actually *presupposes* one. In fact, the argument that eq. (51) is only covariant under linear transformations only works, as Einstein was aware, if it is assumed that $\boldsymbol{t}^{\mu\nu}$ has the same transformational behavior as $\boldsymbol{T}^{\mu\nu}$, i.e. if it is a generally-covariant object. However, $\boldsymbol{t}^{\mu\nu}$ is a coordinate-dependent expression.[140] While the assumption that it is generally covariant may have at first appeared plausible to Einstein in the light of his conviction that gravitational and other forms of energy should behave in the same way as sources of the gravitational field, this assumption becomes much less plausible once eq. (51) is seen as imposing a coordinate restriction on generally-covariant equations. That perspective requires in fact a much closer examination of its ingredients, since it is now the content rather than the form of the equation that matters. Apart from checking more closely the character of $\boldsymbol{t}^{\mu\nu}$, Einstein's earlier experience with what we have called the conservation compatibility check in the sense of eq. (XLIII) suggested expressing $\boldsymbol{T}^{\mu\nu} + \boldsymbol{t}^{\mu\nu}$ in this equation by means of the field equations so that eq. (51) becomes a condition merely in terms of the metric tensor and its derivatives. One thus obtains eq. (54) as a condition for the class of admissible coordinate systems.

This equation played a central role in Einstein's new approach to the problem of the covariance properties of the *Entwurf* field equation. It first appeared in Einstein's and Grossmann's 1914 paper[141] and expresses in fact a physically motivated coordinate restriction in a sense that was quite familiar to him from his experiences along the mathematical strategy in the Zurich Notebook. In distinction from the original mathematical strategy, however, the generally-covariant equation from which the *Entwurf* field equation should be derivable by means of this coordinate restriction was unknown. But finding this generally-covariant equation may have been precisely Einstein's point in formulating eq. (54). In summary, a reconsideration from the perspective of the mathematical strategy of the argument for restricted covariance based on energy-momentum conservation could have led Einstein both to see the fallacy of his original argument and to cast it into a new form.

This reconstruction is supported by the timing of the transformation of the original argument for a linear covariance of the *Entwurf* equations into an argument about a coordinate restriction in the sense of the Zurich Notebook. In the manuscript of a popular exposition on his theory,[142] which Einstein completed by the end of January 1914,[143] he still included the argument in its original form. When he submitted the paper for publication by March, 21, 1914, that is, before he left Zurich, the passage arguing for the linear covariance of the *Entwurf* field equation was cancelled. By the beginning of March, Einstein had already achieved a breakthrough along the varia-

---

140 There are a number of arguments by which Einstein could have seen his fallacy: 1) There are no generally-covariant tensors involving only the metric and its first order derivatives. 2) In a suitably chosen coordinate system, the stress-energy complex of the gravitational field $t_{\mu\nu}$ can be made to vanish.

141 See (Einstein and Grossmann 1914, 218).

142 The published version is (Einstein 1914c). For references to and transcriptions of the manuscript version, see the annotations in (CPAE 4, 621–622).

143 See Einstein to Heinrich Zangger, ca. 20 January 1914, (CPAE 5, Doc. 507).

tional approach, as we know from his correspondence. The paper by Einstein and Fokker on Nordström's theory, on the other hand, in which, as we have also seen, the application of the mathematically strategy to the *Entwurf* theory is formulated as a program, was submitted on 19 February 1914. In other words, Einstein must have reformulated his argument based on the conservation principle at some point between the end of January and the beginning of March 1914, at the time or shortly after he was working on the application of the mathematical strategy to Nordström's theory.

With the discovery of the fallacy in the original argument, the question of the covariance properties of the *Entwurf* field equation was open again. While it was obvious that eq. (54) imposes a necessary condition on the coordinates systems "adapted" to this theory, it remained to be clarified whether this condition is also a sufficient one and how it related the *Entwurf* field equation to its generally-covariant counterpart. Einstein had thus arrived at a point where it made sense for him to take up the second point raised in Besso's notes:

> Is *every* system *that satisfies the conservation laws* a *justified* system?[144]

In the context of Einstein's reconsideration of the *Entwurf* field equation from the perspective of the mathematical strategy, the relation between conservation laws and "justified" coordinate systems must have assumed a new significance. Exploring whether or not this field equation actually retained its form under transformations between the "preferred" coordinate systems characterized by eq. (54) now became a pressing task. Unfortunately, the absolute differential calculus offered little help in addressing this task.

7.10 A New Turn for the Mathematical Strategy: Variational Calculus

When Einstein took up the mathematical strategy once again and adapted it to the *Entwurf* theory, he faced difficulties achieving concrete results along this strategy, and must have searched out mathematical advice. It is unclear at exactly which point Grossmann re-entered the story. Perhaps he was already instrumental in recognizing the fallacy of Einstein's original argument for restricted covariance from energy-momentum conservation. Perhaps he entered the picture only when Einstein needed help in exploring the consequences of the new coordinate restriction eq. (54). But Grossmann was, it seems, as little successful as Einstein in establishing relations between the *Entwurf* theory and absolute differential calculus. At some point they both turned to another Zurich mathematician colleague, Paul Bernays, for help.[145] Bernays advised Einstein and Grossmann to bring the field equation of the *Entwurf* theory into the form of a variational principle.[146]

The reformulation of the *Entwurf* theory in terms of a variational principle did not, however, provide any clue concerning the relation of this theory to absolute dif-

144 "Ist *jedes* System, *welches den Erhaltungssätzen genügt*, ein *berechtigtes* System?" For a facsimile of this passage, see Fig. 2 on p. 300 of "What Did Einstein Know ..." (vol. 2 of this series).

ferential calculus. The latter would have suggested, as it later did to Hilbert, to take the Ricci scalar as a starting point for such a reformulation. But, by analyzing the relation of this scalar to the B_μ of their coordinate restriction, Einstein and Grossmann convinced themselves that the B_μ do not form a generally-covariant vector and that the *Entwurf* field equation has nothing to do with the invariant Ricci scalar.[147] Nevertheless, a variational reformulation of the *Entwurf* had, for Einstein and Grossmann, one chief advantage: instead of having to study the covariance properties of a complex tensorial field equation, they could instead explore the invariance group of a single scalar object, the action integral (cf. eq. (LXIV)). Much later Einstein still considered the simplification due to the introduction of the more familiar scalar quantities the main advantage of the variational calculus.[148]

In early 1914, the suggestion to make use of the variational calculus brought Einstein and Grossmann back to the initial bold approach of exploring by direct calculation the covariance properties of the *Entwurf* field equation. In pursuing this approach they could rely on their experience from the Zurich Notebook where they had attempted to study the covariance properties of objects found along the physical strategy or of coordinate restrictions by means of infinitesimal transformations.

However, one crucial presupposition of the new approach had to be established first, the expression for the action integral from which the *Entwurf* field equation could be derived by means of the variational formalism. In their 1914 paper Einstein and Grossmann only give the end result, without mentioning what had motivated them to introduce a particular Langrangian, other than its successful employment in deriving the field equations.[149] Probably they found the Lagrangian of the *Entwurf* theory by starting from an expression quadratic in the fields in analogy to classical and special-relativistic physics according to the default setting eq. (LXIII). With Einstein's default setting for the components of the gravitational field, the Lagrangian required for deriving the *Entwurf* field equation was found to be:

$$L = g^{\mu\nu} g_{\beta\mu,\alpha} g_{\alpha\nu,\beta} = g^{\mu\nu} \tilde{\Gamma}^{\alpha}_{\beta\mu} \tilde{\Gamma}^{\beta}_{\alpha\nu}. \tag{85}$$

145 Bernays later became known for his work in mathematical logic and set theory, was in Zurich from 1912 to 1919 after completing a mathematical doctoral thesis on the analytic theory of binary quadratic forms under the supervision of the mathematician E. Landau. Before coming to Zurich, Bernays had spent two years in Göttingen studying mathematics and physics chiefly with Hilbert, Landau, Weyl, Klein, Voigt and Born. Bernays was at the time concerned with an extension of the special theory of relativity.

146 In their paper, Einstein and Grossmann acknowledge the stimulation received from Bernays in a footnote (Einstein and Grossmann 1914, 218).

147 See (Einstein and Grossmann 1914, 225).

148 Einstein to Lorentz, 19 January 1916 and Einstein to T. De Donder, 23 July 1916 (CPAE 8, Docs. 184 and 240).

149 See (Einstein and Grossmann 1914, 219).

The *Entwurf* theory was thus solidified by connecting its field equation in yet another way with the established knowledge of classical and relativistic physics, in this case about the canonical form of a Lagrangian.

The identification of both the Langrangian for the *Entwurf* field equation and of a physically motivated coordinate restriction now gave Einstein and Grossmann a clear definition of their next goal, the establishment of a relation between the coordinate restriction and the transformational properties of this Lagrangian. Does the coordinate restriction eq. (54) resulting from the conservation principle indeed constitute not only a necessary but also a sufficient condition for the covariance of the *Entwurf* Lagrangian? In that case, by establishing a connection between conservation and covariance, Einstein would have achieved a result effectively preparing the later Noether theorem.[150] The close connection between conservation and covariance, first suggested by the ill-fated argument for a restriction of the *Entwurf* theory to linear transformations, became a heuristic guiding principle for Einstein's further exploration and a criterion that he expected a satisfactory theory to fulfill.

The means to answer his question concerning the covariance properties of the *Entwurf* Lagrangian was provided by the infinitesimal coordinate transformations explored earlier in the Zurich Notebook. With their help, Einstein and Grossmann succeeded in establishing a connection between the transformational properties of the action integral for the *Entwurf* Lagrangian and the physically motivated coordinate restriction eq. (54). Their bold approach had finally given them what they had failed to achieve with the help of the absolute differential calculus—a link between the physical and the mathematical strategies.

7.11 Looking Back on a Breakthrough: The General Relativity of the Entwurf *Theory*

With their proof of the covariance properties of the *Entwurf* field equations, Einstein and Grossmann had finally closed the crucial gap in their 1913 publication. But in Einstein's view, they had achieved much more. In the time between the completion of this proof by early March 1914 and the discovery of a problem with transformations to rotating frames of reference in September 1915, he was convinced that he had finally reached not only a generalization of the relativity principle but a truly general theory of relativity. In early March he wrote to Paul Ehrenfest that the proof of the existence of "most general transformations" leaving the field equations covariant demonstrated the validity of the principle of equivalence as well:

> The work on gravitation progresses, but at the cost of extraordinary efforts; gravitation is coy and unyielding! The equivalence principle is valid after all in the sense that there exist highly general transformations that transform the gravitational equations into themselves. What has been found is simple, but the search is hell![151]

In a similar vein he expressed himself in a contemporary letter to Heinrich Zangger:

---

150 See "Untying the Knot ..." (in this volume).

> I was toiling again on the gravitation theory to the point of exhaustion, but this time with unheard-of success. That is to say that I succeeded in proving that the gravit. equations hold for arbitrarily moving reference systems, and thus that the hypothesis of the equivalence of acceleration and the gravitational field is absolutely correct, in the widest sense. Now the harmony of the mutual relationships in the theory is such that I no longer have the slightest doubt about its correctness.[152]

In these passages Einstein left it somewhat open what he meant by qualifying the covariance he had reached as "most general" or "in the widest sense." In another contemporary passage he showed himself convinced that the transformation to a rotating coordinate system was comprised by this covariance:

> By means of a *simple* calculation I have been able to prove *that the gravitation equations hold for every reference system that is adapted to this condition*. From this it follows that there exist acceleration transformations of the most varied kind that transform the equations to themselves (e.g., also rotation), so that the equivalence hypothesis is preserved in its original form, even to an unexpectedly large extent.[153]

The "simple calculation" to which Einstein refers must be the demonstration of the covariance properties of the *Entwurf* equation published jointly with Grossmann, since he emphasizes the crucial element of this demonstration, the condition for adapted coordinate systems. He obviously perceived the more specific properties of the field equation, such as its covariance under rotation, as being merely a trivial consequence of this proof. Einstein thus believed he had achieved a full implementation of the generalized principle of relativity.

Yet, the exact relation of the *Entwurf* field equation to the absolute differential calculus had not been clarified. It seems, however, that Einstein did not bother too much about this problem. When he learned that Grossmann had finally succeeded in establishing such a relation, Einstein viewed this result as a nice complement to what they had already achieved earlier but not more. In late March or early April 1914 he wrote to Ehrenfest:

---

151 "Die Gravitation macht Fortschritte, aber unter ausserordentlichen Anstrengungen; sie ist spröde! Das Aequivalenzprinzip gilt nun doch in dem Sinne, dass es höchst allgemeine Transformationen gibt, die die Gravitationsgleichungen in sich überführen. Das Gefundene ist einf[a]ch, aber das Suchen ganz verflucht." Einstein to Paul Ehrenfest, before 10 March 1914, (CPAE 5, Doc. 512).

152 "Ich habe mich wieder bis zur Erschöpfung geplagt mit der Gravitationstheorie, aber diesmal mit unerhörtem Erfolge. Es ist nämlich der Beweis gelungen, dass die Gravit. Gleichungen für beliebig bewegte Bezugssysteme gelten, dass also die Hypothese von der Aequivalenz der Beschleunigung und des Gravitationsfeldes durchaus richtig ist, im weitesten Sinne. Nun ist die Harmonie der gegenseitigen Beziehungen in der Theorie eine derartige, dass ich an der Richtigkeit nicht mehr im Geringsten zweifle." Einstein to Heinrich Zangger, 10 March 1914, (CPAE 5, Doc. 513).

153 "Ich habe beweisen können durch eine *einfache* Rechnung, *dass die Gleichungen der Gravitation für jedes Bezugssystem gelten, welches dieser Bedingung angepasst ist*. Hieraus geht hervor, dass es Beschleunigungstransformationen mannigfaltigster Art gibt, welche die Gleichungen in sich selbst transformieren (z.B. auch Rotation), sodass die Aequivalenzhypothese in ihrer ursprünglichen Form gewahrt ist. sogar in ungeahnt weitgehendem Masse." Einstein to Michele Besso, ca. 10 March 1914, (CPAE 5, Doc. 514), Einstein's emphasis. The "einfache Rechnung" probably refers to the covariance proof. For an alternative interpretation, see (Janssen 1999, n. 125).

> Grossmann wrote me that now he also is succeeding in deriving the gravitation equations from the general theory of covariants. This would be a nice addition to our examination.[154]

The more Einstein thought about the proof of the covariance properties of the *Entwurf* field equation he had jointly developed with Grossmann, the more he became convinced that what he had reached was general covariance. This is apparent from his ever more optimistic assessments of their result. In June 1914 Einstein wrote to Wien:

> In Zurich I had found the proof for covariance in the gravitation equations. Now the theory of relat[ivity] really is extended to arbitrarily moving systems.[155]

In July Einstein wrote to Planck, also claiming that he had now his theory covered every possible manifold and that the restriction was only one of the coordinate system:

> Then also a brief reply to a comment you made recently at the Academy in the welcoming speech. There is an essential difference between the reference system restriction introduced by classical mechanics for the theory of relativity and that which I apply in the theory of gravitation. For the latter can always be adopted no matter how the $g_{\mu\nu}$'s may be selected. To the contrary, the specialization introduced by the principle of the constancy of the velocity of light presupposes differential correlations between the $g_{\mu\nu}$'s, that is, correlations that ought to be very difficult to interpret physically. Satisfaction of these correlations cannot be forced by the appropriate choice of a reference system for any given manifold.[156]

According to the explanation given to Planck, Einstein considered the principal distinction between the specialization of the reference system in classical mechanics and in the special theory of relativity, on the one hand, and that which he had introduced in his new gravitation theory, on the other hand, to be the fact that in the latter case the specialization of the reference system refers only to the choice of the coordinate system in an otherwise arbitrarily given manifold. Einstein made this point particu-

---

154 "Grossmann schreibt mir, dass es ihm nun auch gelingt, die Gravitationsgleichungen aus der allgemeinen Kovariantentheorie abzuleiten. Es wäre dies eine hübsche Ergänzung zu unserer Untersuchung." Einstein to Paul Ehrenfest, 10 April 1914, (CPAE 8, Doc. 2).

155 "In Zürich fand ich noch den Nachweis der Kovarianz der Gravitationsgleichungen. Nun ist die Relat[ivitäts]theorie wirklich auf beliebig bewegte Systeme ausgedehnt." Einstein to Wilhelm Wien, 15 June 1914, (CPAE 8, Doc. 14).

156 "Sodann noch eine kurze Beantwortung einer Bemerkung, die Sie neulich in der Akademie in der Begrüssungsrede geäussert haben. Es gibt einen prinzipiellen Unterschied zwischen derjenigen Spezialisierung des Bezugssystems, welche die klassische Mechanik bezw. die Relativitätstheorie einführt und zwischen derjenigen, welche ich in der Gravitationstheorie anwende. Die letztere kann man nämlich stets einführen, wie auch die $g_{\mu\nu}$ gewählt werden mögen. Diese durch das Prinzip der Konstanz der Lichtgeschwindigkeit eingeführte Spezialisierung dagegen setzt Differenzialbeziehungen zwischen den $g_{\mu\nu}$ voraus, und zwar Beziehungen, deren physikalische Interpretation sehr schwierig sein dürfte. Das Erfülltsein dieser Beziehungen kann nicht für jede gegebene Mannigfaltigkeit durch passende Wahl des Bezugssystems erzwungen werden." Einstein to Max Planck, 7 July 1914, (CPAE 8, Doc. 18).

larly clear in a letter he wrote to Lorentz a few months later. In this letter Einstein explained in what sense the restriction to adapted coordinate systems in his understanding was compatible with the claim that the theory would be a "general" theory of relativity. He referred to an analogous situation in the Gaussian theory of surfaces:

> Although I prefer certain reference systems, the fundamental difference to the Galilean preference is, however, that my coordinate selection makes no physical assumptions about the world; let this be illustrated by a geometric comparison. I have a plane of unknown description which I want to subject to geometric analysis. If I require that a coordinate system (p, q) on the plane be selected in such a way that
>
> $$ds^2 = dp^2 + dq^2,$$
>
> I therefore assume that then the surface can be unfolded on to a [Euclidean] plane. Were I only to demand, however, that the coordinates be chosen in such a way that
>
> $$ds^2 = A(p, q)dp^2 + B(p, q)dq^2$$
>
> i.e., that the coordinates be orthogonal, then I am assuming nothing about the nature of the surface; this can be obtained on any surface.[157]

The analogy with Gaussian surface theory suggests a geometrical interpretation of the coordinate restriction introduced in Einstein's theory of gravitation. A letter Einstein wrote in 1915 to Paul Hertz shows that he had searched in vain for such an interpretation and that for elucidating the meaning of this coordinate restriction, he had little more to offer than the comparisons he mentioned in the letter to Lorentz.

> He who has wandered aimlessly for so long in the chaos of possibilities understands your trials very well. You do not have the faintest idea what I had to go through as a mathematical ignoramus before coming into this harbor. Incidentally, your idea is very natural and would by all means be worth following up, if it could be carried through at all, which, based upon my experiences gathered during my wayward wanderings, I doubt very much.
>
> Given an arbitrary manifold of 4 dimensions (given $g_{\mu\nu}(x_\sigma)$). How can one distinguish a coordinate system or a group of such? This appears not to be possible in any way simpler than the one chosen by me. I have groped around and tried all sorts of possibilities, e.g., required: The system must be chosen such that the equations
>
> $$\sum_\nu \frac{\partial g^{\mu\nu}}{\partial x_\nu} = 0 \qquad (\mu = 1-4)$$

157 "Ich bevorzuge zwar auch gewisse Bezugssysteme, aber der fundamentale Unterschied gegenüber der Galileischen Bevorzugung besteht darin, dass meine Koordinatenwahl nichts über die Welt voraussetzt; dies sei durch einen geometrischen Vergleich erläutert. Es liegt mir eine Fläche unbekannter Art vor, auf der ich geometrische Untersuchungen machen will. Verlange ich, es solle auf der Fläche ein Koordinatensystem (p, q) so gewählt werden, dass $ds^2 = dp^2 + dq^2$, [s]o setze ich damit voraus, dass die Fläche auf eine Ebene abwickelbar sei. Verlange ich aber nur, dass die Koordinaten so gewählt seien, dass $ds^2 = A(p, q)dp^2 + B(p, q)dq^2$ ist, d.h. dass die Koordinaten orthogonal seien, so setze ich damit über die Natur der Fläche nichts voraus; man kann dies auf jeder Fläche erzielen." Einstein to H.A. Lorentz, 23 January 1915, (CPAE 8, Doc. 47).

> are satisfied throughout.
> At least it seemed definite to me *a priori* that a transformation group exceeding the Lorentz group must exist, because those observations summed up in the words "relativity principle" and "equivalency principle" point to it.
> The coordinate limitation that was finally introduced deserves particular trust because it establishes a link between it and the postulate of the event's complete determination.
> A theoretical differential geometric interpretation of preferred systems would be of great value. The weakest point of the theory as it stands today consists precisely in this, that the group of justified transformations are by no means closely assessable. There is not even any *exact* proof that arbitrary motions can be transformed to motionlessness.[158]

The letter shows that Einstein saw all coordinate restrictions he had examined to function essentially on the same level, that is, to be generally imposed as conditions supplementary to the field equations; this is made clear by his formulation that he assumed what we have called the "Hertz restriction" eq. (60) to be satisfied "everywhere." He evidently treated the Hertz restriction on the same level as the condition eq. (54) for adapted coordinate systems, despite their different form. Both conditions were motivated, as we have seen, by the conservation principle. But, as Einstein points out in his letter, the condition for adapted systems could also be justified on a deeper level; the causality considerations were related to the hole argument, and therefore inspired more confidence. The letter to Hertz furthermore confirms that Einstein was convinced that this condition just implies a particular choice of the coordinate system without restricting the range of possible manifolds. He implicitly claimed that, in the *Entwurf* theory, all motions can be transferred to rest, although he admitted that he had been unable to demonstrate this "exactly."

---

158 "Wer selber im Chaos der Möglichkeiten sich so viel herumgetrieben hat, begreift Ihre Schicksale sehr gut. Sie haben ja keine blasse Ahnung, was ich als mathematischer Ignorant habe durchmachen müssen, bis ich in diesen Hafen eingelaufen bin. Übrigens ist Ihre Idee sehr natürlich und wäre auf jeden Fall ernster Verfolgung wert, wenn sie sich überhaupt durchführen liesse, was ich auf Grund meiner im Herumirren allmählich angesammelten Erfahrung sehr bezweifle.
Gegeben eine beliebige Mannigfaltigkeit von 4 Dimensionen ($g_{\mu\nu}(x_\sigma)$ gegeben). Wie kann man ein Koordinatensystem bezw. eine Gruppe von solchen auszeichnen? Es scheint dies auf einfacher als die von mir gewählte Art nicht möglich zu sein. Ich habe herum getastet und alles Mögliche versucht, z.B. verlangt: Das System soll so gewählt werden, dass überall die Gleichungen [eq.] erfüllt seien.
Immerhin schien es mir a priori sicher, dass eine über die Lorentz-gruppe hinausgehende Transformationsgruppe vorhanden sein müsse, da jene Erfahrungen, die mit den Worten Relativitätsprinzip, Aequivalenzprinzip zusammengefasst werden, darauf hinweisen.
Die schliesslich eingeführte Koordinatenbeschränkung verdient deshalb besonderes Vertrauen, weil sie sich mit dem Postulat der vollständigen Bedingtheit des Geschehens in Zusammenhang bringen lässt.
Eine flächentheoretische Interpretation der bevorzugten Systeme wäre von sehr grossem Werte. Der schwächste Punkt der Theorie bei ihrem heutigen Stande besteht nämlich gerade darin, dass man die Gruppe der berechtigten Transformationen durchaus nicht scharf übersieht. *Exakt* ist nicht einmal der Beweis geliefert, dass beliebige Bewegungen auf Ruhe transformiert werden können." Einstein to Paul Hertz, 22 August 1915, (CPAE 8, Doc. 111). For an extensive discussion of this letter, see (Howard and Norton, 1993).

In spite of such reservations, Einstein was nevertheless convinced that he had reached all his major original heuristic goals within the *Entwurf* theory. What remained were only some minor problems, such as the establishment of the connection of the methods used by Einstein and Grossmann to the absolute differential calculus, and a clarification of some other mathematical aspects of the theory. Einstein also had the impression that the crucial proof of the covariance properties of the *Entwurf* field equation still required improvement. This was the task he set himself in mid-1914 in the context of composing a major review article, finished by the end of October and providing a full exposition of the finally complete theory which now was called, for the first time, the "general theory of relativity" (Einstein 1914a).

7.12 The Revised Covariance Proof and the Definitive Formulation of the Hole Argument[159]

When at the end of 1914 Einstein looked back on his first review article on general relativity, entitled "The Formal Foundations of General Relativity" and submitted on 29 October 1914, the revision of the covariance proof appeared to him as the central achievement, as is suggested in a letter he wrote in December to Paul Ehrenfest:

> In recent months I reworked extremely carefully the basis of the general theory of rel. The covariance proof of last spring was not yet completely right. Otherwise, I have also been able to penetrate a few things more clearly. Now I am entirely satisfied with that matter. You will soon receive the paper; read it, you will find it very enjoyable.[160]

The proof of the covariance properties of the *Entwurf* field equation as originally conceived by Einstein and Grossmann was based on the idea that an infinitesimal, adapted coordinate transformation leaves the variation of the action integral invariant. The variation of the manifold giving rise to the variation of this integral had to be performed in two steps, an "adapted" variation, making it possible to vary the coordinate system along with the manifold so that it remains adapted to it, and a variation of the manifold that merely corresponds to the introduction of a new coordinate system, a "coordinate variation." The problem with the original proof was that the first of these two variations was not clearly defined. In fact, Einstein and Grossmann had obvious difficulties in arguing for the possibility of an appropriate variation of the adapted coordinate system "following" that of the manifold. This variation of the coordinate system was introduced rather as an afterthought to the variation of the manifold, an afterthought which left open exactly how the variation of the manifold is restricted by the condition that it must be possible to vary the adapted coordinate system along

---

159 See (Cattani and De Maria 1989b).

160 "In den letzten Monaten habe ich die Grundlage der allgemeinen Rel Theorie nochmals höchst sorgfältig bearbeitet. Der Kovarianzbeweis vom letzten Frühjahr war noch nicht ganz in Ordnung. Auch sonst habe ich manches klarer durchdringen können. Nun bin ich aber völlig zufrieden mit der Angelegenheit. Du erhälst bald die Arbeit, lies sie, Du wirst grosse Freude daran finden." Einstein to Paul Ehrenfest, December 1914 (CPAE 8, Doc. 39).

with it. The origin of Einstein's difficulties was his concept of a manifold being closely tied to its representation by the metric tensor and hence lacked the clear-cut distinction from the representation of a manifold in terms of coordinates.[161]

At the outset of his new approach, Einstein distinguished more clearly a variation of the manifold and a variation of the coordinate system. It was probably for the purposes of such a cleaner separation of the different kinds of variations that he treated coordinate systems—as suggested by the hole argument and in contrast to the modern understanding—as being essentially given independently from the manifold for whose description they serve. Einstein believed that in this way he could refer to two different manifolds, or rather one manifold before and, after the variation, to one and the same coordinate system.[162] The ensuing challenge to refer changes of the values of the metric tensor due to a coordinate transformation to one and the same coordinate system, given independently from the manifold, may well have induced him to formulate more clearly than he had done before the artifice of transposing values of the metric tensor characteristic of the hole argument.

Einstein's treatment of the hole argument in his 1914 review paper is in fact the first published version of this argument that makes plain how values of the metric tensor given at one point of the manifold are to be referred to another point, a notion implicit in its original formulation in late August 1913 but obscured in the subsequent published presentations. It is also the first version that introduces a distinct notation for the metric tensor and its representation in a particular coordinate system.[163] The mature and more elaborate formulation of the hole argument was hence closely associated with the reworking of the covariance proof. Revisiting, together with Marcel Grossmann, the covariance properties of the *Entwurf* field equation, Einstein arrived at a formulation of this argument that now pointed to philosophical questions concerning the mathematical representation of the physical properties of space and time.

7.13 A Shaky Mathematical Derivation and a Spin-off with Consequences

In the introduction to his 1914 review paper Einstein mentioned the peculiar combination of physical and mathematical arguments that led him to the *Entwurf* theory and announced a purely mathematical derivation:

> In recent years I have worked, in part together with my friend Grossmann, on a generalization of the theory of relativity. During these investigations, a kaleidoscopic mixture of postulates from physics and mathematics has been introduced and used as heuristical tools; as a consequence it is not easy to see through and characterize the theory from a formal mathematical point of view, that is, only based upon these papers. The primary

---

161 See (Norton 1992b).

162 See (Einstein 1914a, 1071–1073). Einstein conceived a variation of the metric tensor generated by a coordinate transformation, referring its result to the same original coordinate system. His transformation can thus not be an ordinary coordinate transformation, but must be the kind of transport of values of the metric tensor from one coordinate system to the other as it is essential to the hole argument.

163 See (Einstein 1914a, 1067).

> objective of the present paper is to close this gap. In particular, it has been possible to obtain the equations of the gravitational field in a purely covariance-theoretical manner ... .[164]

The basis for this derivation was provided by the variational formalism. In their paper of early 1914, Einstein and Grossmann had, as we have seen, solved the problem of identifying an appropriate Lagrangian from which the *Entwurf* equations were derived. Since their Lagrangian now represented the natural starting point for building up the entire edifice of Einstein's theory, the question arose whether this Lagrangian could be justified by reasons other than that of generating the desired field equation. In order to answer this question, Einstein generalized the formalism jointly developed with Grossmann to apply to an arbitrary Lagrangian. While this generalization was rather straightforward, it was a more challenging task to pinpoint the assumptions by which the resulting formalism could be specialized again so as to determine the Lagrangian appropriate for the *Entwurf* field equation. Einstein's approach was effectively guided by the mathematical strategy presupposing a generic mathematical object, which is then specialized in light of concrete physical requirements. In the generalized formalism of his 1914 review paper, such physical requirements had to be formulated as mathematical criteria serving to select the *Entwurf* Lagrangian.

It was a combination of two criteria that helped Einstein to achieve this goal, one derived from the conservation principle, the other from the generalized principle of relativity. He formulated both criteria in terms of differential conditions for the Lagrangian and concluded, erroneously as it later turned out, that the requirement of their compatibility singles out a particular candidate. In the *Entwurf* theory, the implementation of the conservation principle imposed, as we have seen, a coordinate restriction $B_\mu = 0$ (cf. eq. (54)). This condition played, as we have also seen, the double role of ensuring the satisfaction of the conservation principle and of determining the covariance properties of the field equation. The exact same equation $B_\mu = 0$ also played a role in Einstein's interpretation of the generalized theory, but here only in the context of analyzing its covariance properties. The formulation of the conservation principle within the generalized framework yielded a slightly different equation, now comprising two terms instead of one:

$$\sum_\nu \frac{\partial S_\mu^\nu}{\partial x_\nu} - B_\mu = 0. \tag{86}$$

164 "In den letzten Jahren habe ich, zum Teil zusammen mit meinem Freunde Grossmann, eine Verallgemeinerung der Relativitätstheorie ausgearbeitet. Als heuristische Hilfsmittel sind bei jenen Untersuchungen in bunter Mischung physikalische und mathematische Forderungen verwendet, so daß es nicht leicht ist, an Hand jener Arbeiten die Theorie vom formal mathematischen Standpunkte aus zu übersehen und zu charakterisieren. Diese Lücke habe ich durch die vorliegende Arbeit in erster Linie ausfüllen wollen. Es gelang insbesondere, die Gleichungen des Gravitationsfeldes auf einem rein kovarianten-theoretischem Wege zu gewinnen ... ." (Einstein 1914a, 1030)

The existence of two similar but not identical conditions could be turned into a compatibility argument identifying the *Entwurf* theory as a special case of the generalized formalism. He thus demanded:

$$S_\mu^\nu \equiv 0, \tag{87}$$

and claimed that this condition, together with the additional requirement that the Lagrangian be a homogeneous function of second degree in the gravitational fields, determines uniquely the *Entwurf* Lagrangian. Einstein's additional "mathematical" requirement has, as is the case for his other constraints, also physical aspects, here the analogy with the Lagrangian for a free electromagnetic field (cf. the default setting eq. (LXIII)). In his paper Einstein did not explicitly prove his claim. It may well have been his long-held conviction that the conservation principle determines uniquely the *Entwurf* field equation, that simply correlated with his belief that the *Entwurf* theory can be uniquely characterized with the help of eq. (87).

It turned out later that Einstein's reasoning was flawed. A more careful analysis of his formalism later showed him that eq. (87) did not actually impose a strong additional selective criterion, but could be easily fulfilled by simply requiring that the Lagrangian be an invariant under linear transformations, a criterion that does not help to single out the *Entwurf* theory. Einstein had imposed this requirement implicitly in the context of his analysis of the covariance properties and had effectively suppressed the condition involving S_μ^ν in this context, thus coming up with what appeared to be two different sets of conditions, one derived from the generalized principle of relativity, the other from the conservation principle. A deeper exploration of his formalism, first achieved about a year later, eventually offered him the insight that the analysis of the covariance and the conservation aspects actually implied the same set of conditions, an important step towards what later became Noether's theorem. This step was prepared by Einstein's seemingly successful attempt to derive the *Entwurf* theory along a mathematical strategy in which, alongside the conservation principle, covariance considerations had assumed the role of the correspondence principle in restricting the admissible candidate Lagrangians. More than anything else, it was the supposed achievement of being able to renounce the correspondence principle as part of Einstein's derivation that gave it the appearance of being largely independent of specific physical knowledge about gravitation (Einstein 1914a, 1076).

Nevertheless, what Einstein had achieved was satisfactory also from a physical point of view. In the course of his mathematical elaboration of the *Entwurf* theory, he had brought its field equation into a form satisfying all structural requirements following from the conservation principle. In particular, he succeeded in identifying, even for a generic Lagrangian, an expression for the stress-energy tensor of the gravitational field, i.e. for **FIELDMASS**. Furthermore, he was able to write the field equation in a form corresponding to eqs. (XLIV) and (XLV), thus demonstrating the structural analogy with classical field theory as well as the parallelism between the stress-energy tensor of matter and of the gravitational field as sources of the field. In this way, Einstein had built up a mathematical framework lending itself to direct physical interpretation:

> The system of equations (81) allows for a simple physical interpretation in spite of its complicated form. The left-hand side represents a kind of divergence of the gravitational field. As the right-hand side shows, this is caused by the components of the total energy tensor. A very important aspect of this is the result that the energy tensor of the gravitational field itself acts field-generatingly, just as does the energy tensor of matter.[165]

7.14 From Consolidation to Exploration

7.14.1 Living with the Less than Perfect

If considered in hindsight of general relativity, the *Entwurf* theory has considerable flaws: it does not comply with the only astronomical test available before 1919[166] for a relativistic gravitation theory, the anomalous perihelion advance of Mercury by ca. 43” per century, which is inexplicable in terms of Kepler’s laws; it does not include the Minkowski metric in rotating coordinates as a solution and hence disappointed Einstein’s Machian expectations; and the mathematical derivation from general principles was based on an error. Einstein’s discovery of these flaws in the *Entwurf* theory may appear to constitute a step-by-step refutation, clearing the way for a new approach. However, uncovering these flaws did not immediately shatter the *Entwurf* theory. As was shown above, the *Entwurf* theory had emerged as a theory firmly grounded in the knowledge of classical physics, incorporating, in particular, both the correspondence and the conservation principles. At the same time, the theory allowed for a limited extension of the generalized relativity principle to at least general linear transformations, this limitation being, however, justified by both physical and mathematical arguments. Whatever Einstein achieved in the second phase beyond this state—in terms of an astronomical confirmation of the theory, of a further generalization of the relativity principle, or in terms of its mathematical elaboration—was not necessary to support the theory. The successes and failures beyond the core established in the consolidation period concerned the more ambitious part of Einstein’s heuristics, in particular the extension of the generalized principle of relativity, which from the beginning was a less stringent criterion for the validity of his new gravitation theory than its relation to the knowledge of classical physics.

In this section, we shall briefly assess the impact of the discovery of flaws in the *Entwurf* theory on Einstein’s attitude with respect to his theory. It demonstrates his

165 “Das Gleichungssystem (81) [cf. eq. (52)] läßt trotz seiner Kompliziertheit eine einfache physikalische Interpretation zu. Die linke Seite drückt eine Art Divergenz des Gravitationsfeldes aus. Diese wird—wie die rechte Seite zeigt—bedingt durch die Komponenten des totalen Energietensors. Sehr wichtig ist dabei das Ergebnis, daß der Energietensor des Gravitationsfeldes selbst in gleicher Weise felderregend wirksam ist wie der Energietensor der Materie.” (Einstein 1914b, 1077)

166 For a discussion of the status of the other two classical tests, gravitational light bending and gravitational redshift, by 1919, see the introduction to (CPAE 9).

ability to live with the less than perfect or, more specifically, his resistance to abandoning an elaborate edifice because of damage it suffered on one of its floors.

7.14.2 The Mercury Problem

In temporal order, Einstein's discovery of the failure of the *Entwurf* theory to yield the correct perihelion shift of Mercury came first; it was made as early as the summer of 1913. The extensive research notes by Einstein and Besso, which document their joint effort to calculate Mercury's perihelion motion, show that this endeavor was actually part of a broader program that included not only the *Entwurf* theory, but also Nordström's gravitation theory, and not only the perihelion shift of Mercury, but also other possible checks of a non-Newtonian gravitation theory, such as its compatibility with the effects anticipated on the basis of Einstein's Machian heuristics. This broader perspective may have shaped Einstein's reaction to finding that the *Entwurf* theory could not account for the astronomically observed value of the perihelion shift. First of all, this anomaly could not be explained by other contemporary gravitation theories; second, there might have been a purely astronomical explanation for it; and third, there was a range of other possible checks of the *Entwurf* theory, such as the deflection of light in a gravitational field and gravitational redshift. In view of this situation, the negative finding on Mercury's perihelion shift was not a result of immediate significance for the validity of the *Entwurf* theory. It had required some effort to perform the perihelion calculation, but from the beginning it must have been at best a hope that a relativistic gravitation theory could actually account for this effect. Einstein himself did not publish his negative result. He encouraged Besso to complete a paper offering a comparative evaluation of contemporary gravitation theories both on empirical and epistemological grounds.[167] In his contemporary letters, he appeared more convinced of or worried by, as the case might be, the theory's internal consistency or lack thereof.

The failure of the perihelion calculation was not mentioned in Einstein's publications and hardly ever in his contemporary correspondence. It only played a role in Einstein's later justifications of his abandonment of the *Entwurf* theory. If his and Besso's extensive manuscript notes had not survived, one would not have known how much effort they had invested into this calculation. And yet, this calculation had a profound impact on the genesis of general relativity, which is discussed more extensively below, by affecting the speed with which Einstein could calculate the perihelion shift on the basis of his later generally-covariant theory.[168] This was possible because the formalism he had developed for the *Entwurf* theory turned out to be more generally applicable and hardly required any modification when used in the context of another gravitation theory. But the perihelion calculation also had more subtle effects which, as we shall see, later turned out to be beneficial to Einstein's renewed

---

167 Einstein to Michele Besso, after 1 January 1914 (CPAE 5, Doc. 499).
168 For detailed historical discussion, see (Earman and Janssen 1993).

exploration of generally-covariant candidate field theories. It led, in particular, to an improved understanding of the Newtonian limit.

7.14.3 The Rotation Problem

The method developed by Einstein and Besso for calculating the perihelion advance was based on an iterative procedure for finding approximate solutions of the field equation. It also turned out to be applicable to the investigation of another question of great heuristic significance for Einstein's attempt to generalize the relativity principle.[169] As we have discussed earlier, this attempt was guided, from the beginning, by the idea of conceiving rotation as being equivalent to a state of rest, interpreting the inertial forces arising in a rotating frame of reference as a special gravitational field. If the *Entwurf* theory were actually compatible with this heuristics, the Minkowski metric in rotating coordinates should be a solution of its field equations.

The inertial forces arising in a rotating frame, centrifugal and Coriolis forces, are of a different order in the angular velocity, the centrifugal force depending on its square, the Coriolis forces depending linearly on this velocity. Einstein and Besso's approximation scheme could thus be used to check whether one can obtain from a first-order approximation of a Minkowski metric in rotating coordinates, containing only the components relevant for the Coriolis forces, the correct second order term relevant for the centrifugal forces. The result of this calculation could then be compared with that of the direct transformation of the Minkowski metric in rotating coordinates.

In a scratch notebook Einstein first wrote down the one component relevant for the centrifugal force and then two components relevant for the Coriolis force. Underneath he wrote:

> Is the first equation [concerning the centrifugal force] a consequence of the other two [concerning the Coriolis force] on the basis of the theory?[170]

In a page of the bundle of manuscripts used jointly by Einstein and Besso for their calculations on the perihelion shift, Einstein actually performed this check (CPAE 4, Doc. 14, [41–42]). Although the approximation scheme applied to the *Entwurf* theory does not yield the correct value for the 4–4 component of a Minkowski metric in rotating coordinates, he at first came to the erroneous conclusion that it actually does, and ended his calculation with the remark "stimmt" (CPAE 4, Doc. 14, [41]).

There is, however, as early as 1913 evidence that this was not Einstein's last word. In the draft for his paper on contemporary gravitation theories, Besso listed the failure of the *Entwurf* theory to yield the correct combination of centrifugal and Coriolis forces, in other words, its failure to include the Minkowski metric in rotating coordi-

---

169 For a discussion of this procedure, see the editorial note on the Einstein-Besso manuscript in (CPAE 4, 346–349), as well as (Earman and Janssen 1993, 142–143).

170 "Ist die erste Gleichung Folge der beiden letzten auf Grund der Theorie?" (CPAE 3, Appendix A, [66]).

nates as a solution, as problems to be thought about and to be discussed with Einstein. From these notes it seems that at some point by the end of August 1913, Besso was aware of this problem. Einstein's contemporary correspondence suggests that he as well had realized by mid-August that the positive result mentioned above was actually based on an error.[171]

The problem resurfaced only when Einstein had convinced himself, after the discovery of a flaw in his original conservation argument (cf. subsection "Einstein's Reinterpretation of the Conservation Principle," p. 235, that there actually exists a large variety of transformation to accelerated reference systems under which the *Entwurf* theory was covariant. On 20 March 1914, Michele Besso wrote to Einstein, after the latter had reported his progress in analyzing the covariance properties of the *Entwurf* field equations:

> Does the result obtained also give a clue, perhaps, for a more complete treatment of the rotation problem, so that one can get the correct value of the centrifugal force? Unfortunately, my brain, at least the way it has been trained, is much too feeble for me to answer this question myself, or even to guess from what side it could be attacked. For reasons already discussed, it seems to me that it (?) is of importance for the astronomical problem as well (for until now it at least seemed that a system in which no Coriolis forces flit about could still be a seat of centrifugal forces, or the reverse).[172]

The passage clearly confirms that Besso was aware by spring 1914—and also assumed Einstein to be aware—that the "incomplete treatment" of the problem of rotation (probably referring to the use of an approximation procedure) did not yield the correct result for the Coriolis force. Besso also claimed that the solution to this problem might be relevant for the calculation of the perihelion shift of Mercury (possibly the "astronomical problem" to which he alluded). Einstein did not, however, at this point check the compatibility of his general insights into the covariance properties of the *Entwurf* field equations with concrete calculations on the level of his approximation scheme.

It was only in September 1915 that Einstein rediscovered, to his surprise, the result that the iterative solution of the *Entwurf* field equations does not yield the correct Minkowski metric in rotating coordinates. This is known from a letter he wrote

---

171 See, e.g., Einstein to H. A. Lorentz, 14 August 1913, (CPAE 5, Doc. 467). Einstein did not allude to anything like having established the Minkowski metric in rotating coordinates as a solution of the *Entwurf* equations (at least in second-order approximation) in his letters to Lorentz nor in those he wrote to other colleagues and friends, who would have been interested in the issue, such as Ernst Mach, Erwin Freundlich, Heinrich Zangger, Paul Ehrenfest, and Michele Besso. See also the extended discussion in "What Did Einstein Know ..." (in vol. 2 of this series).

172 "Gibt das erreichte Resultat vielleicht auch einen Wink für eine vollständigere Behandlung des Drehungsproblems, so dass man den richtigen Wert der Centrifugalkraft bekommen kann? Leider ist mein Kopf, wenigstens so wie er einmal erzogen ist, viel zu schwach, um mir die Frage selbst zu beantworten, oder auch nur zu ahnen, wo man sie angreifen könnte. Aus schon besprochenen Gründen scheint sie (?) mir auch für das astronomische Problem von Bedeutung (weil es früher wenigstens so aussah, ein System in welchem keine Corioliskräfte huschen, doch Sitz von Centrifugalkräften sein könnte, oder umgekehrt)." Michele Besso to Einstein, 20 March 1914, (CPAE 5, Doc. 516, 606).

on 30 September 1915 to Erwin Freundlich, in which he now also connected this finding with the perihelion problem, just as Besso had done in the letter quoted above. Evidently Einstein was quite concerned by his finding:

> I am writing you now about a scientific matter that electrifies me enormously. For I have come upon a logical contradiction of a quantitative nature in the theory of gravitation, which proves to me that there must be a calculational error somewhere within my framework. [...]
>
> Either the equations are already numerically incorrect (numerical coefficients), or I am applying the equations in a principally incorrect way. I do not believe that I myself am in the position to find the error, because my mind follows the same old rut too much in this matter. Rather, I must depend on a fellow human being with unspoiled brain matter to find the error. If you have time, do not fail to study the topic.[173]

The letter leaves open in which context Einstein redid the earlier calculation. It is plausible to assume that it was once more Besso who stimulated the reconsideration of this problem. In fact, Besso and Einstein probably discussed the Mercury as well as the rotation problem during the latter's stay in Switzerland in September 1915.[174] The letter to Freundlich was sent only a week after Einstein's return to Berlin.[175] It was probably written as a reaction to a request for political support and represented one of the first occasions for Einstein to present the revived rotation problem to a colleague who must have been interested in it because of its implication for the understanding of the Mercury problem.[176]

Evidently, this time Einstein found the rotation problem much more alarming than he did in the summer of 1913. In his letter to Freundlich he still did not substantially question the *Entwurf* field equation but merely took into consideration that he applied

---

173 "Ich schreibe Ihnen jetzt in einer wissenschaftlichen Angelegenheit, die mich ungeheuer elektrisiert. Ich bin nämlich in der Gravitationstheorie auf einen logischen Widerspruch quantitativer Art gestossen, der mir beweist, dass in meinem Gebäude irgendwo eine rechnerische Unrichtigkeit stecken muss. [...] Entweder sind die Gleichungen schon numerisch unrichtig (Zahlenkoeffizienten) oder ich wende die Gleichungen prinzipiell falsch an. Ich glaube nicht, dass ich selbst imstande bin, den Fehler zu finden, da mein Geist in dieser Sache zu ausgefahrene Gleise hat. Ich muss mich vielmehr darauf verlassen, dass ein Nebenmensch mit unverdorbener Gehirnmasse den Fehler findet. Versäumen Sie nicht, wenn Sie Zeit haben, sich mit dem Gegenstande zu beschäftigen." Einstein to Erwin Freundlich, 30 September 1915 (CPAE 8, Doc. 123), extensively discussed in (Janssen 1999), here just a summary,

174 He wrote to Elsa Einstein from Lucerne: "In Zurich I was together with Besso very often; my stay in Zurich was very much improved by it, but thus I neglected my duties to others." ("In Zürich war ich sehr viel mit Besso zusammen, mein Aufenthalt wurde dadurch sehr verschönert, doch vernachlässigte ich so meine Pflicht gegen andere.") Einstein to Elsa Einstein, 11 September 1915, (CPAE 8, Doc. 116).

175 See Calender (CPAE 8, 998).

176 Einstein had written to Freundlich in March of the same year on the perihelion problem, see Einstein to Erwin Freundlich, 19 March 1915 (CPAE 8, Doc. 63). He had written a letter to Lorentz a day after his return from Switzerland, Einstein to H. A. Lorentz, 23 September 1915, (ibid., Doc. 122), in which he did not mention this problem, probably because he was ashamed.

it incorrectly—probably a reference to the unclear status of the approximation procedure—or that some numerical coefficients were wrong.

That the discovery of this flaw, taken by itself, did not amount to a refutation of the *Entwurf* theory is made evident also by Einstein's immediate reaction to the rotation problem. Apparently encouraged by his general results on the covariance properties of the *Entwurf* theory, which, as we have seen, amounted for him to the claim that there was no physical restriction of the generalized relativity principle but only on the choice of admissible coordinates, he attempted to show that the *Entwurf* field equation could be solved by a rotating system in a different set of coordinates; but this attempt failed as well.[177] Shortly afterwards, he developed a new derivation of the *Entwurf* field equation, to which we will turn below, effectively confirming its immunity with regard to the rotation problem. It was only after his return to the November tensor, that he listed the problem of rotation as one of the three flaws which undermined his trust in the *Entwurf* theory.

Einstein's diverse reactions over the course of time to the same problem of the *Entwurf* theory are correlated with his changing perspectives during the elaboration of this theory. When he first believed that his rotation calculation worked, it seemed like progress on the bold approach, without providing him with a general insight into the covariance properties of the *Entwurf* theory. When it turned out that it does not actually work, this negative result became irrelevant because Einstein successfully developed his defensive approach with the supposed consequence that the *Entwurf* field equation is covariant only under linear transformations. When Einstein then, in the second phase of the consolidation period of the *Entwurf* theory, achieved more general insights into its covariance properties, these insights seemed to make a check on the level of concrete calculations superfluous. Eventually, Einstein took it for granted that rotation did not present a problem for the *Entwurf* theory. Only when he reviewed the problem in September 1915, possibly at Besso's prompting, he finally connected his general considerations with his concrete calculations—and rediscovered the problem. This result now questioned not only his earlier conviction concerning rotation, but also more generally the significance of his insights into the theory's covariance properties. Still, the discovery of this failure implied little more than a return to the status of the *Entwurf* theory at the end of the first phase of the consolidation period, its covariance being guaranteed only for general linear transformations.

7.14.4 The Failure of the Covariance Proof

We now turn to the last flaw that Einstein discovered, probably in early October, some weeks before he gave up the *Entwurf* theory. This flaw concerns Einstein's attempt to derive the *Entwurf* field equation along the mathematical strategy. As our earlier discussion of this endeavor has made clear, one of its problems was the neces-

177 See the calculations on the back of the draft of letter Einstein wrote to Otto Naumann after 1 October 1915, (CPAE 8, Doc. 124).

sity to adapt general tools, such as variational calculus, to the requirement of restricted covariance. This aspect had been at the center of Einstein's discussion in early 1915 with the Italian mathematician Tullio Levi-Civita, which apparently was triggered by a letter from Max Abraham.

Abraham may well have been one of the first to discover a problem with Einstein's derivation of the *Entwurf* field equation from a Lagrangian function. On 23 February 1915 he wrote to Levi-Civita:

> Really I did not understand on which hypotheses his new demonstration is based. Among all possible invariants that could be used to construct the [Lagrangian] function *H* he chooses very arbitrarily the one that yields his field equations.[178]

Abraham thus succinctly summarized the crucial weakness of Einstein's proof.

But Levi-Civita's exchange with Einstein did not touch upon this crucial problem. Levi-Civita focused instead on a specific technical problem in Einstein's derivation; his proof of the claim that the candidate for the left-hand side of the field equations is a tensor.[179] He produced a counter-example which Einstein, however, declared irrelevant by pointing to the fact that Levi-Civita's example does not satisfy the condition of being covariant under the linear transformations that he had explicitly stipulated.[180] As we shall see, he only later would realize the questionable role of this condition in his proof. Levi-Civita, in any case, did not insist on this aspect. Einstein had more difficulties in responding to other problems in his proof to which Levi-Civita drew his attention. In spite of his attempts to rebut the latter's criticism, Einstein eventually had to admit that his derivation was incomplete, without, however, losing faith in it actually fulfilling its purpose in yielding the *Entwurf* field equations.[181] On the contrary, as Einstein wrote to Levi-Civita during their controversy:

> I must even admit that, through the in-depth considerations to which your interesting letters have led me, I have become only more firmly convinced that the proof of the tensor character of $\mathfrak{E}_{\mu\nu}/\sqrt{-g}$ is correct in principle.[182]

Further objections by Levi-Civita did not shatter this conviction. Nevertheless Einstein and Levi-Civita agreed upon a shortcoming of Einstein's proof; eventually when Levi-Civita proposed an alternative gravitation theory involving a scalar gravitational potential, Einstein lost interest in the exchange and broke it off.[183]

Einstein's discovery of the crucial flaw in his proof was not stimulated by Levi-Civita's criticism but by a reconsideration of this proof in light of a paper by Lorentz about six months later. When Einstein returned from Switzerland on 22 September

---

178 Quoted after (Cattani and De Maria 1989b, 184–185).

179 For an extensive discussion, see (Cattani and De Maria 1989b).

180 See Einstein to Tullio Levi-Civita, 5 March 1915 (CPAE 8, Doc. 60).

181 See Einstein to Tullio Levi-Civita, 5 May 1915, (CPAE 8, Doc. 80).

182 "Ich muss sogar gestehen, dass ich durch die tieferen Überlegungen, zu denen mich Ihre interessanten Briefe brachte, noch fester in der Überzeugung wurde, dass der Beweis vom Tensorcharakter von $\mathfrak{E}_{\mu\nu}/\sqrt{-g}$ im Prinzip richtig ist." Einstein to Tullio Levi-Civita, 8 April 1915 (CPAE 8, Doc. 71)

183 See Einstein to Tullio Levi-Civita, 5 May 1915, (CPAE 8, Doc. 80).

1915, he found Lorentz's recently published paper on Hamilton's principle in the theory of gravitation, including a treatment of electromagnetic fields. In a letter to Lorentz dated the following day Einstein wrote:

> Your article delighted me. I have also found a proof for the validity of the [relativistic] energy-momentum conservation principle for the electromagnet. field taking gravitation into consideration, as well as a simplified covariant theoretical representation of the vacuum equations, in which the "dual" six tensor [*Sechservektor*] concept proves unessential. At the moment I am occupied with studying your paper.[184]

Lorentz's paper on generally-covariant Maxwell theory introduced a generic Hamiltonian principle without deriving Einstein's specific choice from it.[185] In a first reaction to this paper, Einstein attempted to convince Lorentz that the theory of invariants actually leads to such a specific choice. Although the letter in which Einstein expressed this conviction is not preserved, this much can be concluded from a subsequent letter in which Einstein revoked his claim:

> Subsequent reflections on the last letter I sent you have revealed that I made erroneous assertions in that letter. In actual fact the invariant theory method does not yield more than Hamilton's principle when determining your function $Q(= H\sqrt{-g})$.[186]

Evidently, it was the thorough comparison with Lorentz's approach that directed Einstein's attention to a flaw in his derivation of the *Entwurf* field equations.

On reexamining his 1914 derivation, Einstein found that the condition of linearity, which had apparently entered his argument as an unproblematic default setting, was less innocent than it first appeared to him:

> The reason why I did not notice this last year is that on p. 1069 of my article I had frivolously introduced the condition that H was invariant against linear transformation.[187]

As we discussed earlier, it was by introducing this condition that Einstein had found the condition $B_\mu = 0$ for the choice of an adapted coordinate system, while he had derived the condition $\sum_\nu \partial S_\mu^\nu / \partial x_\nu - B_\mu = 0$ as a consequence of energy-momentum conservation—without taking into account the linear covariance of the Lagrangian.

---

184 "Über Ihre Abhandlung habe ich mich sehr gefreut. Ich habe auch einen Beweis für die Gültigkeit des Impuls Energ[ie]satzes des elektromagnet. Feldes mit Berücksichtigung der Gravitation gefunden sowie eine kovariantentheoretisch vereinfachte Darstellung der Vakuumgleichungen, indem sich der Begriff des "dualen" Sechservektors als entbehrlich erweist. Ich bin gerade mit dem Studium Ihrer Arbeit beschäftigt." Einstein to H. A. Lorentz, 23 September 1915, (CPAE 8, Doc. 122).

185 See (Lorentz 1915).

186 "Nachträgliche Überlegungen zu dem letzten Briefe, den ich an Sie richtete, haben gezeigt, dass ich in diesem Briefe Unrichtiges behauptete. Thatsächlich liefert die invariantentheoretische Methode nicht mehr als das Hamilton'sche Prinzip wenn es sich um die Bestimmung der Ihrer Funktion $Q(= H\sqrt{-g})$ handelt." Einstein to H. A. Lorentz, 12 October 1915, (CPAE 8, Doc. 129).

187 "Dass ich dies letztes Jahr nicht merkte liegt daran, dass ich auf Seite 1069 meiner Abhandlung leichtsinnig die Voraussetzung einführte, H sei eine Invariante bezüglich *linearer* Transformationen." (CPAE 8, Doc. 129)

The basic error which Einstein discovered thus consisted in requiring the compatibility of two conditions derived under different assumptions, once with linearity and once without. An acceptable derivation of the *Entwurf* field equation from the Lagrangian formalism had therefore to be based on additional assumptions. In his letter to Lorentz, Einstein reintroduced the correspondence principle in order to justify the selection of the *Entwurf* Lagrangian among the lengthy list of candidates given in his 1914 paper:

> That $Q/\sqrt{-g}$ had been set by me as equal to the fourth expression given there can be justified by the fact that only with this choice does the theory contain Newton's in approximation. That I believed it possible to support this selection on the equation S^{λ}_{μ} was based on error.[188]

Einstein's derivation along a mathematical strategy was thus reduced, in its substance, to that of the original 1913 *Entwurf* paper. He had come back to his starting point—but with one important difference: In spite of the failure of his hope to achieve a derivation essentially from covariant theory only, he had effectively found a derivation in which mathematical principles came first and were supplemented by the physical requirements of energy-momentum conservation and Newtonian limit embodied in the default settings of his field equation. In other words, Einstein had established a derivation which follows precisely the pattern of his attempted derivations along the mathematical strategy in the Zurich Notebook. But instead of taking a generally-covariant object suggested by absolute differential calculus as a starting point, Einstein's point of departure was now the variational calculus he had developed himself.

From Einstein's correspondence it becomes clear that he did not yet consider the state of affairs just described as a reason for abandoning the *Entwurf* theory. There is no trace of this in his letter to Lorentz. Also in a letter written to Zangger a few days later,[189] Einstein treated gravitation as one among several topics, clearly considering his current work on it as business as usual:

> It has unfortunately become clear to me now that the "new stars" have nothing to do with the "lens effect," moreover that, taking into account the stellar densities existing in the sky, the latter must be such an incredibly rare phenomenon that it would probably be futile to expect one of the like.[190]

---

188 "Dass $Q/\sqrt{-g}$ von mir gleich dem vierten der dort angegebenen Ausdrücke gesetzt werde, lässt sich dadurch rechtfertigen, dass die Theorie nur bei dieser Wahl die Newton'sche als Näherung enthält. Dass ich glaube, diese Auswahl auf die Gleichung S^{λ}_{μ} stützen zu können, beruhte auf einem Irrtum." (CPAE 8, Doc. 129)

189 The letter was dated by the editors of (CPAE 8) as 15 October 1915, but since it is explicitly dated only as "Friday" and other indications leave a window between 30 September and 21 October, it may well have been written on 8 October, i.e. before the letter to Lorentz.

190 "Seit ich hier bin, habe ich sehr fest auf meiner Bude gearbeitet. Es ist mir nun leider klar geworden, dass die "neuen Sterne" nichts mit der "Linsenwirkung" zu thun haben, das ferner letztere mit Rücksicht auf die am Himmel vorhandenen Sterndichten ein so ungeheuer seltenes Phänomen sein muss, dass man wohl vergeblich ein solches erwarten würde. Ich schrieb eine ergänzende Arbeit zu meiner letztjährigen Untersuchung über die allgemeine Relativität. Gegenwärtig arbeite ich etwas in Wärmetheorie." Einstein to Heinrich Zangger, 15 October 1915, (CPAE 8, Doc. 130).

Einstein evidently considered what he believed to be the impossibility of relating the observation of nova stars to the gravitational lensing effect he had predicted in 1912[191] to be more important than the problem he had discovered in the mathematical proof of the *Entwurf* theory. Even the paper that he mentions as being supplementary to his 1914 review is probably not a reference to a planned revocation of his proof[192] but to a paper mentioned in his earlier letter to Lorentz on generally-covariant electrodynamics, which was eventually published in 1916 (Einstein 1916b). Thus, also the third flaw which Einstein discovered in the *Entwurf* theory did not immediately lead to its refutation.[193] It nevertheless must have had a subversive effect on his belief in this theory as will be discussed in the next section.

In summary, we have seen that the endeavor to develop a mathematical strategy for deriving the *Entwurf* field equation had to face tensions between the non-covariance of the theory and the properties of a mathematical apparatus naturally tuned to generally-covariant objects. These tensions implied the necessity to repeatedly rework the mathematical analysis of the covariance properties of the theory and the derivation of the field equation based on it; they became particularly evident in the criticism by Abraham and Levi Civita. The latter's critique, however, never questioned the goal of Einstein's derivation, to show the supposed uniqueness of the *Entwurf* field equation. Einstein's exposition of his theory to the Göttingen mathematicians and physicists in a series of six Wolfskehl lectures, delivered in late June and early July 1915 at the invitation of David Hilbert,[194] may also have induced a renewed reflection on the justification of the *Entwurf* theory along the mathematical strategy.

The problematic character of Einstein's derivation of the *Entwurf* field equation from a Lagrangian formalism may have emerged, as we have also seen, in the context of applying the formalism to a different context, that of a generally-covariant Maxwell theory. It was here that the formalism developed specifically for the *Entwurf* theory first proved its greater generality. Against this background Einstein's discovery of a flaw in his derivation could have had a double effect: It pointed out the familiar physical arguments for deriving the *Entwurf* theory, and it suggested taking up once more the mathematical strategy of trying out the newly empowered techniques on different candidates. Initially Einstein chose the first option, but his rederivation of the *Entwurf* theory from a variational formalism plus one extra physical condition may

191 See (Renn and Sauer 2003b).

192 As conjectured in (Janssen 1999, note 51).

193 We leave aside here the question of a possible, direct or indirect, influence by David Hilbert, who had also found a flaw in Einstein's reasoning around this time. It is, however, unclear whether and if so when and what Einstein may have learned about Hilbert's work prior to 7 November 1915 when their extant correspondence on this issue begins. See (Corry 2004, ch. 7) and further references cited therein for a discussion of the interaction between Einstein and Hilbert in this period. See also the discussion below in sec. 7.18.1

194 Fragments of an auditor's notes taken during one of the lectures are published in (CPAE 6, Appendix B). For further historical discussion of Einstein's Göttingen visit in June and July 1915, see (Corry 2004, pp. 320–329).

have made it clear to him how much more general this formalism was, particularly since the conservation principle no longer presented a major obstacle.

7.15 Einstein's November Revolution: the Restoration of an Old Candidate

7.15.1 Looking Back in Anger and Hope

With his publication of 11 November 1915, submitted on the 4th of November, Einstein made his definite rejection of the *Entwurf* theory public:

> My efforts in recent years were directed toward basing a general theory of relativity, also for nonuniform motion, upon the supposition of relativity. I believed indeed to have found the only law of gravitation that complies with a reasonably formulated solution in a paper that appeared last year in the *Sitzungsberichte*.
> Renewed criticism showed to me that this truth is absolutely impossible to show in the manner suggested. That this seemed to be the case was based upon a misjudgment.[195]

Among the three major flaws he had meanwhile found in the *Entwurf* theory, the Mercury failure, the rotation failure, and the breakdown of its mathematical derivation, the latter was publicly the most visible, documented as it was by Einstein's lengthy 1914 review paper. It was, in any case, the only failure explicitly mentioned in his first 1915 article:

> The postulate of relativity—*as far as I demanded it there*—is always satisfied if the Hamiltonian principle is chosen as a basis. But in reality, it provides no tool to establish the Hamiltonian function H of the gravitational field.[196]

As we have seen, none of the problems of the *Entwurf* theory, taken by themselves or together, resulted in an immediate rejection of this theory. Therefore it is not self-evident that Einstein, in view of these problems, finally decided to give up the *Entwurf* theory, as he pointed out in the 1915 paper:

> For these reasons I lost trust in the field equations I had derived, and instead looked for a way to limit the possibilities in a natural manner.[197]

---

195 "In den letzten Jahren war ich bemüht, auf die Voraussetzung der Relativität auch nicht gleichförmiger Bewegungen eine allgemeine Relativitätstheorie zu gründen. Ich glaubte in der Tat, das einzige Gravitationsgesetz gefunden zu haben, das dem sinngemäß gefaßten, allgemeinen Relativitätspostulate entspricht, und suchte die Notwendigkeit gerade dieser Lösung in einer im vorigen Jahre in diesen Sitzungsberichten erschienenen Arbeit darzutun.
Eine erneute Kritik zeigte mir, daß sich jene Notwendigkeit auf dem dort eingeschlagenen Wege absolut nicht erweisen läßt; daß dies doch der Fall zu sein schien, beruhte auf Irrtum." (Einstein 1915c, 778)

196 "Das Postulat der Relativität, *soweit ich es dort gefordert habe*, ist stets erfüllt, wenn man das Hamiltonsche Prinzip zugrunde legt; es liefert aber in Wahrheit keine Handhabe für eine Ermittelung der Hamiltonschen Funktion H des Gravitationsfeldes." (Einstein 1915c, 778)

197 "Aus diesen Gründen verlor ich das Vertrauen zu den von mir aufgestellten Feldgleichungen vollständig und suchte nach einem Wege, der die Möglichkeiten in einer natürlichen Weise einschränkte." (Einstein 1915c, 778)

This step cannot be exclusively accounted for, we believe, on the basis of the failures of that theory but only becomes plausible in view of the unexploited resources that he still had at his disposal from his earlier work in the Zurich Notebook. This is also suggested by the sentences following immediately in the paper:

> In this pursuit I arrived at the demand of general covariance, a demand from which I parted, though with a heavy heart, three years ago when I worked together with my friend Grossmann. As a matter of fact, we were then quite close to that solution of the problem, which will be given in the following.[198]

In fact, after abandoning the *Entwurf* theory Einstein returned to one of the mathematical objects he had encountered along the mathematical strategy three years ago in the Zurich Notebook, the November tensor (cf. eq. (82)). As the above passage suggests, he did not insist on the specific requirement of general covariance, but merely required a "more general covariance of the field equations." It was thus, above all, a return to the mathematical strategy applied to the absolute differential calculus that marked the turning point of 11 November 1915, rather than a radical break with his earlier experiences concerning the restriction of covariance if it turned out to be necessary. Indeed, Einstein now emphatically embraced the absolute differential calculus:

> Nobody who really grasped it can escape from its charm, because it signifies a real triumph of the general differential calculus as founded by Gauss, Riemann, Christoffel, Ricci, and Levi-Civita.[199]

In the context of our account, the crucial questions for understanding Einstein's shift in late 1915 are:

1. What eventually convinced him to give up the *Entwurf* theory and return to the mathematical strategy applied to the absolute differential calculus?
2. How could *any* of the tensors explored and discarded in the course of Einstein's work on the Zurich Notebook now again represent a resource for a renewed exploration?
3. What made the November tensor particularly suitable for such a renewed exploration? And, finally:
4. How exactly did Einstein find his way back to the November tensor?

The answer to the first question follows from our analysis of the third flaw Einstein discovered in the *Entwurf* theory, the erroneous derivation. His last derivation of the *Entwurf* field equation had effectively reinstalled the mathematical strategy in its

198 "So gelangte ich zu der Forderung einer allgemeineren Kovarianz der Feldgleichungen zurück, von der ich vor drei Jahren, als ich zusammen mit meinem Freunde Grossmann arbeitete, nur mit schwerem Herzen abgegangen war. In der Tat waren wir damals der im nachfolgenden gegebenen Lösung des Problems bereits ganz nahe gekommen." (Einstein 1915c, 778)

199 "Dem Zauber dieser Theorie wird sich kaum jemand entziehen können, der sie wirklich erfaßt hat; sie bedeutet einen wahren Triumph der durch Gauss, Riemann, Christoffel, Ricci und Levi-Civiter begründeten Methode des allgemeinen Differentialkalküls." (Einstein 1915c, 779)

original sense, starting from a generally-covariant object which is then checked and, if necessary, modified according to physical criteria. However, both the starting point and the check list of criteria now looked somewhat different from how they did in 1912–1913. If Einstein's starting point had then been a second-rank tensor representing a candidate for the left-hand side of the field equation, it was now a scalar representing the Lagrangian of the theory. And if the conservation principle was then a criterion that had to be laboriously checked for each single candidate, it was now automatically fulfilled for any candidate fitting into the general framework. The correspondence principle therefore remained the crucial criterion for choosing the right candidate. If Einstein at any point after his letter to Lorentz of 12 October 1915 decided to actually check his claim that the *Entwurf* theory was determined uniquely by this criterion, his search would have been governed by the renewed mathematical strategy. In a word, checking the *Entwurf* theory and pursuing the mathematical strategy simply coincided in the end.

To answer the second question of why, in general, tensors earlier discarded could now be considered worthy of further examination, we must turn once again to the Zurich Notebook. Einstein's examination in the notebook of candidate gravitation tensors extracted from the Riemann tensor was restricted to the weak-field form of the field equation. In the notebook, Einstein mastered energy-momentum conservation only for weak fields—with the exception of the *Entwurf* operator at the end of the notebook. Therefore, all candidates extracted from the Riemann tensor were left only partially explored in the notebook when Einstein decided to move on to the next candidate. This unexplored potential of the candidates encountered along the mathematical strategy was one of the essential reasons why he considered it worthwhile to reexamine them in 1915. Another crucial reason for such progress in a loop—or by reflection—was the fact that Einstein's renewed mathematical strategy now drew on more resources, in particular, the variational calculus as applied for the *Entwurf* theory.

In answering the third question of what made the November tensor particularly suitable for a renewed exploration, we again must look at the Zurich Notebook. There Einstein, with Grossmann's help, had derived the November tensor from the Ricci tensor under the stipulation of unimodular coordinate transformations. Contrary to the Ricci tensor, the November tensor satisfies Einstein's physically motivated criteria because it could be reduced to a form suitable for obtaining the Newtonian limit by assuming a coordinate restriction (the Hertz restriction) that was the same as the restriction required by energy-momentum conservation in the weak field limit (cf. eq. (LXXXIV)). Contrary to the Einstein tensor, the harmonization of these two restrictions for the November tensor did not require a change in the view of how the Newtonian limit was to be achieved. Clearly, the November tensor offered a natural starting point for a renewal of the mathematical strategy, since in its case the check of the weak field equation with regard to the conservation and correspondence principles produced a positive outcome—shadowed only by the restricted covariance properties of the candidate that resulted from the November tensor by imposing the required coordinate restriction. The questions that remained were whether or not this result could be extended to the full field equation as well, and which restrictions of covariance were implied for the full equation by the conservation principle. These questions could now be addressed with the help of an improved mathematical apparatus.

The answer to the fourth question, concerning the actual path that Einstein took in rediscovering the November tensor in the fall of 1915 is suggested by several comments pointing to the crucial role of the default setting for the gravitational field eq. (XXII) which, in hindsight, played the role of a "fateful prejudice" with its substitution by the default setting eq. (XXIII) being the "key to the solution."[200] The path leading from the *Entwurf* field equation to a field equation based on the November tensor can be reconstructed with fair confidence in view of the default setting for the Lagrangian in terms of the field eq. (LXIII). It was, as we have discussed, this default setting, rooted in classical field theory, which had also made the *Entwurf* Lagrangian look particularly promising.

It must have been tempting for Einstein to look for other Lagrangians that could be interpreted in this way as involving a "square" of the gravitational fields, experimenting with the definition of the gravitational field. In fact, the internal logic of the mathematical representation exerted a pressure on Einstein's interpretation, since in that representation the connection coefficients, i.e. the Christoffel symbols, have a central importance in, e.g. the concept of a covariant derivative, the geodesic equations, or the definition of the Riemann and Ricci tensors. This role of the connection coefficients was at odds with the significance that Einstein attached to the coordinate derivatives of the metric. Thus, when Einstein took the mathematical tradition more seriously again, the mathematical knowledge that was accumulated in the representation, forced him to reconsider his physical prejudices. And it was indeed not far fetched to reinterpret the Christoffel symbol as representing the field, a choice that almost immediately leads to the Lagrangian of the November theory. Einstein's earlier experience, documented in the Zurich Notebook, might have helped find this path from the *Entwurf* to the November theory because at that time he had already explored the relation between gravitation tensors expressed by Christoffel symbols and their expression in terms of the derivatives of the metric, e.g. in the context of studying the so-called "theta-restriction."[201]

In summary, in the fall of 1915 Einstein succeeded in combining insights from his earlier mathematical strategy and canonical mathematical knowledge with the achievements of the physically motivated *Entwurf* theory. He had probably omitted the November tensor from the Zurich Notebook because he lacked the mathematical means to build a full-scale theory around it, in particular, with regard to the implementation of the conservation principle. His unsuccessful attempt at deriving the *Entwurf* theory from a mathematical strategy had laid just those means in his hands. The failure of the mathematical derivation of the *Entwurf* theory left Einstein with a formalism that initially seemed tailor-made for this very purpose, but then turned out to

200 See "Untying the Knot ..." (in vol. 2 of this series).
201 See the "Commentary …" (in vol. 2 of this series).

be much more generally applicable. An attempt to rederive the *Entwurf* field equation within this formalism turned almost automatically into a renewal of his search along the mathematical strategy. Einstein's physical expectations, not only of the correspondence and the conservation principles, but also of the role of the gravitational field in the Lagrangian and in the equation of motion, must have quickly led him to identify the November tensor as the most appropriate candidate, which not only was probably the easiest to handle given Einstein's propensity for unimodular coordinates but for which the implications of the mathematical and the physical strategies seemed to coincide.

7.15.2 Removing an Old Stumbling Block and Encountering a New One: The Conservation Principle in 1915

The demonstration of energy-momentum conservation for a theory based on the November tensor and the representation of the gravitational field by the Christoffel symbols were, for Einstein, the hallmark of the turnaround in November 1915. His contemporary comments and later recollections not only confirm that his earlier rejection of candidates derived from the Riemann tensor was just as much associated with the difficulty in demonstrating the validity of the conservation principle as with difficulties related to the correspondence principle.[202] They also confirm that it was his revision of the understanding of the components of the gravitational field that was a crucial turning point associated with his return to the mathematical strategy in November 1915.[203]

The decisive progress from Einstein's earlier exploration of a theory based on the November tensor was made possible by the Lagrangian formalism that allowed him to demonstrate that the theory complies with the conservation principle. He thus succeeded in removing an old stumbling block that had earlier forced him to abandon the November tensor, as well as expressions based on it, in the Zurich Notebook. However, the problem of establishing the compatibility between conservation and relativity principles, which had also been a problem with the November tensor in 1912–1913, continued to challenge him even in 1915 when it presented new insights as well as a new obstacle.

The most consequential new insight was related to the fact that the coordinate restriction resulting from the conservation principle turned out to have a remarkably simple structure, being reduced to the requirement that a certain scalar function is constant (cf. eq. (83)). Instead of the usual four equations, Einstein merely obtained a single condition from the requirement of energy-momentum conservation. In spite of this simplification, his physical interpretation of this condition did not change; he still saw it as defining adapted coordinates and admissible transformations (Einstein 1915c, 785).

---

202 Einstein to Michele Besso, 10 December 1915, (CPAE 8, Doc. 162) quoted in the introduction.
203 For extensive discussion, see "Untying the Knot …" (in vol. 2 of this series).

Due to the technical novelties of the November theory, Einstein's perspective on the role of coordinate restrictions changed. Now the requirements arising from the conservation principle and those related to the correspondence principle began to play different roles. The first kind of requirements only led to a minimal but still global constraint on the choice of coordinate systems, the second kind of requirements essentially fixes the coordinate system—but now only in the context of a specific physical situation without global implications. In his paper, Einstein for the first time introduced coordinate conditions in this modern sense albeit without any further explanation. He simply made use of the opportunity that the formalism of the November theory had opened up for him.

Considering his earlier failures, the November theory implemented Einstein's heuristic requirements without requiring much of an adjustment of these requirements. What had changed was, as we have seen, the default setting for the representation of the gravitational field. Furthermore, the way in which the conservation principle had earlier fully determined adapted coordinate systems was, as we have also seen, now changed into a weak constraint that could be harmonized with the requirements of the correspondence principle, thus giving rise to the idea of coordinate conditions in the modern sense. On the other hand, what had not changed was the view of the conservation principle as imposing additional conditions on the choice of coordinates. Finally, the way in which the Newtonian limit is attained in the November theory, that is, via a weak field equation of the form of eq. (33), also remained, by default, the same. That Einstein, at the time, did not regard the November theory as a first step towards a more complete theory—as it must appear to a modern reader. This is evident by a letter he wrote to his son on the day he submitted his first November paper, the 4th of November 1915:

> In the last few days I completed one of the finest papers of my life; when you are older I'll tell you about it.[204]

The formalism of the November theory generated one, apparently minor novelty that could *not* easily be assimilated to Einstein's expectations and that therefore called for a physical interpretation. This new stumbling block was the scalar condition for the choice of adapted coordinates eq. (84). In view of its derivation from the conservation principle, it could not have been surprising to Einstein that this condition determines the choice of adapted coordinates by the properties of the stress-energy tensor of matter.[205] But this general argument does little to make the precise way of this determination plausible, let alone make it understandable that coordinate systems for which $\sqrt{-g} = 1$ are to be excluded, as is implied by Einstein's condition. This was a point where the new formalism of the November theory confronted him with the challenge

204 "Dieser Tage habe ich eine der schönsten Arbeiten meines Lebens fertig gestellt; wenn Du einmal grösser bist, erzähle ich Dir davon." Einstein to Hans Albert Einstein, 4 November 1915, (CPAE 8, Doc. 134).

205 The very existence of an additional coordinate restriction must have been a puzzle in view of Einstein's earlier insights into the relation between covariance and conservation.

to find an adequate physical interpretation. Einstein did not hesitate to take up this challenge and, less than a week after the submission of his first November paper, submitted a short addendum dedicated to the physical interpretation of this condition.

7.16 A Familiar Candidate in a New Context: Einstein's Return to the Ricci Tensor

Einstein's addendum to his first November paper was submitted on 11 November and published on 18 November 1915 (Einstein 1915d). It does not contain a single novel formula with respect to the earlier paper but merely constitutes a reinterpretation of what had been achieved. Yet, it introduced a new, now generally-covariant field equation, replacing that of the November theory, which was covariant only for unimodular transformations. The new field equation is instead based on the Ricci tensor, a candidate that Einstein had also considered earlier while working on the Zurich Notebook (cf. eq. (55)). How did he reinterpret his earlier results?

The point of departure for this reinterpretation was the scalar condition for the choice of adapted coordinates eq. (84). This equation, together with the requirement of unimodularity, were the only obstacles, it appeared, that separated Einstein from the realization of general covariance. Furthermore, there was no general physical interpretation for these two requirements: Why should coordinate transformations be unimodular and why should it nevertheless be impossible to select a coordinate system so that $\sqrt{-g} = 1$? These must have been questions motivating Einstein's further search, beyond what he had achieved in his first November paper. In the new approach presented in the addendum, these two questions were answered in the context of an issue that at first glance appears to be unrelated to gravitation theory; the question of the fundamental constitution of matter.

In the introductory part of his addendum, Einstein discussed a contradiction arising in an electromagnetic theory of matter. He argued that the inclusion of gravitation in the energy-momentum balance could resolve, at least in principle, the following contradiction: The hypothesis that all matter is of electromagnetic origin, and Maxwell's equations imply that the trace of the energy-momentum tensor vanishes:[206]

$$\sum T^{\mu}_{\mu el} = 0. \tag{88}$$

It is also clear that for the default setting of the source-term, i.e. pressureless dust (cf. eq. (4)), the trace of the energy-momentum tensor does not vanish. The conflict between this implication and eq. (88) seems to indicate that matter if conceived of as pressureless dust cannot be constructed on an electromagnetic basis.

However, it is possible to conceive the energy-momentum tensor as being composed of two parts, as is suggested by the parallelism of the energy-momentum of

206 Cf. (Laue 1911, § 13).

matter and of the gravitational field on right-hand side of the field equation (cf. eq. (XLI)):

$$\sum_{\mu} T^{\mu}_{\mu} = \sum_{\mu} (T^{\mu}_{\mu el} + t^{\mu}_{\mu grav}) = \sum_{\mu} (0 + t^{\mu}_{\mu grav}), \tag{89}$$

where $T^{\mu}_{\mu el}$ is due to the electromagnetic origin of matter and $t^{\mu}_{\mu grav}$ to gravitational fields, which are now assumed to play a role in the constitution of matter as well. It follows that the non-vanishing trace of the energy-momentum tensor for matter no longer necessarily contradicts with eq. (88) since it seems possible that the vanishing of $T^{\mu}_{\mu el}$ is compensated by $t^{\mu}_{\mu grav}$. In other words, the additional assumption that gravitational fields play a role in the constitution of matter might be considered as hinting at the solution of a puzzle in a purely electromagnetic theory of matter (Einstein 1915d, 800).[207]

The discussion of an electromagnetic theory of matter in the introductory part of Einstein's addendum raises the obvious question of its function in his theory of gravitation. In his introductory paragraph he stakes the following claim:

> In a recent investigation I have shown how Riemann's theory of covariants in multidimensional manifolds can be utilized as a basis for a theory of the gravitational field. I now want to show here that an even more concise and logical structure of the theory can be achieved by introducing an admittedly bold additional hypothesis on the structure of matter.[208]

The hypothesis of an electromagnetic constitution of matter on the basis of "a theory more complete than Maxwell's theory" allowed Einstein to invalidate an important implication regarding the source term of the gravitational field equation—its non-vanishing trace. In his first November paper, this default assumption had forced him to introduce the coordinate condition that excluded coordinate systems with $\sqrt{-g} = 1$. Even earlier, in the Zurich Notebook, he had discarded the Ricci tensor as a candidate for the left-hand side of a gravitational field equation (if only on the level of linear approximation) because this candidate implies the vanishing of the trace of the stress-energy tensor in contrast to the default properties of Einstein's standard model of matter (cf. eq. (LXXV)). The hypothesis of an electromagnetic origin of

207 Apart from the fact that Einstein does not elaborate on the question as to how matter might be conceived of on the basis of "einer gegenueber Maxwells Theorie vervollstaendigten Elektrodynamik" (p. 800), Einstein's suggestion suffers, however, from a rather conspicuous difficulty: Since t^{μ}_{ν} is no tensor but a coordinate-dependent expression, it can in fact not replace the stress-energy tensor of matter. In particular, the claim that the coordinate-dependent expression t^{μ}_{μ} is positive remains unproven in Einstein's paper and can be refuted in a rather simple way by considering a coordinate system in which this quantity vanishes as well. See (Earman and Glymour 1978, 298).

208 "In einer neulich erschienenen Untersuchung habe ich gezeigt, wie auf Riemanns Kovariantentheorie mehrdimensionaler Mannigfaltigkeiten eine Theorie des Gravitationsfeldes gegründet werden kann. Hier soll nun dargetan werden, daß durch Einführung einer allerdings kühnen zusätzlichen Hypothese über die Struktur der Materie ein noch strafferer logischer Aufbau der Theorie erzielt werden kann." (Einstein 1915d, 799)

matter made it possible to resolve both these problems at the same time. In fact, under the condition $\sqrt{-g} = 1$, admissible under this hypothesis, the November tensor coincides with the generally-covariant Ricci tensor. Together with the generalized principle of relativity, this mathematical feature led Einstein to choose the Ricci tensor rather than the November tensor as the more appropriate candidate for the left-hand side of a gravitational field equation. With the introduction of the generally-covariant Ricci tensor, the other problem of the November theory—its restriction to unimodular coordinate systems—disappeared.

All aspects of the new Ricci theory are simply straightforward consequences of the November field equation plus the condition $\sqrt{-g} = 1$ which can now be conceived of as a coordinate condition in the sense that its stipulation does not affect the physical validity of the equations. In particular, Einstein did not present a new derivation of the new field equation from a Hamiltonian variation principle, now to be formulated for the Ricci tensor. He did not write down the free field Lagrangian that would produce the Ricci tensor in the field equation. Instead, Einstein still used the technique of reducing his gravitational field equation, using the condition $\sqrt{-g} = 1$, to the November field equation in his conclusive 1915 paper, and even in the 1916 review paper on general relativity. Similarly, in his addendum, he neither provided an independent discussion of energy-momentum conservation nor of the Newtonian limit, but just assumed that everything would carry over unchanged from the November theory. In his paper, he explicitly claimed that the physically relevant relations remain unchanged by the transition from the November to the Ricci theory:

> Based upon this system one can—by retroactive choice of coordinates—return to those laws which I established in my recent paper, and without any actual change in these laws, ...[209]

He emphasized that the only difference was the increased freedom in choosing a coordinate system:

> The only difference in content between the field equations derived from general covariance and those of the recent paper is that the value of $\sqrt{-g} = 1$ could not be prescribed in the latter.[210]

Even more radically, Einstein claimed in his letter to Hilbert of 12 November that his latest modification implied that Riemann's tensor would now directly produce the gravitational equations but would not change the equations of the theory.[211] In short, his new generally-covariant field equation based on the Ricci tensor represented for

209 "Von diesem System aus kann man durch nachträgliche Koordinatenwahl leicht zu dem System von Gesetzmäßigkeiten zurückgelangen, welches ich in meiner letzten Mitteilung aufgestellt habe, und zwar ohne an den Gesetzen tatsächlich etwas zu ändern." (Einstein 1915d, 801)

210 "Der Unterschied zwischen dem Inhalte unserer aus den allgemein kovarianten gewonnenen Feldgleichungen und dem Inhalte der Feldgleichungen unserer letzten Mitteilung liegt nur darin, daß in der letzten Mitteilung der Wert für $\sqrt{-g} = 1$ nicht vorgeschrieben werden konnte." (Einstein 1915d, 801)

211 Einstein to David Hilbert, 12 November 1915, (CPAE 8, Doc. 139).

Einstein largely a reinterpretation of his earlier results from the November theory.[212] Only the condition $\sqrt{-g} = 1$ had changed its status from being an excluded special case to a key relation for translating results from the older theory into the new one. What had also changed was the physical interpretation of the theory, in particular with regard to its implications for physics outside of gravitation theory.

Concerning Einstein's new gravitation theory, the only significant property of his new model of matter that replaces his standard instantiation of the slot for the source term of pressureless dust is the vanishing of the trace of the stress-energy tensor. All other aspects of such a theory of matter were irrelevant. As we have seen, he had merely two tenuous arguments to support his audacious new approach. According to the first argument, the inclusion of gravitation in an electromagnetic theory of matter could help to avoid the conflict between the vanishing of the trace of the stress-energy tensor for electromagnetic fields and its non-vanishing for matter. This argument was, however, only an unelaborated idea and quite problematic. According to the second argument, an electromagnetic theory of matter was rendered plausible by the greater consistency of the theory of gravitation that it made possible.

Since all essential equations, according to Einstein's assertion, remain the same in the Ricci and November theories, the question of which theory was to be given preference was thus a matter of choice between the following two options:

1. to rely on a standard model of matter and to accept the physically unmotivated restriction of the theory to unimodular transformations and an inexplicable exclusion of certain coordinate systems (November theory);
2. to achieve a generally-covariant theory without special requirements on coordinate systems and with a logically simple structure, but to accept the introduction of non-trivial consequences for a highly problematic fundamental theory of matter (Ricci theory).

Einstein's preference for the second option was affected by the context in which he formulated his new approach, in particular, by the contemporary discussion about an electrodynamic worldview and the parallel work of David Hilbert, which constituted serious competition for Einstein.[213] The context of this discussion lent some credibility to the introduction of speculative assumptions about a fundamental theory of matter.

---

212 In his paper, Einstein did not address the conflict between the Ricci tensor and the correspondence principle, cf. (Stachel 1989; Norton 1984). This conflict was somewhat hidden by the fact that the physical consequences of the Ricci theory were elaborated in terms of the November theory in which the Newtonian limit can be attained via the Hertz condition which is not in conflict with Einstein's default assumption about the static metric. The conflict here arises from to the condition $\sqrt{-g} = 1$, mediating between the two theories and implying, together with the Hertz condition, the harmonic condition. It seems that Einstein was, either at that time or when working on the Zurich Notebook, not aware of this conflict.

213 See (Corry, Renn and Stachel 1997; Sauer 1999).

7.17 The Mercury Problem as a Theoretical Laboratory for the Ricci Tensor

7.17.1 Einstein's Motivation

Only seven days after his last note, on 18 November 1915, Einstein presented an application of his newly found field equation, his "Explanation of the Perihelion Motion of Mercury from the General Theory of Relativity" (*Erklärung der Perihelbewegung des Merkur aus der allgemeinen Relativitätstheorie*) to the Academy. It was the only one of his November papers which he submitted as a manuscript to the assembly of the Academy accompanied by oral comment, as is documented by the protocols.[214] He saw the excellent agreement between his calculated value of the perihelion shift of Mercury (43") and astronomical observations (45" +/– 5") as a breakthrough for his new theory. Einstein may also have commented publicly on his note because he hoped that his achievement would attract the attention of the astronomers attending his presentation to the Academy, such as Karl Schwarzschild.[215]

Einstein achieved his result in only seven days, a very short time for an involved calculation. David Hilbert showed himself impressed by Einstein's rapid success:

> Many thanks for your postcard and cordial congratulations on conquering perihelion motion. If I could calculate as rapidly as you, in my equations the electron would correspondingly have to capitulate, and simultaneously the hydrogen atom would have to produce its note of apology about why it does not radiate.[216]

Einstein must have been eager to quickly find convincing physical consequences for his new theory for three main reasons:[217]

1. he had produced several candidate theories among which no definite decision had yet been possible; the *Entwurf* theory, the November theory, and the Ricci theory,
2. he was in close competition with Hilbert who had just sent him a manuscript about his own gravitational field theory and had to make an effort in order to secure his priority, and
3. he may have been looking for further confirmation for his bold hypothesis of a combined electromagnetic and gravitational origin of matter, a hypothesis which so far had been based mainly on reasons of internal consistency or on general philosophical arguments.

All three motivations for Einstein's concern with the perihelion problem are well documented by his contemporary correspondence with Hilbert.[218] Einstein consid-

---

214 See (Archiv der Berlin-Brandenburgischen Akademie der Wissenschaften, II–V, vol. 91, sheet 64).

215 See previous note, and for historical discussion, see (Renn, Castagnetti and Damerow 1999).

216 "Vielen Dank fuer Ihre Karte und herzliche Gratulation zu der Ueberwältigung der Perihelbewegung. Wenn ich so rasch rechnen könnte wie Sie, muesste bei meinen Gleichg entsprechend das Elektron kapituliren und zugleich das Wasserstoffatom seinen Entschuldigungszettel aufzeigen, warum es nicht strahlt." David Hilbert to Einstein, 19 November 1915, (CPAE 8, 149).

217 See also the discussion in (Earman and Janssen 1993), on which the following relies.

218 See Einstein to David Hilbert, 12 November 1915, Einstein to David Hilbert, 18 November 1915, and David Hilbert to Einstein, 19 November 1915, (CPAE 8, Docs. 139, 148, 149).

ered the Mercury calculation as a piece of evidence in favor of an electromagnetic theory of matter is also confirmed by the abstract of his paper in the Academy proceedings:

> It is shown that the general theory of relativity explains qualitatively and quantitatively the perihelion motion of Mercury, which was discovered by Leverrier. The hypothesis of the vanishing of the stress-energy tensor of matter is thus confirmed. Furthermore, it is shown that the examination of the bending of light rays in the gravitational field makes it also possible to verify this important hypothesis.[219]

7.17.2 The Advantages of a Second Attempt

What enabled Einstein to check this physical consequence of the anomalous perihelion advance of Mercury on the basis of the new field equation so rapidly was his earlier attempt in 1913, undertaken jointly with Michele Besso, to calculate the perihelion shift for the *Entwurf* theory.[220] This earlier attempt had given him a quantitative result that is too small (18") if compared to the empirical value. But this attempt had given Einstein the tools that could now be applied without any essential modification to the new field equation based on the Ricci tensor. The additional resource which the earlier work had laid in Einstein's hands not only allowed him to achieve quick success by applying his new theory to a challenging problem. Even more remarkably, this application also had a profound repercussion on the applied theory itself. The employment of the Mercury calculation scheme in the context of the Ricci theory effectively changed the heuristic criteria of Einstein's search for the field equation and resulted in a more sophisticated understanding of the correspondence principle.

This far-reaching consequence emerged only after Einstein's theory was explored in greater depth with his new calculation of the Mercury problem. His first two November papers were short and contain hardly any discussion of the physical consequences of the postulated field equations. The addendum of November 11 refers entirely to the considerations, including the Newtonian limit, that are presented in the main paper of November 4. The study of the Mercury problem hence constitutes the first elaboration of the Ricci theory, giving it a justification beyond the field equation and its immediate consequences. This holds even if one takes into account Einstein's earlier consideration of this theory in the Zurich Notebook. But now, in mid-November 1915, a fully fledged calculation scheme permitted the determination of approximate solutions to the gravitational field equation. This calculation scheme was inherited form the earlier calculation of the Mercury problem in the context of the

219 "Es wird gezeigt, daß die allgemeine Relativitätstheorie die von Leverrier entdeckte Perihelbewegung des Merkurs qualitativ und quantitativ erklärt. Dadurch wird die Hypothese vom Verschwinden des Skalars des Energietensors der "Materie" bestätigt. Ferner wird gezeigt, daß die Untersuchung der Lichtstrahlenkrümmung durch das Gravitationsfeld ebenfalls eine Möglichkeit der Prüfung dieser wichtigen Hypothese bietet." (Einstein 1915e)

220 See (CPAE 4, Doc. 14; Earman and Janssen 1993).

Entwurf theory; it had been developed after Einstein's struggle with various candidate field equations in the Zurich Notebook at a time when he believed in the validity of the *Entwurf* theory. This scheme had therefore not yet been applied to different candidate field equations and had thus not have had effect on the balance between Einstein's heuristic criteria. The lasting impact of the Mercury problem on the development of the field equations of general relativity in 1915 was to provide the judgement about candidate field equations with knowledge about the formalism of a gravitational field theory that was essentially independent of the field equations and that had been acquired as early as 1913. The accumulation of this knowledge triggered a process of reflection which guided Einstein to the definite field equation of general relativity.

Einstein's calculation of Mercury's perihelion shift was based on finding an approximate solution to the gravitational field equation by an iterative procedure. To find the solution of first order, Einstein and Besso, in 1913, turned directly to the first-order field equation with which Einstein was familiar from his consideration of the Newtonian limit (cf. eq. (33)), (CPAE 4, 360). The solution to this equation was hence given in terms of the canonical metric for a static field (25) which Einstein used to obtain the Newtonian limit. To obtain the second approximation, Einstein and Besso wrote down the general form of a spherically symmetric metric in Cartesian coordinates in terms of three unknown functions so as to immediately satisfy one of the constraints of the problem with an appropriate ansatz (CPAE 4, 364). These functions were then determined by the iterative procedure, starting from the first approximation.

In his 1915 paper, Einstein no longer proceeded in two separate steps but immediately started from the generic ansatz for a spherically symmetric metric (Einstein 1915b, 833). In 1915, this approach was not only natural but also necessary. It was natural because the procedure Einstein and Besso had constructed in 1913 worked just as well for the first as for the second approximation so that there was really no reason for proceeding in two steps as they had done when first developing their method for calculating the Mercury problem. In 1915 it was necessary to begin right away with the second step since the first step of 1913 no longer worked for the field equation of the Ricci theory. While Einstein's default assumption about the metric for a static field presented no manifest problem in the November theory, as it was compatible with the Hertz condition that served to obtain its Newtonian limit, this default assumption was no longer acceptable in the Ricci theory due to the additional condition $\sqrt{-g} = 1$. The conflict between this condition and Einstein's earlier understanding of the Newtonian limit is also addressed in a letter Einstein wrote to Schwarzschild in early 1916, probably referring to a problem analogous to the conflict represented by eq. (LXXV):

> My comment in this regard in the paper of November 4 no longer applies according to the new determination of $\sqrt{-g} = 1$, as I was already aware. [At this point he added in footnote: The choice of coordinate system according to the condition $\sum \partial g^{\mu\nu}/\partial x_\nu = 0$ is not consistent with $\sqrt{-g} = 1$.] Since then, I have handled Newton's case differently, of course, according to the final theory.[221]

In the context of the Ricci theory, Einstein's generic ansatz for a spherically symmetric metric pointed almost without any further calculation to the existence of non-trivial values for $g_{11}...g_{33}$ in contrast to his default assumption about the metric for static gravitational fields. Einstein considered the difference to his earlier assumption about such a metric a remarkable consequence of the application of his methods for solving the perihelion problem to the new field equation. This is evident from his contemporary correspondence. After the completion of the final version of general relativity, he repeatedly mentioned this fact in letters to Michele Besso. In a letter from 10 December, he remarked:

> You will be surprised by the appearance of the $g_{11}...\dot{g}_{33}$.[222]

A little more than a week later, Einstein returned to this point, again emphasizing the remarkable nature of the deviation from what he expected to be the metric for weak static fields. He was now able to point out how the conflict between this deviation and the Newtonian limit could be avoided:

> Most gratifying is the agreement with perihelion motion and the general covariance; strangest, however, is the circumstance that Newton's theory of the field is incorrect already in the 1st order eq. (appearance of the $g_{11}...g_{33}$). It is just the circumstance that the $g_{11}...g_{33}$ do not appear in first-order approximations of the motion eqs. which determines the simplicity of Newton's theory.[223]

The scheme for calculating a spherically symmetric static metric did not in itself lead to a way in which the deviation from Einstein's standard metric could be reconciled with the correspondence principle. However, it was clear that gravitational fields cannot be observed directly but only via the motion of bodies within these fields—a point stressed in Besso's 1913 memo[224]—so that the equation of motion, at second glance, suggested a natural way out of this dilemma. This second glance showed that in first-order approximation only the g_{44}-component of the metric tensor determines the motion of a material point and that, accordingly, non-trivial values for $g_{11}...g_{33}$

---

221 "Meine diesbezügliche Bemerkung in der Arbeit von 4. November gilt gemäss der neuen Festsetzung $\sqrt{-g} = 1$ nicht mehr, wie mir schon bekannt war. [At this point he added in footnote: Die Wahl des Koordinatensystems gemäß der Bedingung $\sum \partial g^{\mu\nu}/\partial x_\nu = 0$ ist nicht vereinbar mit $\sqrt{-g} = 1$.] Seitdem habe ich ja den Newton'schen Fall nach der endgültigen Theorie ja anders behandelt." Einstein to Karl Schwarzschild, 19 February 1916 (CPAE 8, Doc. 194). For a discussion of the role of coordinate conditions in general relativity, see also Einstein's paper on gravitational waves (Einstein 1916c).

222 "Du wirst über das Auftreten der $g_{11}...g_{33}$ überrascht sein." Einstein to Michele Besso, 10 December 1915, (CPAE 8, Doc. 162).

223 "Das Erfreulichste ist das Stimmen der Perihelbewegung und die allgemeine Kovarianz, das Merkwürdigste aber der Umstand, dass Newtons Theorie des Feldes schon in Gl. 1. Ordnung unrichtig ist (auftreten der $g_{11}...g_{33}$). Nur der Umstand, dass die $g_{11}...g_{33}$ nicht in den ersten Näherungen der Bewegungsgleichungen des Punktes auftreten, bedingt die Einfachheit von Newtons Theorie." Einstein to Michele Besso, 21 December 1915, (CPAE 8, Doc. 168).

224 For a facsimile of the relevant passage, see Fig. 2 on p. 300 of "What Did Einstein Know..." (in vol. 2 of this series) and (Renn 2005a, 128).

do not affect the first-order equation of motion and hence the Newtonian limit of Einstein's theory.

In a third letter to his friend Besso, Einstein once more returned to this point, now in order to explain that the new way of obtaining the Newtonian limit is closely related to the perihelion shift of Mercury and hence to the excellent agreement between theory and experiment.

> The great magnification of the effect against our calculation stems from that, according to the new theory, the $g_{11} \dots g_{33}$'s also appear in the first order and hence contribute to the perihelion motion.[225]

This close connection between the empirical success of the theory and the deviation from the correspondence principle, as originally conceived by Einstein, stabilized the modified understanding of this principle and freed it from the aura of a dubious technical trick.

7.17.3 A New Problem Meets an Old Solution

Einstein repeatedly stressed the fact that only g_{44} matters for the equation of motion, a circumstance that must have seemed a strange but lucky coincidence to him. This solution to the dilemma created by the occurrence of non-trivial diagonal components in the first-order static metric was in itself as little new in 1915 as the dilemma itself. We have seen that in 1912–1913 the harmonically reduced and linearized Einstein tensor had been discarded because it led to a metric for weak static fields with non-trivial diagonal components (cf. eq. (74)). Furthermore, the very ansatz for a spherically symmetric static metric used to treat the Mercury problem pointed to the possibility of such non-trivial components. When Einstein first developed this ansatz in 1913, he did not give this possibility serious consideration because he was convinced of the validity of the *Entwurf* equation which does not give rise to such components.

It is remarkable is that, even though the dilemma of a non-spatially flat static metric was in mid-1913 no longer (and not yet) a real one for Einstein, it was nevertheless at that time already considered and resolved by Besso, and probably also by Einstein. This is documented by a page in the Einstein-Besso manuscript, written by Michele Besso on the back of a letter to Einstein. The page can be dated to June 1913 when both worked together in Zurich (CPAE 4, 392). It is one of a couple of pages on which Besso recapitulated the procedure he and Einstein had applied in order to determine the perihelion shift of Mercury. The purpose of this recapitulation was evidently not only Besso's wish to understand more thoroughly a method that in essence had probably been developed by Einstein, but also his intention to apply this method to more complex cases such as the field of a rotating sun or the inclusion of the sun's pressure in the calculation. Given the reflective character of Besso's notes, we also

225 "Die starke Vergrösserung des Effektes gegenüber unserer Rechnung führt daher, dass gemäss der neuen Theorie auch die $g_{11} \dots g_{33}$). in Grössen erster Ordnung auftreten und so zur Perihelbewegung beitragen." Einstein to Michele Besso, 3 January 1916, (CPAE 8, Doc. 178).

find, along with the recapitulation of essentials, general remarks concerning the nature and plausibility of assumptions which Einstein and Besso had made in the course of their application of the method. It is among such remarks that one finds the brief reflection on the assumption of the spatially flat static metric quoted above, which shows that the assumption of this form of the metric was not an unquestioned prejudice.[226] Besso also considered the more general possibility of a weak-field metric in which components other than the 4–4 one deviated from the Minkowski metric. He came to the same conclusion as Einstein in his perihelion paper of November 1915, namely that, to first order, only the g_{44}-component is relevant for the equation of motion:

> The values so derived and inserted in the equations for the motion of the material point lead to the result that in the latter, [as far as] deviations of the magnitudes g from the relativity scheme [i.e. the Minkowski metric] [are concerned], only the elements g_{44} have any influence.[227]

This observation from June 1913 shows that the possibility of attaining the Newtonian limit also for spatially non-flat static metrics was not new. What was new was the necessity to bring this knowledge to bear on a field equation which seemed to forclude any other way of satisfying the correspondence principle. The novelty on 18 November 1915 was thus the combination of two chunks of knowledge that had been available independently for years, i.e., to base field equations on the Ricci tensor and to attain the Newtonian limit also for spatially non-flat static metrics.[228] In the following, we shall see that this combination triggered a new development that would very soon lead Einstein beyond the Ricci tensor.

7.18 Completing the Circle: Einstein's Return to the Einstein Tensor

7.18.1 Finding the Capstone of General Relativity by Double-Checking a New Theory of Matter

Einstein's completion of general relativity in November 1915 was essentially a solitary phase during which he had little correspondence and no collaboration on this subject, except for the mathematician David Hilbert, with whom Einstein's corresponded on the progress of their respective efforts. Hilbert had a long-standing interest in physics and was especially interested in foundational issues within his program of an axiomatization of the natural sciences.[229] When Gustav Mie published a special

---

226 See (CPAE 4, Doc. 14 [p. 16]). The mention of § 1 is probably a reference to (Einstein and Grossmann 1913).

227 "Die so ermittelten Werte in die Gleichg[ung]en für die Beweg[ung] des Materiellen Punktes eingesetzt, ergeben dass in denselben Abweichngen der Grössen g vom Relativitätsschema nur die Glieder g_{44} von Einfluss sind." (CPAE 4, Doc. 14 [p. 16])

228 For Einstein's continued concern with the problem of the appearance of other components of the metric tensor than g_{44}, see Albert Einstein to Erwin Freundlich, 19 March 1915, (CPAE 8, Doc. 63).

229 See (Corry 2004).

relativistic, electromagnetic theory of matter in 1912, he was particularly intrigued by it, and after Einstein's visit to Göttingen, at Hilbert's invitation, in the summer of 1915, Hilbert engaged in an attempt to find a synthesis between Mie's theory and Einstein's approach to gravitation. In November 1915, he was close to finishing his work and became Einstein's competitor for priority of the field equation of general relativity. The two scientists exchanged criticism and preliminary results, directly and possibly also indirectly via others, so that the question arises of the extent to which their results can be considered independent achievements. A set of proofs of Hilbert's "First Communication on the Foundations of Physics" (Hilbert 1915) rules out the possibility that Einstein took the last and crucial step in completing general relativity from the work of David Hilbert. Since this issue is discussed elsewhere in detail,[230] we limit ourselves here to the analysis of how Einstein completed this last step along the pathways of his own prior research.

Einstein considered the calculation of the perihelion shift of Mercury as the success of a generally-covariant theory of gravitation based on the Ricci tensor, but also as confirming the possibility of a new theory of matter. This is clear from the abstract of his paper quoted above and also from a letter he wrote to his friend Besso:

> In these last months I had great success in my work. *Generally covariant* gravitation equations. *Perihelion motions explained quantitatively.* The role of gravitation in the structure of matter. You will be astonished. I worked horrendously intensely; it is strange that it is sustainable.[231]

But the agreement between theoretical and empirical values for the perihelion shift of Mercury supported the new theory of matter only through the condition $\sqrt{-g} = 1$. The precarious role of this condition for further considerations by Einstein is made evident by a footnote appended to the perihelion paper:

> In a forthcoming communication it will be shown that this hypothesis is unnecessary. It is because such a choice of reference frame is possible that the determinant $|g_{\mu\nu}|$ takes on the value –1. The following investigation is independent of this choice.[232]

Einstein reexamined the connection between this determinant condition and his new theory of matter, which found its essential expression in the vanishing of the trace of the stress-energy tensor of matter. The requirement of the vanishing trace resulted

---

230 See (Corry, Renn, and Stachel 1997; Sauer 1999, 2002, 2005; Corry 2004) and further references cited therein. For a facsimile reproduction of both the proofs and the published version of (Hilbert 1915), see (Renn 2005, 146–173). In this series, the relation between Einstein's and Hilbert's work is further discussed in the section "Including Gravitation in a Unified Theory of Physics" (vol. 4 of this series).

231 "Ich habe mit grossem Erfolg gearbeitet in diesen Monaten. *Allgemein kovariante* Gravitationsgleichungen. *Perihelbewegungen quantitativ erklärt*. Rolle der Gravitation im Bau der Materie. Du wirst staunen. Gearbeitet habe ich schauderhaft angestrengt; sonderbar, dass man es aushält." Einstein to Michele Besso, 17 November 1915, (CPAE 8, Doc. 147).

232 "In einer bald folgenden Mitteilung wird gezeigt werden, daß jene Hypothese entbehrlich ist. Wesentlich ist nur, daß eine solche Wahl des Bezugssystems möglich ist, daß die Determinante $|g_{\mu\nu}|$ den Wert –1 annimmt. Die nachfolgende Untersuchung ist hiervon unabhängig." (Einstein 1915b, 831)

from the comparison of two equations Einstein had derived in his paper of 4 November 1915, one with the help of the energy-momentum balance, the other directly from the field equation. With the help of the trace t of the energy-momentum expression of the gravitational field, these equations can be rewritten as:[233]

$$\sum_{\alpha\beta} \frac{\partial^2 g^{\alpha\beta}}{\partial x_\alpha \partial x_\beta} - \kappa t = 0, \text{ and} \tag{90}$$

$$\sum_{\alpha\beta} \frac{\partial^2 g^{\alpha\beta}}{\partial x_\alpha \partial x_\beta} - \kappa t + \frac{\partial}{\partial x^\alpha}\left(g^{\alpha\beta} \frac{\partial}{\partial x^\beta} \log \sqrt{-g}\right) = \kappa t. \tag{91}$$

Comparing these two requirements Einstein had derived the scalar coordinate restriction of his first November paper, eq. (84). Both the empirical success of his perihelion calculation and the support for his new theory of matter were hinging on this condition. But there was another way of bringing the trace of the full field equation into agreement with eq. (83). Possibly feeling uneasy about the far-reaching consequences that this delicate compatibility argument had to support, Einstein reexamined his earlier reasoning.

From this perspective, the system of equations (90) and (91) provided a representation in which to explore the optimal way of putting together the pieces of his puzzle. This exploration led to yet another modification of the field equation. A reflection on how conditions (90) and (91) had been derived from the November field equation may have sufficed for the identification of an appropriate modification of this field equation by adding a multiple of the trace of the stress-energy tensor of matter to its right-hand side so as to yield instead:[234]

$$\sum_{\alpha\beta} \frac{\partial^2 g^{\alpha\beta}}{\partial x_\alpha \partial x_\beta} - \kappa(t + T) = 0, \text{ and} \tag{92}$$

$$\sum_{\alpha\beta} \frac{\partial^2 g^{\alpha\beta}}{\partial x_\alpha \partial x_\beta} - \kappa(t + T) + \frac{\partial}{\partial x^\alpha}\left(g^{\alpha\beta} \frac{\partial}{\partial x^\beta} \log \sqrt{-g}\right) = 0. \tag{93}$$

This set of equations no longer gives rise to problematic additional conditions.

The compatibility of Einstein's two conditions could thus be achieved without requiring $\sqrt{-g} = 1$ to imply that the trace of the stress-energy tensor must vanish, i.e. without eq. (61). This was the final step by which Einstein arrived at the definitive field equations of general relativity, which were presented in his paper of 25 November 1915 (cf. eqs. (69), (70)).

The modified source term in the new field equation violated the default assumption eq. (XLI) about the right-hand side of Einstein's mental model of a gravitational field equation. But Einstein could accept this violation since the energy-momentum

233 Cf. (Einstein 1915a, eqs. 9 and 10).
234 See "Untying the Knot …" (in vol. 2 of this series), sec. 7, eqs. 85–91.

tensor of matter and the energy-momentum expression of the gravitational field now entered the right-hand side of the field equation in a completely analogous way, (cf. eq. (68) to eq. (81)). It may also have played a role that the step from the Ricci to the Einstein tensor was, after all, not unfamiliar given his earlier experience in the Zurich Notebook. In his paper, Einstein lapidarily noted:

> I now quite recently found that one can get away without this hypothesis about the energy tensor of matter merely by inserting it into the field equations in a slightly different way than is done in my earlier papers.[235]

What had earlier prevented Einstein from accepting the (harmonically reduced and linearized) Ricci and Einstein tensors—his understanding of the correspondence principle—had meanwhile been transformed in the context of the perihelion calculation. The success of his solution to the Mercury problem included a solution to the problem of the Newtonian limit, and this solution now effectively replaced the correspondence principle as a criterion for an acceptable field equation.

In summary, the final phase of Einstein's work in November 1915 was not so much a phase in which new results challenged old prejudices, but rather one of reflection on the knowledge that was already available to him and in which different options were weighed against each other. One of the results of this process of reflection was that there was no support for a new theory of matter as Einstein had believed, possibly following Hilbert, in his addendum of 11 November. In the conclusion of his last November paper Einstein explicitly revoked his earlier claim:

> With this, we have finally completed the general theory of relativity as a logical structure. The postulate of relativity in its most general formulation (which makes spacetime coordinates into physically meaningless parameters) leads with compelling necessity to a very specific theory of gravitation that also explains the movement of the perihelion of Mercury. However, the postulate of general relativity cannot reveal to us anything new and different about the essence of the various processes in nature than what the special theory of relativity taught us already. The opinions I recently voiced here in this regard have been in error. Every physical theory that complies with the special theory of relativity can, by means of the absolute differential calculus, be integrated into the system of general relativity theory—without the latter providing any criteria about the admissibility of such physical theory.[236]

---

235 "Neuerdings finde ich nun, daß man ohne Hypothese über den Energietensor der Materie auskommen kann, wenn man den Energietensor der Materie in etwas anderer Weise in die Feldgleichungen einsetzt, als dies in meinen beiden früheren Mitteilungen geschehen ist." (Einstein 1915a, 844)

236 "Damit, ist endlich die allgemeine Relativitätstheorie als logisches Gebäude abgeschlossen. Das Relativitätspostulat in seiner allgemeinsten Fassung, welches die Raumzeitkoordinaten zu physikalisch bedeutungslosen Parametern macht, führt mit zwingender Notwendigkeit zu einer ganz bestimmten Theorie der Gravitation, welche die Perihelbewegung des Merkur erklärt. Dagegen vermag das allgemeine Relativitätspostulat uns nichts über das Wesen der übrigen Naturvorgänge zu offenbaren, was nicht schon die spezielle Relativitätstheorie gelehrt hätte. Meine in dieser Hinsicht neulich an dieser Stelle geäußerte Meinung war irrtümlich. Jede der speziellen Relativitätstheorie gemäße physikalische Theorie kann vermittels des absoluten Differentialkalküls in das System der allgemeinen Relativitätstheorie eingereiht werden, ohne daß letztere irgendein Kriterium für die Zulässigkeit jener Theorie lieferte." (Einstein 1915a, 847)

7.18.2 Reorganizing the Structure of General Relativity

The further history of general relativity shows that this theory could not yet be considered "logically complete," as Einstein formulated in the last paragraph of his conclusive paper. Even if one disregards later developments such as his modification of the field equations with a cosmological term, fundamental issues such as the status of energy-momentum conservation as an independent postulate of the theory still remained to be clarified. Without this clarification, the theory was initially unconvincing even to those physicists, such as Ehrenfest and Lorentz, who supported Einstein and closely followed his work.

Ehrenfest argued that one can eliminate the stress-energy tensor of matter from the two postulates of the theory, the conservation equation and the field equation, and thus arrive at a new differential equation, which the metric tensor has to satisfy in addition to the field equations. He therefore doubted, apparently following Einstein's earlier line of argumentation, that the new field equation was actually generally covariant. On 1 January 1916 Einstein wrote to Lorentz:

> I am conducting a discussion with Ehrenfest at present essentially on whether the theory really does fulfill the general covariance requirement. He also indicated to me that you had encountered problems or objections to it as well; you would do me a great favor if you were to inform me of them briefly. I have broken in my hobbyhorse so thoroughly that with a short hint I certainly also would notice where the crux of the problem lies.[237]

It was in the exchange with Ehrenfest that Einstein arrived at the conclusion that energy-momentum conservation was not an independent postulate but a consequence of the field equation.[238] The substantial clarification of the conservation principle that Einstein achieved in this debate became a starting point for a rearrangement of the foundational elements of his theory. The first step was taken in a lengthy letter that Einstein wrote to Ehrenfest.[239] In this letter he presented a derivation of the field equation from scratch and showed how energy-momentum conservation can be derived from it. Einstein proceeded in four steps:

1. He first derived the Lagrangian form of the field equation.
2. He next turned to the conservation principle. However, he did not yet derive the conservation of energy and momentum from the field equation. Rather, he assumed an equation that includes an unspecified function that has the form of energy-momentum conservation of matter, as he had postulated it in the earlier

237 "Mit Ehrenfest stehe ich in einer Diskussion im Wesentlichen darüber, ob die Theorie die Forderung der allgemeinen Kovarianz wirklich erfülle. Er deutete mir auch an, dass Sie Schwierigkeiten bezw. Einwendungen gefunden hätten; Sie würden mir große Freude machen, wenn Sie mir dieselben kurz mitteilten. Mein Steckenpferd habe ich so gründlich eingeritten, dass ich gewiss auch nach kurzer Andeutung merke, wo das Wesen der Schwierigkeit liegt." Einstein to H. A. Lorentz, 1 January 1916 (CPAE 8, Doc. 177).

238 Einstein to Paul Ehrenfest, 29 December 1915, Einstein to Paul Ehrenfest, 3 January 1916, Einstein to Paul Ehrenfest, 24 January 1916 (CPAE 8, Docs. 174, 179,185).

239 Einstein to Paul Ehrenfest, 24 January 1916 or later (CPAE 8, Doc. 185).

versions of his theory, and then derived from this equation another equation that has the form of energy-momentum conservation for matter *and* gravitation.

3. In the next step, Einstein wrote the gravitational field equation in terms of mixed tensor densities. He had apparently two reasons for doing so, the first being the possibility of an immediate physical interpretation of the equation in this form. The second reason was the preparation of the fourth and final step of his argument in which the conservation principle is demonstrated.
4. In his last step, Einstein derived energy-momentum conservation with the help of an indirect proof. He showed that one obtains a contradiction with the field equation in the mixed form if one does not assume that the unspecified function in the hypothetical equation for energy-momentum conservation (step 2) vanishes.

Einstein considered this line of argument as a new achievement clarifying the foundations of the theory, as becomes evident from the final passage of his letter:

> You will certainly not encounter any more problems now. Show this thing to Lorentz as well, who also does not yet perceive the need for the structure on the right-hand side of the field equations. I would appreciate it if you would then give these pages back to me, because nowhere else do I have these things so nicely in one place.[240]

Einstein made this new derivation the basis for his exposition in the 1916 review paper (Einstein 1916a), submitted on 20 March 1916.[241]

In the 1916 review, however, Einstein introduced a further rearrangement of the foundational elements of his theory. His main new results were a transformation of the indirect proof of the letter to Ehrenfest into a direct proof of energy-momentum conservation and the establishment of a connection between this derivation and a mathematical theorem by Hilbert, which was later generalized by Emmy Noether. The latter result is particularly important as it amounted, in effect, to a recognition of the contracted Bianchi identities and their role as integrability conditions for the sources of the field equation of general relativity.

In his review Einstein proceeded in six steps. We will briefly review these steps and show how a new deductive structure of general relativity emerged from Einstein's reflection on his discovery process and from the insights obtained in the controversy with Ehrenfest:

1. Einstein first introduced the field equations for the source-free case. In this step he transformed his own pathway from the Ricci to the Einstein tensor into a strategy for justifying the foundations of his theory. He introduced the Ricci equation as the appropriate gravitational field equation for empty space conceiving it as a weakening of an equation based on the Riemann tensor (Einstein 1916a, 803).

240 "Du wirst nun wohl keine Schwierigkeit mehr finden. Zeige die Sache auch Lorentz, der die Notwendigkeit der Struktur der rechten Seite der Feldgleichungen auch noch nicht empfindet. Es wäre mir lieb, wenn Du mir diese Blätter dann wieder zurückgäbest, weil ich die Sachen sonst nirgends so hübsch beisammen habe." (CPAE 8, Doc. 185)

241 For historical discussions of this paper, see (Janssen 2005, Sauer 2005).

2. He then developed the Lagrangian formalism and derived an equation for energy-momentum conservation of the gravitational field alone using the pseudo-tensor for the stress-energy of the gravitational field.
3. Einstein next reformulated the field equation in “mixed” form, including the trace term of the pseudo-tensor that suggested the new default setting eq. (81). The peculiar way in which the matter tensor has to be introduced as the source term of the Einstein field equation was thus prepared.
4. Einstein then supplemented the source-free field equation with this matter term, which was introduced in analogy to the pseudo-tensor for the stress-energy of the gravitational field, and thus arrived at the complete Einstein field equation. With respect to the paper of 25 November 1915, the context for Einstein’s justification of his new field equations had changed: Equations corresponding to eq. (92) and eq. (93) no longer appear in the 1916 review paper since energy-momentum conservation is not introduced as an independent postulate. As a consequence, these equations were no longer available as a justification for the new field equation. Instead, Einstein introduced the requirement that the energy of matter and the energy of gravitation enter the field equation on the same footing as the primary motivation for postulating the particular form of the Einstein field equation (Einstein 1916a, 808). He made it additionally clear that the main justification for his postulated field equation were the deductive consequences following from it.
5. Again in analogy to the source-free case, Einstein next showed that an energy-momentum equation holds for matter and the gravitational field. Previously, the equivalents of this equation in the earlier versions of the theory, going back to and including the *Entwurf* theory, were derived from the field equation, together with the independent postulate of energy-momentum conservation. Einstein had now succeeded in deriving this equation from the field equation alone.
6. In his final step, Einstein shows how his usual equation for the energy-momentum conservation of matter in the presence of a gravitational field, which was represented by the vanishing covariant divergence of the stress-energy tensor of matter, actually follows from his field equation. In other words, what had been a heuristic principle useful for selecting appropriate field equations now became a consequence of the field equation that was useful for selecting an appropriate stress-energy tensor of matter suitable to act as a source of the field equation (Einstein 1916a, 809–810).

In the last step of his deductive construction, Einstein also established a bridge to Hilbert’s contemporary work integrating one of its mathematical corner stones into his own newly established framework of general relativity. As we have just seen, within this framework the stress-energy tensor of matter is no longer conceived as an independent ingredient of the theory with properties that affect its physical interpretation (such as the selection of preferred coordinates) but this tensor has itself to satisfy certain constraints imposed by the theory. In a short remark, Einstein characterized this partly dependent and partly independent status of the material process in his theory of gravitation:

> Thus the field equations of gravitation contain four conditions which govern the course of material phenomena. They give the equations of material phenomena completely, if the latter is capable of being characterized by four differential equations independent of one another.[242]

At this point Einstein appended a footnote in which he referred to Hilbert.[243] Einstein provides here a reinterpretation of the mathematical claim central to Hilbert's theory, which constitutes the core of what later became Noether's theorem.[244] In fact on the page referred to by Einstein we find the following passage:

> ... then in this invariant system of *n* differential equations for the n quantities there are always four that are a consequence of the remaining $n-4$ in this sense, that among the *n* differential equations and their total derivatives there are always four linear and linearly independent combinations that are satisfied identically.[245]

By referring this general theorem to the relation between his gravitational field equation and the four differential equations corresponding to the vanishing of the covariant divergence of the stress-energy tensor, Einstein gave a physical interpretation of this theorem that was quite different from Hilbert's. Combining his own results with those of Hilbert, he was able to understand that energy-momentum conservation follows from the field equations. He had thus finally realized the structural role which the four differential equations, expressing energy-momentum conservation and mathematically corresponding to the contracted Bianchi identities, play for the conservation principle in the general theory of relativity as we understand it today.

8. THE TRANSITION FROM CLASSICAL PHYSICS TO GENERAL RELATIVITY AS A SCIENTIFIC REVOLUTION

In the preceding sections, we have reconstructed the complex process by which Einstein's heuristics led to the formulation of the general theory of relativity. We have shown that a key role was played by the interaction between the heuristics guiding the search for the new theory and the concrete representations of intermediate results in terms of physically interpreted mathematical formalisms. These representations opened up new possibilities for further development and often required adjustments

---

242 "Die Feldgleichungen der Gravitation enthalten also gleichzeitig vier Bedingungen, welchen der materielle Vorgang zu genügen hat. Sie liefern die Gleichungen des materiellen Vorganges vollständig, wenn letzterer durch vier voneinander unabhängige Differentialgleichungen charakterisierbar ist." (Einstein 1916a, 810)

243 Cf. (Hilbert 1915, 3). Einstein's page number actually refers to an offprint of Hilbert's paper, not to the published version. Offprints were available to Hilbert already by mid-February 1916, the published paper itself appeared on 31 March 1916, see (Sauer 1999, note 74).

244 See (Sauer 1999). For the roots of this theorem in Einstein's own work, see sec. 3 of "Untying the Knot ..." (in vol. 2 of this series).

245 "... so sind in diesem invarianten System von *n* Differentialgleichungen für die *n* Größen stets vier eine Folge der n – 4 übrigen – in dem Sinne, daß zwischen den *n* Differentialgleichungen und ihren totalen Ableitungen stets vier lineare, von einander unabhängige Kombinationen identisch erfüllt sind." (Hilbert 1915, 3). See vol. 4 of this series.)

of physical concepts and heuristic principles. Einstein's heuristics, together with such concrete intermediate results, was evidently capable of generating enough of those arguments on which the justification of general relativity as an essential part of modern physics, is still based today.

This heuristics itself and some of Einstein's conceptual starting points in classical physics underwent changes that justify the designation of this process as a scientific revolution. In this final section, we shall first review the beginning and the end of the development of Einstein's heuristics in order to highlight the conceptual innovations brought about by this development with respect to classical physics. We shall then summarize our answers to the three epistemic paradoxes raised by this scientific revolution. These answers make use of the key elements for an understanding of a scientific revolution that is suggested by historical epistemology: the long-term character of knowledge development, the architecture of knowledge, and the mechanisms of knowledge dynamics.

8.1 The Lorentz Model Remodelled

Our analysis has shown that for Einstein's search, the Lorentz model was structurally the most significant heuristic element inherited from classical physics. At each stage of its development, the structure of this mental model and its default settings determined the way in which the specific problems of finding the field equations could be addressed. As long as it remained unquestioned, the model thus opened (or closed) the viable paths of further exploration and determined the possibilities of conceptual unfolding. In classical physics, the two basic structures of the Lorentz model, the field equation and equation of motion, are related to each other as independent components of which the first determines the creation of a global field by a local source, while the second determines the effect of the global field on a local probe. Within the classical framework, source and probe are essentially independent entities entering into this model.

In general relativity, this basic structure has changed. First, the source can no longer be independently prescribed from the field. The distribution of matter and energy acting as a source of the gravitational field can only be described in a given geometry of spacetime, which in turn is only another aspect of the gravitational field determined by the field equation. Second, the equation of motion is no longer an independent aspect of the problem, linked to the description of the gravitational field by an overarching force concept, but is constrained and in special cases even completely determined by the field equation.[246] These features of general relativity, which mark its conceptual distinction from classical physics were not yet evident in 1915 when Einstein formulated his field equations. In other words, the corresponding conceptual innovation was not the presupposition but the result of his research. Thus Einstein's heuristics, which was structured by the Lorentz model, led to the develop-

246 For historical discussion of this point, see (Havas 1989, Kennefick 2005).

ment of a theory whose cognitive content can no longer be adequately captured by this mental model.

The discovery of general relativity would, however, have been impossible if the Lorentz model had not at least been adequate for capturing just those partial aspects of the final theory that made its discovery possible. As we have seen in the previous sections, it was even possible to construct and interpret the definitive field equation of general relativity according to this model. Furthermore, the kind of solutions that Einstein had in mind when he searched for the field equation obscured the new relation between matter distribution and geometry mentioned above. The solutions that he seriously considered were given either by Minkowski spacetime (a vacuum solution) described in various coordinate systems, or weak field solutions that could be obtained from it by an iterative procedure. The problem of having to first specify the geometry and then the distribution of matter and energy in order to solve the field equation turns into an approximation procedure. The further elaboration of the consequences of Einstein's field equation revealed the changes with respect to the Lorentz model. That a revision of this mental model was implied by the field equation of general relativity was clear to Einstein as soon as he noticed that the field equation of the new theory would have to be non-linear. As early as 1912, he interpreted this technical feature as representing the conceptual conclusion that the gravitational field possessing energy must also act as its own source. However, at that time, this modification of the model did not appear to be a radical break, since a modification of only a default setting of the mental model ("linearity of the field equation") was sufficient to account for the insight that gravitation can act as its own source.

8.2 The Ill-Conserved Conservation Principle

In classical physics, the conservation of energy and momentum is a consequence of the fundamental laws governing gravitational and electrodynamic interaction. In special relativity, the conservation principle has found an elegant formulation as a tensorial equation that unifies the conservation of momentum and energy. In both classical and special-relativistic physics, momentum and energy are conceived as localizable physical quantities whose conservation can be described by a partial differential equation which describes a local balance between the various contributions to the energy and momentum of a physical system. Einstein's consideration of a particular example (the behavior of a pressureless dust of particles in a gravitational field) formed the basis, as we have seen, for a tentative generalization of the equation expressing the conservation principle in special relativity, which now also included the effect of gravitation, interpreted as the effect of an external force. Two distinct perspectives on this equation exist, one from classical physics and special relativity, the other from general relativity. The possibility of having these two perspectives on the same mathematical expression turned out to be crucial for the emergence of general relativity.

From the point of view of classical physics and of special relativity, Einstein's postulated equation represented a twofold constraint for the gravitational field equation to be found: the resulting field theory of gravitation had to be compatible with this equation, even at the price of restricting its range of applicability, and, furthermore, the field equation should allow this equation to be rewritten as a local, frame-independent balance between the energy-momentum of matter and that of gravitation.

Further elaboration of the consequences of this equation, however, made this latter request questionable. In the course of his search, Einstein was forced to realize that the expression for energy-momentum conservation which he had postulated turned out to be incompatible with the assumption of a frame-independent stress-energy tensor of gravitation. If this postulate is accepted, then energy and momentum of a gravitational field cannot, in contrast to classical physics, be localizable physical quantities. In this way, a feature of general relativity that is incompatible with classical physics was suggested by a framework still anchored in its fundamental concepts. Einstein's insight into the character of the expression representing the stress-energy of the gravitational field might have given him good reason to abandon this entire approach since its results conflicted with his well-founded expectation that the gravitational field has localizable energetic properties just like all the other known physical fields. Why did he hold on to this equation in spite of its, from the point of view of classical physics, problematic implications? His reasons were in any case not an anticipation of those of the later theory of general relativity.

The equation expressing the energy-momentum balance in a gravitational field that Einstein had postulated at the beginning of his search, and from which the problematic conceptual consequences summarized above can be inferred is obtained in general relativity as an integrability condition of the field equation. Technically speaking, it is a condition to be imposed on an admissible energy momentum tensor, representing the right-hand side of the field equation, required in order to be compatible with a mathematical identity—the contracted Bianchi identity—valid for the left-hand side of the field equation. The Bianchi identity ensures that the gravitational field equation determines the dynamics of the geometry of spacetime without determining also the coordinate system. It reduces the 10 components of the field equation for the 10 components of the metric tensor to only 6 component-equations, thus leaving open the choice of four arbitrary functions corresponding to the choice of a coordinate system. The Bianchi identity together with the gravitational field equation then also determines the evolution of energy and momentum in space and time by way of the equation which Einstein interpreted as the expression for the conservation of energy and momentum in the presence of a gravitational field.

Clearly this argument could not have played a role for Einstein when he was still searching for the correct field equation. He was not even familiar with the Bianchi identity at the time when he concluded his search with the publication of the field equation of general relativity in 1915. Instead he only had two comparatively weak arguments to hold on to this equation even when he recognized that it did not lead him to an invariant local expression for the energy-momentum of the gravitational field. The first argument was that, for the special case of a dust-like cloud of particles, it was possible to obtain this equation from the equation of motion of a single point-particle in a gravitational field described by the metric tensor. The second argument was related to the mathematical form of energy-momentum conservation. The corresponding equation has the form of a generally-covariant divergence equation which is not only the precise analogon for the corresponding special relativistic equation but which also reduces to the latter in the absence of a gravitational field. These two arguments reinforced each other and are in turn supported by other aspects of Einstein's heuristics, in particular by the generalized relativity principle and all those aspects which underlay his understanding of motion in a gravitational field and the requirement of a close correspondence between special relativistic insights and their generalizations in the new theory to be constructed.

But in whatever way Einstein could support his understanding of energy-momentum conservation by drawing on special cases and analogies, it was, from the point of view of the deductive structure of the later theory, support for the wrong side of his argumentative construction, in so far as it stabilized the role of energy-momentum conservation as an independent first principle rooted in the conviction of the fundamental status of energy and momentum conservation for any physical theory. This understanding motivated its use both as a compatibility requirement and as an additional constraint on trial field equations. From the perspective of general relativity, Einstein had thus developed an improper argumentative structure around a proper equation, whereas from his own perspective at the time, he had attained a partial insight into the deductive structure of the theory which he attempted to construct. Only after his achievement of 1915 he was able to reverse this deductive structure and obtain the vanishing of the covariant divergence of the energy-momentum tensor as a consequence of the gravitational field equation in the sense explained above. As was the case for the development of the mental model in Einstein's research, the structural and conceptual insights associated with understanding the role of energy-momentum conservation in general relativity were thus the result and not the presupposition of finding the correct equations.

8.3 The Lack of Correspondence between the Correspondence Principle as seen from Classical Physics and from General Relativity

The way in which the classical theory of gravitation is contained in the theory of general relativity could, of course, not be anticipated on the basis of classical physics before that theory was actually formulated. Nevertheless, the same heuristics which led to the introduction of the principal building block of the new theory, the metric tensor, also determined, to a large extent, Einstein's understanding of the relation between the theory of gravitation which he was looking for and Newton's theory. All in all, he developed, as we have seen, in the course of his research three different arguments in favor of the representation of static gravitational fields by a spatially flat metric tensor in which, for an appropriate coordinate representation, only one component is variable and a function of the three space coordinates; this function then, so Einstein's conclusion, corresponded, under certain limiting conditions, to the Newtonian gravitational potential in his new theory, whatever the precise field equation would be.

The first argument was directly related to the introduction of the metric tensor as representing a gravito-inertial field, a step that was, as we have seen, motivated by the equivalence principle. Einstein conceived Newtonian gravitation and inertia as special cases of a more general interaction. For the case of uniform acceleration he was able to directly identify inertial effects with a scalar Newtonian gravitational field and he expected that he would be able to do the same for more general cases by generalizing the notion of the gravitational field. A model for that generalization was delivered by electrodynamics. In spite of the obvious differences between gravitation theory and electrodynamics, the analogy between them was in fact the only available one and hence determined Einstein's view of the general pattern according to which a theory of the static field should be contained as a special case in a general field theory. According to this pattern, the general potential was represented by a many-component object such as a vector or a tensor which, in the special case of a static field, reduces to a single-component object. In the case of gravitation it should naturally be possible to identify this single-component object with Newton's gravitational potential. This expectation was reinforced by the fact that Einstein had developed, even before introducing a metric formalism, a theory of static gravitational fields in which these are represented by a single function. When he began to employ a metric formalism, it was hence natural to describe static fields by a metric with one variable component and to identify this component with the gravitational potential of his theory of static fields.

Einstein's "classical" understanding of the transition from his general theory to Newton's theory was stabilized by further arguments developed in the course of his research. The second argument was based on the role of special relativity as an intermediate step in this transition. In order to describe the gravitational effects known from classical physics as aspects of a more general gravitational field it is necessary to specify also the conditions under which such an identification is possible. These physical circumstances require, in particular, the general field to be weak and static. These conditions are, however, not sufficient for restricting the realm of gravitational effects to that covered by Newton's theory. The case in which the masses involved perform motions of high velocities requires a treatment by the special theory of relativity. According to this line of reasoning, weak fields, and in particular weak static fields, should hence play the role of an intermediate case in the transition to Newtonian gravitation, an intermediate case to which the special theory of relativity should be applicable. It should hence be possible to formulate a special relativistic gravitational field equation which holds under these circumstances. As it turned out, the solutions of such a weak-field equation, as suggested by the appropriate default-settings of the Lorentz model for this case, exactly correspond to Einstein's classical expectations.

Nevertheless, these expectations were, as we have seen, challenged in the course of Einstein's research which pointed on several occasions towards a representation of static fields by a metric tensor whose form does not correspond to the one which he expected. He therefore felt, at some point, the necessity of developing yet another argument in favor of this expectation. His third argument, which we have also discussed above, was completely independent of a particular gravitational field equation. In essence it consisted in a problematic attempt to deriving the form of the metric tensor for static gravitational fields from the postulate underlying the equivalence principle that all bodies—no matter what their energy content—fall with the same acceleration in a gravitational field.

The assertion that the metric for static fields is of the canonical form expected by Einstein does not belong to the realm of classical physics. It rather appears to be a specific technical assumption which entered his preliminary gravitational theories as an inconspicuous and perhaps precisely for this reason fateful prejudice delaying his progress towards the correct field equation. However, the preceding synopsis of the reasoning by which this assumption was actually anchored in Einstein's thinking shows that, once the metric tensor was introduced as a representation of gravitational fields, the association of static fields with a metric tensor of the canonical form was a necessary consequence of Einstein's understanding of classical physics applied to this representation.

This entire network of reasoning, and in particular, the procedure for attaining the Newtonian limit which forms its core, is not compatible with the final theory of general relativity. According to this theory, static fields are, in general, not represented by a metric tensor of the canonical form. A consistent treatment of the problem of the Newtonian limit in general relativity is an intricate problem[247] and indeed requires a mathematical formalism which did not even exist when Einstein first formulated the theory in 1915; it was only introduced much later by Cartan and others (Cartan 1923, 1924). It is only in this formalism, by using the concept of an affine connection, that it is possible to formulate both general relativity and Newton's theory of gravitation in a way that makes them mathematically comparable.[248] In fact, whereas in general relativity, the geometry of spacetime is described by a metric structure, in Newtonian theory of gravitation, the four-dimensional metric structure is degenerate and only an affine structure can be introduced for spacetime. But since a metric determines also an affine structure, both theories can, with the help of this mathematical concept, actually be expressed in the same mathematical terms. Vice versa, the fact that the spacetime of general relativity carries not only an affine but also a metric structure represents a conceptual leap with respect to Newtonian physics that cannot be bridged by considering the special theory of relativity, which also comprises a metric structure of spacetime, as an intermediate case. It is simply impossible to describe Newtonian gravitational fields by a non-degenerate four-dimensional metric tensor.

---

247 See note 44 above.

248 See "The Story of Newstein …" (in vol. 4 of this series).

This conceptual leap between general relativity and Newtonian physics, together with the strong arguments which Einstein had in favor of his classical conception of the correspondence principle, raise the question as to how he could have ever overcome the crucial hurdle of dealing with the Newtonian limit of his new theory. The surprising solution is that it was in fact not the removal of this major stumbling block which freed his way, but rather its circumvention for the specific problems in the focus of his attention at the time, in particular for the treatment of the motion of a point mass in a gravitational field. By showing that, although the spatial curvature is present even under Newtonian conditions, it remains unobservable if only slowly moving particles are considered, Einstein found a technical loop-hole through which he could escape from his dense network of reasons supporting the canonical form of the metric.

8.4 The Ambiguity of the Equivalence Principle

We have identified the beginning and end points of the development of those aspects of Einstein's heuristics which were obviously rooted in classical physics. We have concluded that this classical heuristics was just sufficient to allow for the formulation of the key equations of general relativity, whose exploration then, however, led to conceptual insights with which the original expectations were no longer compatible. We now turn to those elements of Einstein's heuristics which were peculiar to his specific research strategy, the equivalence principle and the generalized relativity principle. No such principles belonged to the accepted core of classical physics at the time when he took up his research.

If one separates, however, the mathematical development from that of the physical theories, then Einstein's introduction of the principle of equivalence appears to be much less of an idiosyncrasy than it may seem at first sight. To make this clear, consider the reformulation of the classical Newtonian theory of gravitation as a spacetime theory with a non-trivial geometry. This geometry can be described in terms of an affine connection determining the notion of parallel transport of vectors and hence also the geodesic lines in that spacetime. It is a fundamental statement of this reformulation of Newton's theory that the geodesic lines represent the motions of freely falling particles according to the law of gravitation. Remarkably, the equality of gravitational and inertial mass has become, in this modern formulation, an in-built feature rather than a contingent fact as in the traditional formulation. In the formulation of the law of motion as being given by the geodesic lines, the notion of mass appears not at all, while only the notion of gravitational mass enters the field equation of the theory which determines the geometry of spacetime.

From the point of view of this mathematically advanced formulation, Einstein's adoption of the principle of equivalence can hence be recognized as expressing a fundamental feature of Newton's theory of gravitation, shaped, however, by the particular mathematical formulation of the classical theory which formed his starting point and which suggested a sharp conceptual and technical distinction between gravitational and inertial forces. This conclusion indeed frees Einstein's adherence to the principle of equivalence from its idiosyncratic appearance. One might even conjecture that, had the development of the appropriate mathematical tools come a little earlier, others as well might have found it attractive to employ them for a new formalization of the classical knowledge on gravitation, thus arriving at the considerations outlined above even before the advent of general relativity. But it now emerges, on the other hand, even more as a riddle how a principle expressing the knowledge of classical mechanics could have served as a crucial heuristic guidance for overcoming this theory in favor of a theory incompatible with it.

The key to resolving this riddle comes from considering the fact that Einstein used the principle of equivalence not in order to reorganize the knowledge of classical mechanics but the knowledge embodied in both, classical mechanics and the special theory of relativity. His theory of the static gravitational field as well as his early attempts to generalize that theory were nothing but a reinterpretation of the special theory of relativity with the help of the introduction of accelerated frames of reference. His systematic consideration of such accelerated frames induced him to make use of generalized Gaussian coordinates in order to describe the coordinate systems adapted to these frames. It was then a short step for him to consider the metric tensor, coming with the introduction of such coordinates, also as the representation of gravitational effects when these could not be generated by acceleration. In other words, with the introduction of the metric tensor Einstein had found an object that was capable of representing gravitational and inertial effects on the same footing, just as is the affine connection within the modern reformulation of Newton's theory.

It was, however, not a mere coincidence governed by the availability of mathematical methods that Einstein directly attempted to implement the principle of equivalence in a theory that was to generalize special relativity rather than concentrating on a reformulation of classical mechanics. He was aiming from the beginning at a new theory of gravitation which was to comprise both the knowledge on gravitation and inertia represented by classical mechanics and the knowledge on the structure of space and time embodied by special relativity. Effectively, the principle of equivalence acted, according to this reconstruction, as a demand for integrating the knowledge on gravitation and inertia from classical mechanics, which in a modern formulation can be expressed by means of an affine connection, with the knowledge on the metric structure of spacetime from special relativity. It thus acted as a particular instance of Einstein's general strategy to exploit the entire range of classical and special-relativistic physics for constructing his new theory of gravitation. The analysis given here does, however, not square with Einstein's own interpretation of the principle of equivalence as guiding the development of classical and special-relativistic physics with its privileged systems of reference towards a theory of gravitation which would have to encompass also a generalized principle of relativity.

8.5 The Relativity Principle Relativized

Einstein's view that it made sense to search for a generalization of the relativity principle of classical mechanics and special relativity was, as we have seen, based on his acceptance of a philosophical critique of classical mechanics raised by Mach and others. According to this critique, the justification of the privileged role of inertial frames of reference by the notion of absolute space was problematic, while the inertial forces experienced in accelerated frames of reference require an explanation in terms of the interaction between physical masses. Such an explanation would then eliminate any need for absolute space as a causal agent in the analysis of motion. The generalized relativity principle would go, so at least was Einstein's expectation, a long way, and might actually go all the way, towards an implementation of Mach's critique of classical mechanics in the new theory of gravitation.

The implementation of Mach's critique of classical mechanics by way of the generalized relativity principle in Einstein's new theory of gravitation was, however, rather indirect. Rather than explaining inertial properties directly by a physical interaction of masses, they were described by a gravito-inertial field represented by the metric tensor in a way that in fact depends on the frame of reference. But the gravito-inertial field itself would be determined only by the distribution of masses in the universe via a generally-covariant field equation. It follows that the question of whether or not this approach would lead to an exhaustive explanation of inertial properties by the relative distribution of masses depends on the precise nature of the field equation and its solutions. While Einstein was initially convinced that his theory would fully do justice to the Machian roots of the generalized relativity principle, he felt eventually forced to introduce what he called Mach's principle as a separate and additional criterion to be satisfied by the field equation and its solutions. With the establishment of the General Theory of Relativity in 1915, Einstein succeeded in formulating a theory which implemented the generalized relativity principle in its utmost form, the theory being generally covariant; whether it also satisfied Mach's principle, demanding a complete determination of the gravito-inertial field by the distribution of matter in the universe, remained, on the other hand, a much debated issue for a long time to come.[249]

On closer inspection, however, even Einstein's realization of the generalized relativity principle by his formulation of a generally-covariant theory of gravitation represented a questionable success of this heuristic principle. In fact, not only the general theory of relativity of 1915, but also several other theories of gravitation and in particular also the classical Newtonian theory can be given a generally-covariant formulation. The demand for general covariance has to be considered as nothing but a minimal requirement to be imposed on any sensible physical theory, namely to make assertions about physical processes which do not depend on the specific coordinates used for describing them. But Einstein's generalized relativity principle—together with its broader Machian understanding—effectively corresponded, as we have seen,

249 For extensive discussion, see "The Third Way to General Relativity …" (in vol. 3 of this series).

to further requirements beyond the demand for a generally-covariant formulation of the theory of gravitation. It also comprised the demand for treating inertia and gravitation as aspects of a more general interaction as well as the demand for the absence of any prior geometry of spacetime. The latter requirement excludes, for instance, Nordström's theory as being not compatible with Einstein's heuristics since it assumes the geometry of spacetime a priori to be Minkowskian, up to a conformal factor representing the gravitational potential.

Nevertheless, Einstein's own interpretation of this heuristics can hardly be vindicated by the modern understanding of general relativity. First, the demand for treating inertia and gravitation as aspects of a more general interaction can, as we have seen in our discussion of the equivalence principle, already be fulfilled by classical mechanics in an appropriate reformulation. Second, the general covariance of Einstein's theory does not embody a generalization of the relativity principle from classical mechanics and the special theory of relativity, since, in the modern understanding, relativity principles are represented by the symmetry properties of a theory and not by their behavior under coordinate transformations. Third, "Mach's principle" in the sense of Einstein's demand that the metric structure of space be completely determined by the material masses makes little sense according to the modern understanding of general relativity, since the very notion of material bodies acting as a "source" of the gravitational field that can be prescribed independently from the field has turned out to be problematic.

8.6 The Long-Term Development of Knowledge

In the preceding discussion we have emphasized the differences between Einstein's heuristics and the conceptual consequences of the theory whose development was guided by this heuristics. These differences were the result of a process covering two eras stretching from the beginning of the relativity revolution in 1905 to the present: The first era comprised the elaboration of the foundational equations of the new theory guided by the original heuristics, a process that was essentially complete with Einstein's formulation of general relativity in 1915 and that also included, as we have seen, adjustments of the original heuristics. The second era consists in the exploration of the conceptual consequences of the new theory on the basis of an interpretation of the results achieved in the first era as well as in the course of its further elaboration, a process that has still not come to a hold today. In view of the often striking differences between the modern interpretation of general relativity and Einstein's original motivations for searching for such a theory, it represents a remarkable challenge for the historical reconstruction to explain how these original motivations could have led him to such a definitive formulation of the new theory of gravitation. In the beginning we have formulated this challenge in terms of the three epistemic paradoxes of the emergence of general relativity, the paradox of missing knowledge, the paradox of deceitful heuristics, and the paradox of discontinuous progress.

As our reconstruction has shown, an adequate response to the missing-knowledge paradox can only be found when the long-term development of scientific knowledge is taken into account. This development led, after all, to the emergence of a theory whose understanding of how gravity affects motion in terms of spacetime structure is closer to Aristotle's concept of natural motion than to Newton's explanation in terms of an anthropomorphic force. The knowledge on which the astonishing stability of general relativity is founded was, as we have seen, accumulated long before its creation by centuries of physics, astronomy, and mathematics. Our modern acceptance of general relativity is not only based on experiments or observations related to some of its special predictions but also on the fact that it incorporates the entire Newtonian knowledge on gravitation, including its relation to other physical interactions, that has been accumulated over a long period of time in classical physics and in the special theory of relativity. This knowledge embraces, among other aspects, Newton's law of gravitation including its implications for the conservation of energy and momentum, the relation between gravitation and inertia, the understanding that no physical action can propagate with a speed greater than that of light, which was first achieved by the field theoretic tradition of classical physics and then finally established with the formulation of special relativity, and, more generally, the local properties of space and time, also formulated in special relativity.

After special relativity had elevated the causality requirements implicit in field theory to a universal status, gravitation, traditionally a subject at the core of mechanics, had effectively turned into a borderline problem between mechanics and field theory. As was the case for other borderline problems, its successful solution depended on the shared knowledge resources taken into account. In the case of the creation of special relativity, Einstein's success depended on his combining the heritage of mechanics, embodied in the relativity principle, with the heritage of electrodynamics, embodied in the principle of the constancy of the speed of light. In the case of a relativistic theory of the gravitational field, the combination of the heritage of mechanics represented by the Newtonian theory of the static gravitational field with what was known about dynamic fields from electrodynamics was, however, insufficient to create a new and satisfactory theory—as Einstein's competitors experienced to their chagrin. There was, in particular, no clue to the properties of dynamical gravitational fields so that the challenge to build a relativistic field theory of gravitation was comparable to the development of the entire theory of electromagnetism knowing only Coulomb's law.

It was at this point that Einstein's broad perspective, including the philosophical critique of classical mechanics by Mach, allowed him to muster additional resources from classical physics. Einstein exploited the Machian interpretation of the inertial forces in an accelerated reference frame as being due to the interaction of moving masses in order to fill the above-described gap in a field theory of gravitation. By conceiving the inertial forces in accelerated reference frames, such as Newton's rotating bucket, as embodying dynamical gravitational fields he managed in fact to anticipate essential properties of the relativistic theory of gravitation he was about to construct, in particular the necessity to generalize the spatio-temporal framework of special relativity, which led to the notion of a curved spacetime.

8.7 The Architecture of Knowledge

The answer to the second paradox of how Einstein could have formulated the criteria for a gravitational field equation years before finding the solution comes, as we have seen, from considering the architecture of the shared knowledge resources available to him. These resources were in fact part of a system of knowledge with active components capable of providing heuristic guidance to his research.

The characteristics of Einstein's search have become comprehensible by realizing that it was guided by a qualitative knowledge representation structure inherited from classical physics: the mental model of a field theory as embodied in an exemplary way by Lorentz's electron theory. Einstein's preliminary research on a relativistic theory of gravitation in the years between 1907 and 1912 had established default-settings for two of its terminals; the field-slot (filled by assuming that the gravitational potential is represented by the metric tensor), and the source-slot (filled by the stress-energy tensor of matter as suggested by relativistic continuum mechanics). In the context of his research the differential operator describing how the source generates the field represented an open slot for which Einstein was unable to identify a satisfactory instantiation.

As we have discussed, Einstein's difficulty did not result from the fact that too little was known but rather from the fact that too much knowledge had to be taken into account to formulate a field equation that responded to the understanding of gravitation as a borderline problem of mechanics and field theory. On the one hand, a physically plausible instantiation for the differential operator was suggested by knowledge of the Newtonian static gravitational field as well as of the relation between static and dynamic fields in electrodynamic field theory. Constructed in this way, the new theory would automatically be compatible with Newton's theory, thus fulfilling the correspondence principle. On the other hand, a mathematically plausible way to obtain an instantiation of the differential operator was offered by the knowledge about dynamic fields incorporated in Einstein's equivalence principle, which suggested taking generally-covariant objects such as the Riemann tensor as the starting point. Constructed in this way, the new theory would automatically fulfill the generalized relativity principle. The equivalence principle and the generalized relativity principle had helped, in addition, to reveal just those elements of the traditional knowledge on whose integration the new theory could be based. In the modern formulation, they posed the problem of the compatibility between chronogeometry and gravito-inertial structure. Within the knowledge system of classical physics, the Lorentz model was, furthermore, embedded in a network of relations to other frames and mental models; this network served as a control structure for any acceptable implementation of the model. In particular, the new theory had to satisfy the conservation principle, generalizing similar principles from classical and special-relativistic physics.

In short, Einstein's heuristics was overdetermined by the knowledge available to him, explaining why it was so powerful and yet so fortuitous at the same time. The compatibility of the various requirements it imposed could not be established *a priori* but had to be checked by elaborating a mathematical representation of the Lorentz model, starting from one or the other default setting and shaping it according to the remaining heuristic criteria. Einstein's oscillation between a physical strategy starting from an implementation of the correspondence principle, and a mathematical strategy starting from an implementation of the generalized relativity principle could thus be interpreted as realizing alternative and ultimately converging pathways with which to integrate the knowledge of classical physics.

8.8 Knowledge Dynamics

The third paradox, of discontinuous progress, could only be resolved by taking into account that the development of knowledge does not only consist of enriching a given architecture but also comprises processes of reflection by which this architecture is being transformed. Einstein's learning experience was, in fact, characterized by a bottom-up process that accommodated the higher-order structures at the core of his heuristic principles to the outcome of the experiences he made implementing these principles. The interplay between assimilation and accommodation mediated by the mathematical representation has turned out to be the crucial process determining the knowledge dynamics leading to the creation of general relativity as a non-classical theory. Against this background four stages of Einstein's search for the gravitational field equation could be distinguished.

The *tinkering phase* of fall 1912 is documented in the early pages of Einstein's Zurich Notebook. It is characterized by his unfamiliarity with the mathematical operations suitable for constructing a field equation for the metric tensor. Nevertheless, reflecting on his first attempts to formulate a field equation that satisfied his heuristic principles, Einstein built up higher-order structures operating on a strategic level that would later guide his systematic implementation of these principles, in particular, the physical and the mathematical strategy.

The *systematic searching phase* from late 1912 to early 1913 is also extensively documented by the Zurich Notebook. In this phase Einstein systematically examined candidates according to his heuristic principles alternating between physical and mathematical strategies. Meanwhile, the relative weight of the heuristic principles kept changing with the conservation principle emerging as the principal challenge. Paradoxically, the main result of the pursuit of the mathematical strategy was the derivation of an erroneous theory—the *Entwurf* theory—along the physical strategy.

The *consolidation phase* is documented by Einstein's publications and correspondence between 1913 and mid-1915. During this phase he elaborated the *Entwurf* theory, essentially following his earlier heuristics but now under the perspective of consolidation rather than exploration. Paradoxically, however, the main result of the consolidation period was the creation of the presuppositions for a renewed exploration of candidate field equations. Adapting the mathematical strategy to legitimize the *Entwurf* theory, Einstein found that the resulting mathematical formalism did not single out this theory but reopened the perspective of examining other candidates, removing, in particular, the difficulty of implementing the conservation principle. Because of the extended network of results meanwhile assembled, this reexamination could now take the form of a reflective reorganization of Einstein's earlier achievements.

The *reflection phase*, decisive in resolving the paradox of discontinuous progress, is documented by the dramatic series of four communications Einstein submitted to the Prussian Academy in November 1915. The essence of Einstein's return in the first of these communications to a field equation related to the Riemann tensor consists in reinterpreting results achieved in the context of the *Entwurf* theory. As a consequence, also Einstein's original heuristic principles received a revised physical interpretation. The crucial step of the transition from the *Entwurf* theory, still rooted in classical physics, to the non-classical theory of general relativity was, however, the shift in the physical interpretation of the representation he had unfolded in the preceding years.[250] This transition was a Copernicus process resembling Einstein's reinterpretation of Lorentz's auxiliary variable for local time as the time measured in a moving reference frame. But in passing from the *Entwurf* theory to general relativity, however, Einstein was, in a sense, his "own Lorentz"—hence the more isolated character of the second phase of the relativity revolution. In the case of the transition to general relativity, it was, in particular, the Christoffel symbol, initially only an auxiliary quantity, that assumed a new physical meaning, now representing the gravitational field.

The synthesis represented by general relativity was not without alternatives at the time of its establishment—nor is it today. Some of these alternatives were even distinguished by consequences which could be tested empirically. The observational consequences which distinguish general relativity from its main competitor at the time, Newton's and Nordström's theory of gravitation, were, however, by no means momentous and could have easily gone unnoticed for a long time, or might have remained irrelevant for a decision between alternative theories of gravitation had Einstein's research not drawn attention to them. The contemporary discussion about these alternatives and their elaboration document a process of equilibration between individual perspectives and shared knowledge resources.[251] Even the most ingenious phase of the relativity revolution—the phase of reflection—was, from the point of view of historical epistemology, not the privilege of an outstanding individual, but just one aspect of the transformation of a system of knowledge.

250 See "Untying the Knot ..." (in vol. 2 of this series).
251 See vols. 3 and 4 of this series.

ACKNOWLEDGEMENTS

First drafts of this chapter resulted from a collaboration in the context of the Arbeitsstelle Albert Einstein, funded in the years 1990–1995 by the Berlin Senate, see (Castagnetti et al. 1994). First and above all, we wish to thank Peter Damerow for his inspiration and support. It is through him that we learned about the potential use of concepts from cognitive science for the history of science. We are also grateful to our colleagues John Norton, John Stachel, and especially to Michel Janssen for many discussions and all their unrelenting criticism. The final version owes much to recent joint work with Michel Janssen (see "Untying the Knot ..." and "Commentary ..." (in this volume)), who also co-edited the text. We also wish to thank Diana Buchwald for her support and understanding. Last not least, we wish to express our sincere gratitude to Lindy Divarci for her reliable help with copy-editing this paper.

REFERENCES

Abraham, Max. 1914. "Die neue Mechanik." *Scientia* 15: 8–27.

Bianchi, L. 1910. *Vorlesungen über Differentialgeometrie*. Translated by Max Lukat. Leipzig: B.G.Teubner.

Buchwald, Jed. 1985. *From Maxwell to Microphysics. Aspects of Electromagnetic Theory in the Last Quarter of the Nineteenth Century.* Chicago: University of Chicago Press.

Capria, Marco Mamone. 2005. "General Relativity: Gravitation as Geometry and the Machian Programme." In M.M. Capria (ed.), *Physics Before and After Einstein*. Amsterdam: IOS, 93–128.

Cartan, Elie. 1923. "Sur les variétés à connection affine et la théorie de la relativité généralisée (première partie)." *Ecole Normale Supérieure* (Paris). *Annales* 40: 325–412.

———. 1924. "Sur les variétés à connexion affine et la théorie de la relativité généralisée (première partie, suite)". *Ecole Normale Supérieure* (Paris). *Annales* 41: 1–25.

Castagnetti, Giuseppe, Peter Damerow, Werner Heinrich, Jürgen Renn and Tilman Sauer. 1994. *Wissenschaft zwischen Grundlagenkrise und Politik: Einstein in Berlin*. Arbeitsbericht der Arbeitsstelle Albert Einstein 1991–1993. Max-Planck-Institut für Bildungsforschung.

Cattani, Carlo and Michelangelo De Maria. 1989a. "Max Abraham and the Reception of Relativity in Italy: His 1912 and 1914 Controversies with Einstein." In (Howard and Stachel 1989, 160–174).

———. 1989b. "The 1915 Epistolary Controversy between Einstein and Tullio Levi-Civita." In (Howard and Stachel 1989, 175–200).

Corry, Leo. 2004. *David Hilbert and the Axiomatization of Physics, 1898–1918: From "Grundlagen der Geometrie" to "Grundlagen der Physik."* Dordrecht: Kluwer.

Corry, Leo, Jürgen Renn and John Stachel. 1997. "Belated Decision in the Hilbert-Einstein Priority Dispute." *Science* 278: 1270–1273.

CPAE 2. 1989. John Stachel, David C. Cassidy, Jürgen Renn and Robert Schulmann (eds.), *The Collected Papers of Albert Einstein*. Vol. 2. *The Swiss Years: Writings, 1900–1909*. Princeton: Princeton University Press.

CPAE 3. 1993. Martin J. Klein, A. J. Kox, Jürgen Renn, and Robert Schulmann (eds.), *The Collected Papers of Albert Einstein*. Vol. 3. *The Swiss Years: Writings, 1909–1911*. Princeton: Princeton University Press.

CPAE 4. 1995. Martin J. Klein, A. J. Kox, Jürgen Renn, and Robert Schulmann (eds.), *The Collected Papers of Albert Einstein*. Vol. 4. *The Swiss Years: Writings, 1912–1914*. Princeton: Princeton University Press.

CPAE 5. 1993. Martin J. Klein, A. J. Kox, and Robert Schulmann (eds.), *The Collected Papers of Albert Einstein*. Vol. 5. *The Swiss Years: Correspondence, 1902–1914*. Princeton: Princeton University Press.

CPAE 6. 1996. A. J. Kox, Martin J. Klein, and Robert Schulmann (eds.), *The Collected Papers of Albert Einstein*. Vol. 6. *The Berlin Years: Writings, 1914–1917*. Princeton: Princeton University Press.

CPAE 8. 1998. Robert Schulmann, A. J. Kox, Michel Janssen, and József Illy (eds.), *The Collected Papers of Albert Einstein*. Vol. 8. *The Berlin Years: Correspondence, 1914–1918*. Princeton: Princeton University Press. [Part A: 1914-1917, pp. 1–590. Part B: 1918, pp. 591–1118.]

CPAE 9. 2004. Diana Kormos Buchwald, Robert Schulmann, József Illy, Daniel Kennefick, and Tilman Sauer (eds.), *The Collected Papers of Albert Einstein*. Vol. 9. *The Berlin Years: Correspondence, January 1919–April 1920*. Princeton: Princeton University Press.

Damerow, Peter. 1996. *Abstraction and Representation: Essays on the Cultural Revolution of Thinking*. Dordrecht: Kluwer.

Damerow, Peter, Gideon. Freudenthal, Peter McLaughlin and Jürgen Renn. 2004. *Exploring the Limits of Preclassical Mechanics,* 2nd edition. New York: Springer.

Darrigol, Oliver. 2000. *Electrodynamics from Ampère to Einstein*. Oxford: Oxford University Press.

Davis, R. 1984. *Learning Mathematics: The Cognitive Approach to Mathematics Education*. New Jersey: Ablex Publishing Corporation.

Dongen, Jeroen van. 2002. *Einstein's Unification: General Relativity and the Quest for Mathematical Naturalness*. Dissertation, University of Amsterdam.

———. 2004. "Einstein's methodology, semivectors and the unification of electrons and protons." *Archive for History of Exact Sciences* 58 (2004) 219–254.

Earman, John and Clark Glymour. 1978. "Lost in the Tensors: Einstein's Struggles with Covariance Principles 1912–1916." *Studies in History and Philosophy of Science* 9(4):251–278.

Earman, John and Michel Janssen. 1993. "Einstein's Explanation of the Motion of Mercury's Perihelion." In (Earman et al. 1993, 129–172).

Earman, John, Michel Janssen and John Norton (eds.). 1993. *The Attraction of Gravitation. New Studies in the History of General Relativity.* (*Einstein Studies,* Vol. 5), Boston/Basel/Berlin: Birkhäuser.

Earman, John and John Norton. 1987. "What Price Space-Time Substantivalism? The Hole Argument." *British Journal for the Philosophy of Science* 38: 515–525.

Ehlers, Jürgen 1981. "Über den Newtonschen Grenzwert der Einsteinschen Gravitationstheorie." In J. Nitsch et al. (eds.), *Grundprobleme der modernen Physik*. Mannheim: Bibliographisches Institut, 65–84.

———. 1986. "On Limit Relations Between, and Approximative Explanations of, Physical Theories." In R. Marcus et al. (eds.), *Logic, Methodology, and Philosophy of Science VII*. New York: Elsevier Science Publishers, 387–403.

Einstein, Albert. 1907. "Über das Relativitätsprinzip und die aus demselben gezogenen Folgerungen." *Jahrbuch der Radioaktivität und Elektronik* 4: 411–462 (CPAE 2, Doc. 47).

———. 1911. "Über den Einfluss der Schwerkraft auf die Ausbreitung des Lichtes." *Annalen der Physik* 35: 898–908 (CPAE 3, Doc. 23).

———. 1912a. "Gibt es eine Gravitationswirkung, die der elektrodynamischen Induktionswirkung analog Ist?" *Vierteljahrsschrift für gerichtliche Medizin und öffentliches Sanitätswesen* 44: 37–40 (CPAE 4, Doc. 7).

———. 1912b. "Lichtgeschwindigkeit und Statik des Gravitationsfeldes." *Annalen der Physik* 38: 355–369 (CPAE 4, Doc. 3).

———. 1912c. "Zur Theorie des statischen Gravitationsfeldes." *Annalen der Physik* 38, 443–458 (CPAE 4, Doc. 4).

———. 1913. "Zum gegenwärtigen Stande des Gravitationsproblems." *Physikalische Zeitschrift* 14:1249–1266 (CPAE 4, Doc. 17). (English translation in vol. 3 of this series.)

———. 1914a. "Die formale Grundlage der allgemeinen Relativitätstheorie." *Sitzungsberichte der Preussischen Akademie der Wissenschaften*. 2. Halbband, XLI: 1030–1085 (CPAE 6, Doc. 9).

———. 1914b. "Prinzipielles zur verallgemeinerten Relativitätstheorie und Gravitationstheorie." *Physikalische Zeitschrift* 15: 176–180 (CPAE 4, Doc. 25).

———. 1914c. "Zum Relativitäts-Problem." *Scientia* 15: 337–348 (CPAE 4, Doc. 31).

———. 1915a. "Die Feldgleichungen der Gravitation." *Sitzungsberichte der Preussischen Akademie der Wissenschaften*. 2. Halbband, XLVIII-XLIX: 844–847 (CPAE 6, Doc. 25).

———. 1915b. "Erklärung der Perihelbewegung des Merkur aus der allgemeinen Relativitätstheorie." *Sitzungsberichte der Preussischen Akademie der Wissenschaften*. 2. Halbband, XLVII: 831–839 (CPAE 6, Doc. 24).

———. 1915c. "Zur allgemeinen Relativitätstheorie." *Sitzungsberichte der Preussischen Akademie der Wissenschaften*. 2. Halbband, XLIV: 778–786 (CPAE 6, Doc. 21).

———. 1915d. "Zur allgemeinen Relativitätstheorie (Nachtrag)". *Sitzungsberichte der Preussischen Akademie der Wissenschaften*. 2. Halbband, XLVI: 799–801 (CPAE 6, Doc. 22).

———. 1915e. "Zusammenfassung der Mitteilung 'Erklärung der Perihelbewegung des Merkur aus der allgemeinen Relativitätstheorie.'" *Königlich Preußische Akademie der Wissenschaften (Berlin). Sitzungsberichte* (XLVII):803, (CPAE 6, Doc. 24).

———. 1916a. "Die Grundlage der allgemeinen Relativitätstheorie." *Annalen der Physik* 49, 7: 769–822 (CPAE 6, Doc. 30).

———. 1916b. "Eine neue formal Deutung der Maxwellschen Feldgleichungen der Elektrodynamik." *Königlich Preußische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte*, 184–188 (CPAE 6, Doc. 27).

———. 1916c. "Näherungsweise Integration der Feldgleichungen der Gravitation." *Königlich Preußische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte*: 668–696, (CPAE 6, Doc. 32).

———. 1933. *About the Origins of General Theory of Relativity.* [George A. Gibson Foundation Lecture, delivered at Glasgow, 20 June 1933.] University of Glasgow: MS general 1314, (a typescript amended in Einstein's handwriting).

Einstein, Albert and A. D. Fokker. 1914. "Die Nordströmsche Gravitationstheorie vom Standpunkt des absoluten Differentialkalküls." *Annalen der Physik* 44: 321–328 (CPAE 4, Doc. 28).

Einstein, Albert and Marcel Grossmann. 1913. *Entwurf einer verallgemeinerten Relativitätstheorie und einer Theorie der Gravitation*. Leipzig: Teubner (CPAE 4, Doc. 13).

———. 1914. "Kovarianzeigenschaften der Feldgleichungen der auf die verallgemeinerte Relativitätstheorie gegründeten Gravitationstheorie." *Zeitschrift für Mathematik und Physik* 63: 215–225 (CPAE 6, Doc. 2).

Einstein, A. and L. Infeld. 1938. *The Evolution of Physics: The Growth of Ideas from Early Concepts to Relativity and Quanta*. New York: Simon & Schuster.

Eisenstaedt, Jean and A.J.Kox (eds.). 1992. *Studies in the History of General Relativity.* (*Einstein Studies,* Vol. 3). Boston/Basel/Berlin: Birkhäuser.

Galison, Peter. 2003. *Einstein's Clocks and Poincaré's Maps: Empires of Time*. New York: Norton.

Gentner, Dedre. and A. L. Stevens. 1983. *Mental Models*. Hillsdale, NJ: Erlbaum.

Gray, Jeremy (ed.). 1999. *The Symbolic Universe: Geometry and Physics, 1890–1930*. Oxford: Oxford University Press.

Goenner, Hubert, Jürgen Renn, Jim Ritter, and Tilman Sauer (eds.). 1999. *The Expanding Worlds of General Relativity* (*Einstein Studies*, Vol. 7). Boston: Birkhäuser.

Havas, Peter. 1989. "The Early History of the "Problem of Motion" in General Relativity." In (Howard and Stachel 1989, 234–276).

Hilbert, David. 1915. "Die Grundlagen der Physik. (Erste Mitteilung)." *Nachrichten der Königlichen Gesellschaft zu Göttingen*, 395–407.

Hoffmann, Banesh. 1972. "Einstein and Tensors." *Tensor* 26: 439–162.

Holmes, Frederick L., Jürgen Renn and Hans-Jörg Rheinberger (eds.). 2003. *Reworking the Bench: Research Notebooks in the History of Science*. Dordrecht: Kluwer.

Howard, Don and John D. Norton. 1993. "Out of the Labyrinth? Einstein, Hertz, and the Göttingen Answer to the Hole Argument." In (Earman et al. 1993, 30–62).

Howard, Don, and John Stachel (eds.). 1989. *Einstein and the History of General Relativity* (*Einstein Studies,* Vol. 1). Boston: Birkhäuser.

———. 1989–2005. *Einstein Studies,* 11 vols. Boston/Basel/Berlin: Birkhäuser.

Janssen, Michel. 1995. *A Comparison between Lorentz's Ether Theory and Special Relativity in the Light of the Experiments of Trouton and Noble*. Dissertation, University of Pittsburgh.

———. 1999. "Rotation as the Nemesis of Einstein's *Entwurf* Theory." In (Goenner et al. 1999, 127–157).

———. 2005. "Of Pots and Holes: Einstein's Bumpy Road to General Relativity." *Annalen der Physik* 14: *Supplement* 58–85. Reprinted in J. Renn (ed.) *Einstein's Annalen Papers. The Complete Collection 1901–1922*. Weinheim: Wiley-VCH, 2005.

Janssen, Michel and John Stachel 2004. *The Optics and Electrodynamics of Moving Bodies*. Preprint 265: Max Planck Institute for the History of Science.

Katzir, Shaul. 2005. "Poincaré's Relativistic Theory of Gravitation." In (Kox and Eisenstaedt 2005, 15–38).

Kennefick, Daniel. 2005. "Einstein and the Problem of Motion: A Small Clue." In (Kox and Eisenstaedt 2005, 109–124).

Kox, A.J. and Eisenstaedt, Jean (eds.). 2005. *The Universe of General Relativity.* (*Einstein Studies,* Vol. 11) Boston/Basel/Berlin: Birkhäuser.

Künzle, H.P. 1976. "Covariant Newtonian Limit of Lorentz space-times." *General Relativity and Gravitation* 7(5): 445–457.

Lanczos, C. 1972. "Einstein's Path From Special to General Relativity." In L. O'Raifertaigh (ed.), *General Relativity: Papers in Honour of J. L. Synge*. Oxford: Clarendon Press, 5–19.

Laue, Max von. 1911. *Das Relativitätsprinzip*. Braunschweig: Friedr. Vieweg & Sohn.

Lorentz, Hendrik A. 1915. "Het beginsel van Hamilton in Einstein's theorie der zwaartekracht." *Koninklijke Akademie van Wetenschappen te Amsterdam. Wis- en Natuurkundige Afdeeling. Verslagen van de Gewone Vergaderingen* 23: 1073–1089. Reprinted in translation as "On Hamilton's Principle in Einstein's Theory of Gravitation." *Koninlijke Akademie van Wetenschappen te Amsterdam. Section of Sciences. Proceedings* 19 (1916–1917): 1341–1354.

———. 1895. *Versuch einer Theorie der electrischen [elektrischen] und optischen Erscheinungen in bewegten Körpern*. Leiden: Brill.
Maltese, Giulio. 1991. "The Rejection of the Ricci Tensor in Einstein's First Tensorial Theory of Gravitation." *Archive for History of Exact Sciences* 41(4): 363-–381.
Maltese, Giulio, and Lucia Orlando. 1995. "The Definition of Rigidity in the Special Theory of Relativity and the Genesis of the General Theory of Relativity." *Studies in History and Philosophy of Modern Physics* 26(3): 263–306.
Mehra, Jagdesh. *Einstein, Hilbert, and the Theory of Gravitation. Historical Origins of General Relativity Theory.* Dordrecht: Reidel.
Miller, Arthur I. 1992. "Albert Einstein's 1907 Jahrbuch Paper: The First Step from SRT to GRT." In (Eisenstaedt and Kox 1992, 319–335).
Minkowski, Hermann. 1908. "Die Grundgleichungen für die elektromagnetischen Vorgänge in bewegten Körpern." *Königliche Gesellschaft der Wissenschaften Zu Göttingen, Nachrichten. Mathematisch-physikalische Klasse*, 53–111.
Minsky, Marvin. 1975. "A Framework for Representing Knowledge." In P. H. Winston (ed.), *The Psychology of Computer Vision*. New York: McGraw-Hill, 211–281.
———. 1987. *The Society of Mind*. London: Heinemann.
Norton, John D. 1984. "How Einstein Found his Field Equations, 1912–1915." *Historical Studies in the Physical Sciences* 14: 253–316. Reprinted in (Howard and Stachel 1989, 101–159). Page references are to this reprint.
———. 1985. "What Was Einstein's Principle of Equivalence?" *Studies in History and Philosophy of Science* 16. Reprinted in (Howard and Stachel 1989, 5–47.)
———. 1992a. "Einstein, Nordström and the Early Demise of Scalar, Lorentz Covariant Theories of Gravitation." *Archive for the History of Exact Sciences* (45) 1: 17–94.
———. 1992b. "The Physical Content of General Covariance." In (Eisenstaedt and Kox 1992, 281–315).
———. 1994. "General Covariance and the Foundations of General Relativity: Eight Decades of Dispute." *Reports on Progress in Physics* 56: 791–858.
———. 1999. "Geometries in Collision: Einstein, Klein, and Riemann." In (Gray 1999, 128–144).
———. 2000. "'Nature is the Realisation of the Simplest Conceivable Mathematical Ideas': Einstein and the Canon of Mathematical Simplicity." *Studies in History and Philosophy of Modern Physics* 31: 135–170.
O'Neill, Barrett. 1983. *Semi-Riemannian Geometry with Applications to Relativity.* New York: Academic Press.
Pais, A. 1982. *'Subtle is the Lord ...' The Science and the Life of Albert Einstein*. Oxford: Oxford University Press.
Planck, Max. 1906. "Das Prinzip der Relativität und die Grundgleichungen der Mechanik." *Deutsche Physikalische Gesellschaft. Verhandlungen* 8: 136–141.
———. 1907. "Zur Dynamik bewegter Systeme." *Annalen der Physik* 26: 1–34.
Reich, Karin. 1994. *Die Entwicklung des Tensorkalküls. Vom absoluten Differentialkalkül zur Relativitätstheorie*. Basel: Birkhäuser.
Renn, Jürgen. 1993. "Einstein as a Disciple of Galileo: A Comparative Study of Conceptual Development in Physics." *Science in Context* (special issue) 6: 311–341.
———. 1997. "Einstein's Controversy with Drude and the Origin of Statistical Mechanics in his Atomism: A New Glimpse from the 'Love Letters'." *Archive for History of Exact Sciences* 4: 315–354.
———. 2000. "Mentale Modelle in der Geschichte des Wissens: Auf dem Wege zu einer Paläontologie des mechanischen Denkens." In E. Henning (ed.), *Dahlemer Archivgespräche*. Berlin: Archiv zur Geschichte der Max-Planck-Gesellschaft, 83–100.
———. 2004. "Die klassische Physik vom Kopf auf die Füße gestellt: Wie Einstein die Spezielle Relativitätstheorie fand." *Physik Journal* 3: 49–55.
——— (ed.). 2005a. *Documents of a Life's Pathway.* Weinheim: Wiley-VCH.
———. 2005b. "Before the Riemann Tensor: The Emergence of Einstein's Double Strategy." In (Kox and Eisenstaedt 2005, 53–66).
———. 2005c. "Standing on the Shoulders of a Dwarf: General Relativity—a Triumph of Einstein and Grossmann's Erroneous Entwurf Theory." In (Kox and Eisenstaedt 2005, 39–52).
Renn, Jürgen, Giuseppe Castagnetti and Peter Damerow. 1999. "Albert Einstein: alte und neue Kontexte in Berlin." In J. Kocka (ed.), *Die Königlich Preußische Akademie der Wissenschaften zu Berlin im Kaiserreich*. Berlin: Akademie Verlag, 333–354.
Renn, Jürgen, Peter Damerow and S. Rieger. 2001. "Hunting the White Elephant: When and How did Galileo Discover the Law of Fall? (with an Appendix by Domenico Giulini)." In J. Renn (ed.), *Galileo in Context*. Cambridge: Cambridge University Press, 29–149.

Renn, Jürgen and Tilman Sauer. 1996. "Einstein's Züricher Notizbuch. Die Entdeckung der Feldgleichungen der Gravitation im Jahre 1912." *Physikalische Blätter* 53 (1996) 856–872.

———. 1999. "Heuristics and Mathematical Representation in Einstein's Search for a Gravitational Field Equation." In (Goenner et al. 1999, 87–125).

———. 2003a. "Errors and Insights: Reconstructing the Genesis of General Relativity from Einstein's Zurich Notebook." In (Holmes et al. 2003, 253–268).

———. 2003b. "Eclipses of the Stars - Mandl, Einstein, and the Early History of Gravitational Lensing." In A. Ashtekar et al. (eds.), *Revisting the Foundations of Relativistic Physics - Festschrift in Honour of John Stachel*. Dordrecht: Kluwer Academic Publishers.

Renn, Jürgen, Tilman Sauer and John Stachel. 1997. "The Origin of Gravitational Lensing. A Postscript to Einstein's 1936 Science Paper." *Science* (275) 5297: 184–186.

Ricci, G. and T. Levi-Civita. 1901. "Méthodes de calcul différentiel absolu et leurs applications." *Mathematische Annalen* 54: 125–201.

Salmon, Merrilee H. et al. 1999. *Introduction to the Philosophy of Science: A Text by Members of the Department of the History and Philosophy of Science of the University of Pittsburgh*. Indianapolis: Hackett.

Sauer, Tilman. 1999. "The Relativity of Discovery: Hilbert's First Note on the Foundations of Physics." *Archive for History of Exact Sciences* (53) 529–575.

———. 2002. "Hopes and Disappointments in Hilbert's Axiomatic 'Foundations of Physics." In M. Heidelberger and F. Stadler (eds.), *History of Philosophy and Science: new trends and perspectives*. Dordrecht: Kluwer, 225–237.

———. 2005a. "Albert Einstein, Review Paper on General Relativity Theory." In I. Grattan-Guiness (ed.), *Landmark Writings in Western Mathematics* 1640–1940. Amsterdam: Elsevier, 802–822.

———. 2005b. "Einstein Equations and Hilbert Action. What is Missing on Page 8 of the Proofs for Hilbert's First Communication on the Foundations of Physics?" *Archive for History of Exact Sciences* 59 (2005) 577–590.

———. 2006. "Field Equations in Teleparallel Spacetime. Einstein's *Fernparallelismus* Approach Toward Unified Field Theory." *Historia Mathematica*, in press.

Scheibe, Erhard. 1997. *Die Reduktion physikalischer Theorien. Ein Beitrag zur Einheit der Physik. Teil I: Grundlagen und elementare Theorie*. Berlin: Springer.

———. 1999. *Die Reduktion physikalischer Theorien. Ein Beitrag zur Einheit der Physik. Teil II: Inkommensurabilität und Grenzfalreduktion*. Berlin: Springer.

Sommerfeld, A. 1910a. "Zur Relativitätstheorie. I. Vierdimensionale Vektoralgebra." *Annalen der Physik* 32(9): 749–776.

———. 1910b. "Zur Relativitätstheorie. II. Vierdimensionale Vektoranalysis." *Annalen der Physik* 33(14): 649–689.

Stachel, J. 1980. "Einstein and the Rigidly Rotating Disk." In A. Held (ed.), *General Relativity and Gravitation: A Hundred Years after the Birth of Einstein*. New York: Plenum, 1–15.

———. 1982a. "The Genesis of General Relativity." In H. Nelkowski, H. Poser, R. Schrader and R. Seiler (eds.), *Einstein Symposion Berlin aus Anlass der 100. Wiederkehr seines Geburtstages*. Berlin: Springer, 428–442.

———. 1987b. "How Einstein Discovered General Relativity." In M. A. H. MacCallum (ed.), *General Relativity and Gravitation: Proceedings of the 11th International Conference on General Relativity and Gravitation*. Cambridge: Cambridge University Press, 200–208.

———. 1989a. "The Rigidly Rotating Disk as the 'Missing Link" in the History of General Relativity." In (Howard and Stachel 1989, 48–62).

———. 1989b. "Einstein's Search for General Covariance, 1912–1915." In (Howard and Stachel, 63–100).

———. 1994. "Changes in the Concepts of Space and Time Brought About by Relativity." In C. C. Gould and R. S. Cohen (eds.), *Artifacts, Representations, and Social Practice*. Dordrecht: Kluwer Academic Publishers, 141–162.

———. 1995. "History of Relativity." In L. M. Brown, A. Pais and B. Pippard (eds.), *Twentieth Century Physics*, Vol. 1. Bristol: Institute of Physics Pub., Philadelphia: American Institute of Physics Press, 249–356.

———. 2002. *Einstein from 'B' to 'Z'*. (*Einstein Studies,* Vol. 9) Boston/Basel/Berlin: Birkhäuser.

Vizgin, Vladimir P. 2001. "On the discovery of the gravitational field equations by Einstein and Hilbert: new materials." *Physics-Uspekhi* 44(12): 1283–1298.

Vizgin, Vladimir P. and Y. A. SmorodinskiÎ. 1979. "From the Equivalence Principle to the Equations of Gravitation." *Soviet Physics-USPEKHI* 22 (1979) 489–513.

Walter 1999. "Minkowski, Mathematicians, and the Mathematical Theory of Relativity." In (Goenner et al. 1999, 45–86).

Whittaker, Edmund. 1951. *A History of the Theories of Aether and Electricity. Vol. 1: The Classical Theories*. London: Nelson.
———.1953. *A History of the Theories of Aether and Electricity. Vol. 1: The Modern Theories 1900–1926*. London: Nelson (both volumes reprinted by Dover in 1989).
Wright, J. E. 1908. *Invariants of Quadratic Differential Forms*. Cambridge: Cambridge University Press.

EINSTEIN’S ZURICH NOTEBOOK

Descriptive Note

The document known as the Zurich Notebook is a blue bound notebook, 17.5 x 21.5 cm, containing 45 sheets of squared white paper (not counting two sheets that were torn out, probably by Einstein). Three sheets were left blank. The document thus consists of 84 pages. Two typed notes are taped to the front cover (reproduced on p. 316), probably by Einstein's secretary, Helen Dukas. One says: "Notes for lecture on Relativity probably Zurich." The other describes the back cover (reproduced on p. 317): "s. back of this notebook "Relativitaet" in aE's hand." The first and the last page are glued to the insides of the covers.

The number assigned to the Zurich Notebook in the Einstein Archive is "3–006." This numbering is also reflected in the numbering of the images of the online facsimile reproduction of the document (Einstein Archives Online at http://www.alberteinstein.info). In copies of the Einstein Archive, pairs of facing pages have been numbered "1" through "43." The labels "L" and "R" are used to distinguish pages on the left in such pairs from pages on the right. Einstein used the notebook from both ends. This is reflected in the numbering of the pages. Pages 1L through 31L are numbered from the back of the notebook, pages 32L through 43L are numbered from the front. Between the two parts (i.e., between 31L and 43L) three sheets were left blank. Both between pages 29L and 29R and between pages 38L and 38R a sheet was torn out. Since the part starting from the front contains the earliest notes on gravity in the notebook, we present that part (32L–43L) before the part starting from the back (1L–31L). Although we are concerned only with the material on gravity in these volumes, we present a facsimile reproduction and a transcription of all entries in the notebook. The notebook touches on the following topics:

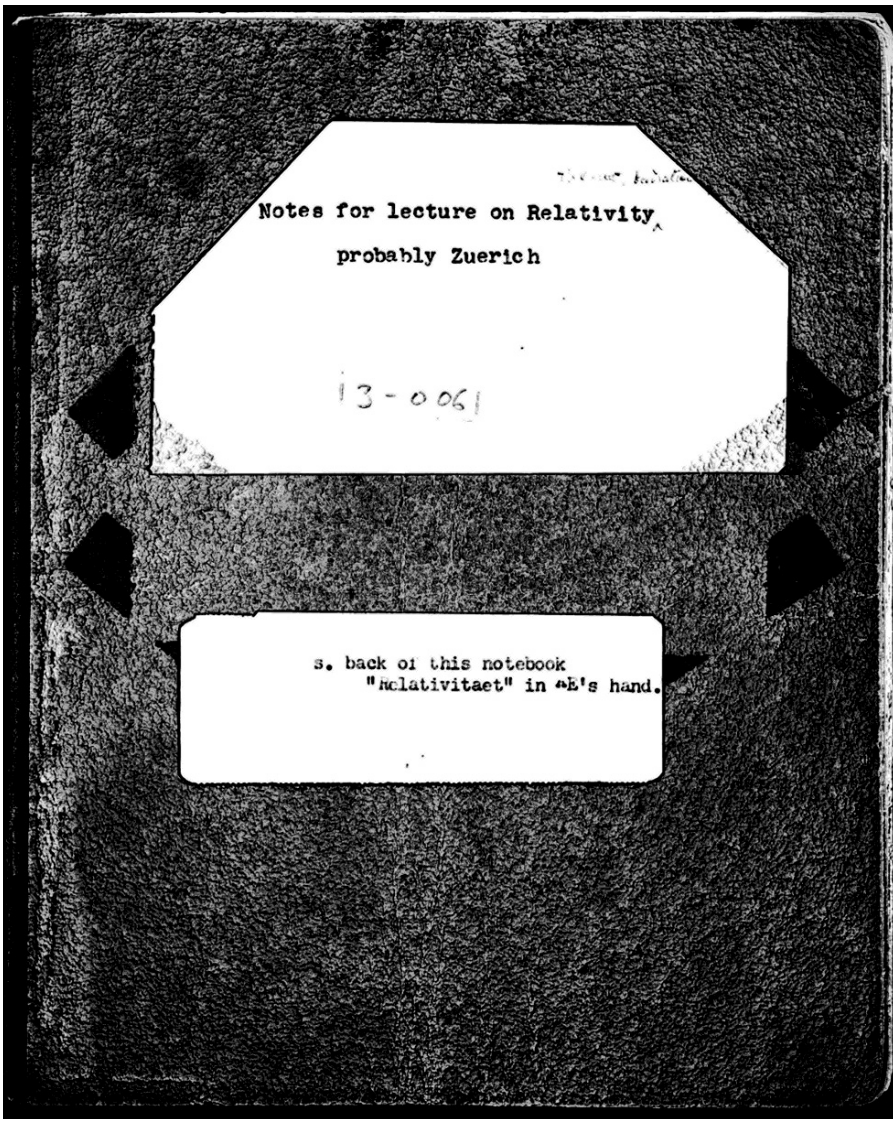
- Electrodynamics (32L–38L)
- Quantum theory (38L–38R)
- Gravity (39L–43L)
  - · First exploration of a metric theory (39L–41R)
  - · Auxiliary calculations (41L–43L)
- Quantum theory (1L–5L)
- Gravity (5R–29L)
  - · Energy-momentum conservation (5R)
  - · Beltrami invariants and core operator (6L–13R)
  - · Riemann tensor, Ricci tensor, and November tensor (14L–25R)
  - · *Entwurf* theory (26L–29L)
- Thermodynamics

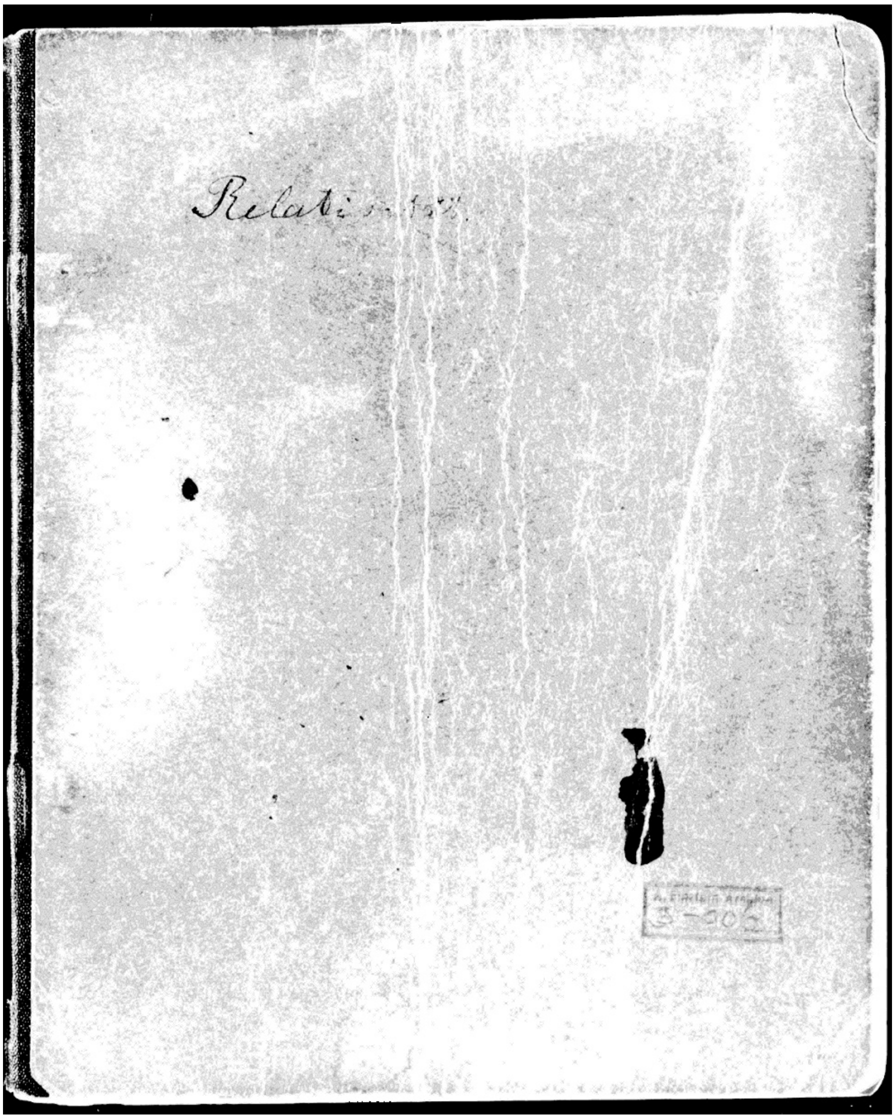
The 57 pages of the notebook dealing with gravity (39L–43L, 5R–29L) were presented in transcription as Doc. 10 in Vol. 4 of *The Collected Papers of Albert Einstein* (CPAE 4). The editors numbered these pages 1 through 58. Page 43L contains material written from the top and material written from the bottom. The former is presented as p. 9 of Doc. 10 in CPAE 4, the latter as p. 58. We likewise present 43L twice, as 43La, with the facsimile the way it appears in the notebook and with a transcription only of the material written from the top, and as 43Lb, with the facsimile upside down and a transcription only of the material written from the bottom.

The authors and editors thank the Albert Einstein Archives at the Hebrew University of Jerusalem and the Einstein Papers Project at the California Institute of Technology for generous permission to reproduce Einstein's Zurich Notebook.

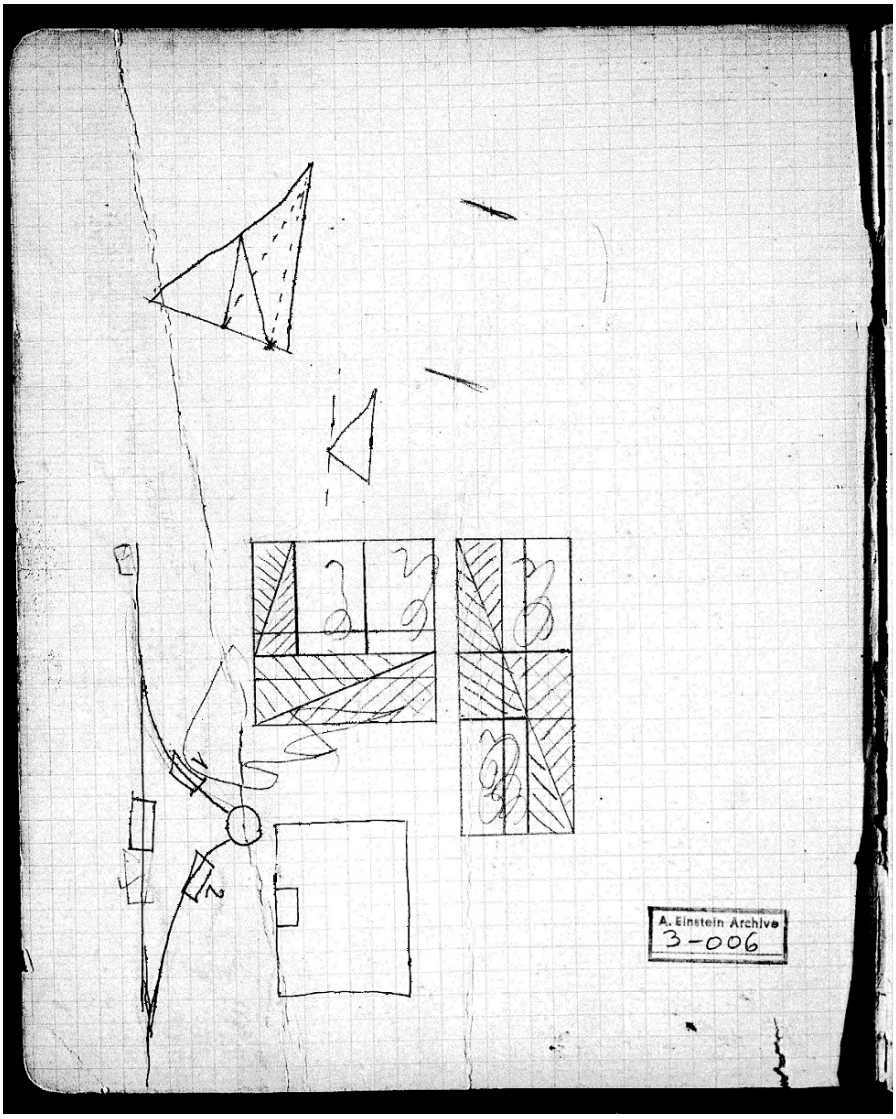
The conversion table below should make it easy to compare our presentation and discussion of the notes on gravity in the notebook to the annotated transcription of this material in CPAE 4.

CPAE 4, Doc. 10	Einstein Archive (3–006)	CPAE 4, Doc. 10	Einstein Archive (3–006)	CPAE 4, Doc. 10	Einstein Archive (3–006)
1	39L	21	11L	41	21L
2	39R	22	11R	42	21R
3	40L	23	12L	43	22L
4	40R	24	12R	44	22R
5	41L	25	13L	45	23L
6	41R	26	13R	46	23R
7	42L	27	14L	47	24L
8	42R	28	14R	48	24R
9	43La	29	15L	49	25L
10	5R	30	15R	50	25R
11	6L	31	16L	51	26L
12	6R	32	16R	52	26R
13	7L	33	17L	53	27L
14	7R	34	17R	54	27R
15	8L	35	18L	55	28L
16	8R	36	18R	56	28R
17	9L	37	19L	57	29L
18	9R	38	19R	58	43Lb
19	10L	39	20L		
20	10R	40	20R		

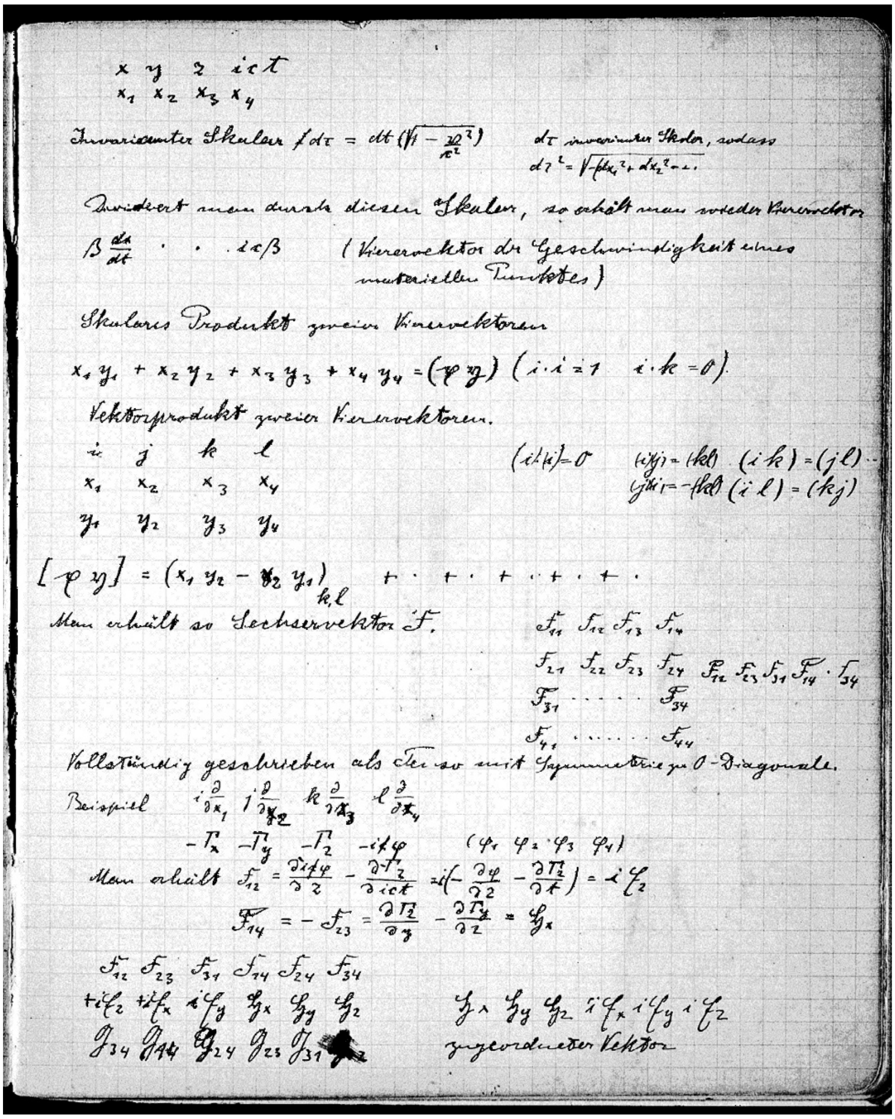

Notes for lecture on Relativity
probably Zuerich
13-0061
s. back of this notebook
"Relativitaet" in E's hand.



[p. 32 L]



[p. 32 R]



$x \quad y \quad z \quad ict$
$x_1 \quad x_2 \quad x_3 \quad x_4$

Invarianter Skalar $f d\tau = dt\left(\sqrt{1-\frac{v^2}{c^2}}\right)$ $d\tau$ invarianter Skalar, sodass $d\tau^2 = \sqrt{-(dx_1^2 + dx_2^2 \ldots}$

Dividiert man durch diesen Skalar, so erhält man wieder Vierervektor

$\beta \frac{dx}{dt} \quad \cdot \quad \cdot \quad . ic\beta$ (Vierervektor der Geschwindigkeit eines materiellen Punktes)

Skalares Produkt zweier Vierervektoren

$x_1 y_1 + x_2 y_2 + x_3 y_3 + x_4 y_4 = (\varphi y) \quad (i \cdot i = 1 \quad i \cdot k = 0)$.

Vektorprodukt zweier Vierervektoren.

$i \quad j \quad k \quad l$ $\qquad (i i) = 0 \quad (i j) = (k l) \quad (i k) = (j l) \cdot$
$x_1 \quad x_2 \quad x_3 \quad x_4$ $\qquad (j i) = -(k l) \quad (i l) = (k j)$
$y_1 \quad y_2 \quad y_3 \quad y_4$

$[\varphi y] = (x_1 y_2 - \cancel{x_2} y_1)_{k,l} \quad + \cdot \ + \cdot \ + \cdot + \cdot + \cdot$

Man erhält so Sechservektor $\mathfrak{F}$.

$\mathfrak{F}_{11} \ \mathfrak{F}_{12} \ \mathfrak{F}_{13} \ \mathfrak{F}_{14}$
$\mathfrak{F}_{21} \ \mathfrak{F}_{22} \ \mathfrak{F}_{23} \ \mathfrak{F}_{24} \quad \mathfrak{F}_{12} \ \mathfrak{F}_{23} \ \mathfrak{F}_{31} \ \mathfrak{F}_{14} \cdot \mathfrak{F}_{34}$
$\mathfrak{F}_{31} \ \cdots \ \mathfrak{F}_{34}$
$\mathfrak{F}_{41} \ \cdots\cdots \ \mathfrak{F}_{44}$

Vollständig geschrieben als Tensor mit Symmetrie zu 0-Diagonale.

Beispiel $\quad i\frac{\partial}{\partial x_1} \quad j\frac{\partial}{\partial x_2} \quad k\frac{\partial}{\partial x_3} \quad l\frac{\partial}{\partial x_4}$

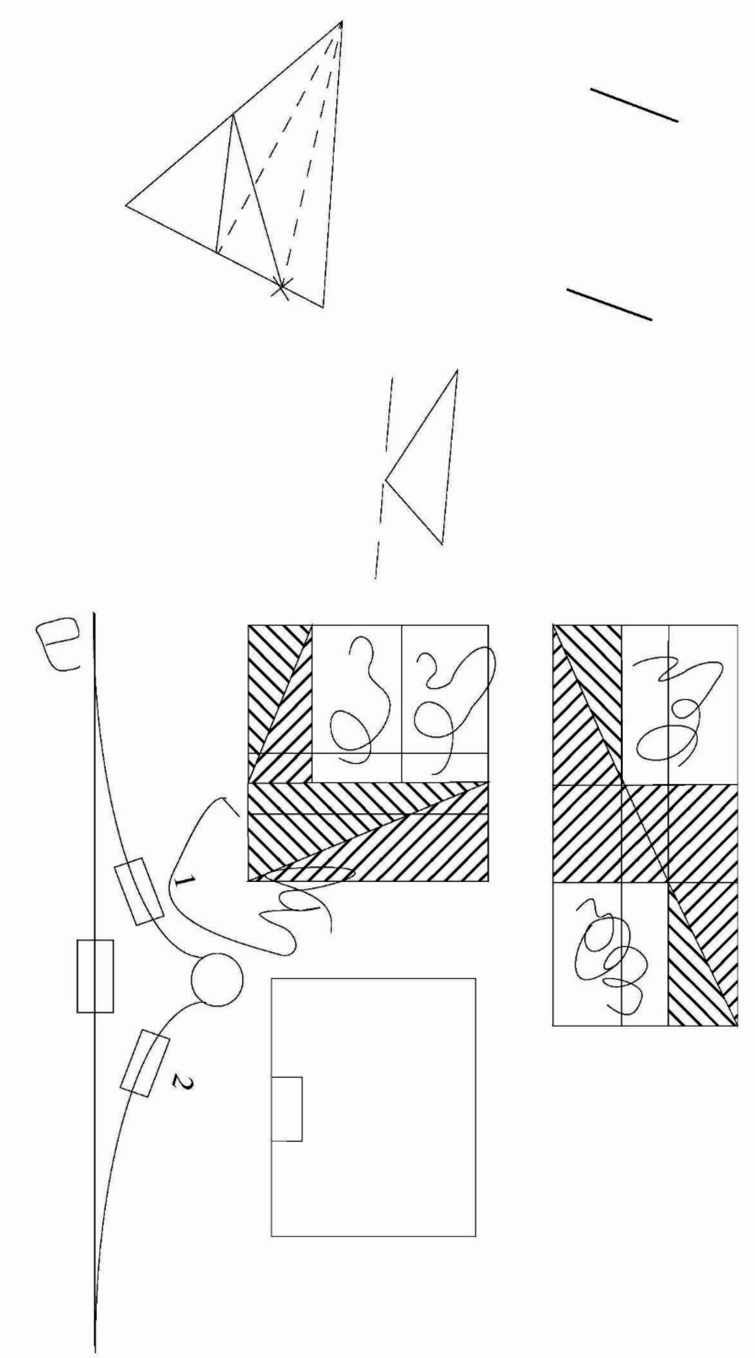
$-\Gamma_x \quad -\Gamma_y \quad -\Gamma_z \quad -i\varphi \qquad (\varphi_1 \ \varphi_2 \ \varphi_3 \ \varphi_4)$

Man erhält $\mathfrak{F}_{12} = \frac{\partial i\varphi}{\partial z} - \frac{\partial \Gamma_z}{\partial ict} = i\left(-\frac{\partial \varphi}{\partial z} - \frac{\partial \Gamma_z}{\partial t}\right) = i\mathfrak{E}_z$

$\mathfrak{F}_{14} = -\mathfrak{F}_{23} = \frac{\partial \Gamma_z}{\partial y} - \frac{\partial \Gamma_y}{\partial z} = \mathfrak{H}_x$

$\mathfrak{F}_{12} \ \mathfrak{F}_{23} \ \mathfrak{F}_{31} \ \mathfrak{F}_{14} \ \mathfrak{F}_{24} \ \mathfrak{F}_{34}$
$+i\mathfrak{E}_z \ +i\mathfrak{E}_x \ i\mathfrak{E}_y \ \mathfrak{H}_x \ \mathfrak{H}_y \ \mathfrak{H}_z$ $\qquad \mathfrak{H}_x \ \mathfrak{H}_y \ \mathfrak{H}_z \ i\mathfrak{E}_x \ i\mathfrak{E}_y \ i\mathfrak{E}_z$
$\mathfrak{G}_{34} \ \mathfrak{G}_{14} \ \mathfrak{G}_{24} \ \mathfrak{G}_{23} \ \mathfrak{G}_{31}$ $\qquad$ zugeordneter Vektor

[p. 32 L]



[p. 32 R]

$$\begin{array}{cccc} x & y & z & ict \\ x_1 & x_2 & x_3 & x_4 \end{array}$$

Invarianter Skalar $\cancel{i}d\tau = dt\left(\sqrt{1-\frac{v^2}{c^2}}\right)$ $d\tau$ invarianter Skalar, sodass

$$d\tau^2 = \sqrt{-(dx_1^2 + dx_2^2 + + \cdot}$$

Dividiert man durch diesen Skalar, so erhält man wieder Vierervektor

$\beta\frac{dx}{dt} \cdot \cdot \cdot \; ic\beta$ (Vierervektor der Geschwindigkeit eines materiellen Punktes)

Skalares Produkt zweier Vierervektoren

$$x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4 = (xy) \qquad (i \cdot i = 1 \qquad i \cdot k = 0).$$

Vektorprodukt zweier Vierervektoren.

$$\begin{array}{cccc} i & j & k & l \\ x_1 & x_2 & x_3 & x_4 \\ y_1 & y_2 & y_3 & y_4 \end{array}$$

$$(i)\cdot(i) = 0 \quad (i)(j) = (kl) \quad (ik) = (jl)\cdots$$
$$(j)(i) = -(kl) \quad (il) = (kj)$$

$$(xy) = (x_1y_2 - y_2y_1)_{k,l} \quad + \cdot + \cdot + \cdot + \cdot + \cdot$$

Man erhält so Sechservektor *F*.

$$\begin{array}{llll} F_{11} & F_{12} & F_{13} & F_{14} \\ F_{21} & F_{22} & F_{23} & F_{24} \\ F_{31} & \ldots\ldots & & F_{34} \\ F_{41} & \ldots\ldots & & F_{44} \end{array} \qquad F_{12}F_{23}F_{31}F_{14} \cdot F_{34}$$

Vollständig geschrieben als Tenso[r] mit Symmetrie zu 0-Diagonale.

Beispiel $i\frac{\partial}{\partial x_1} \quad j\frac{\partial}{\partial x_2} \quad k\frac{\partial}{\partial x_3} \quad l\frac{\partial}{\partial x_4}$

$$-\Gamma_x \quad -\Gamma_y \quad -\Gamma_z \quad -i\cancel{c}\varphi \qquad (\varphi_1 \;\; \varphi_2 \;\; \varphi_3 \;\; \varphi_4)$$

Man erhält $F_{12} = \frac{\partial i\cancel{c}\varphi}{\partial z} - \frac{\partial -\Gamma_z}{\partial ict} = i\left(-\frac{\partial\varphi}{\partial z} - \frac{\partial\Gamma_z}{\partial t}\right) = iE_z$

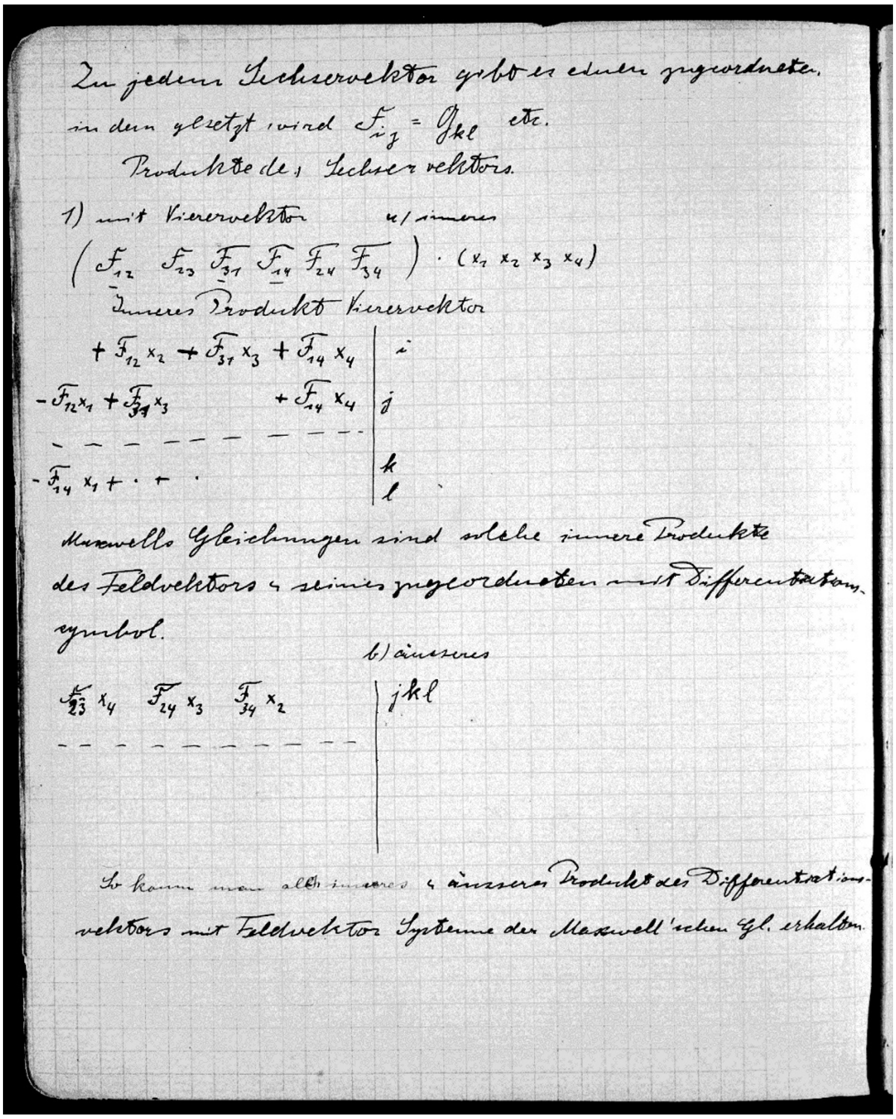
$$F_{14} = -\mathrm{F}_{23} = \frac{\partial\Gamma_z}{\partial y} - \frac{\partial\Gamma_y}{\partial z} = H_x$$

$$\begin{array}{cccccc} F_{12} & F_{23} & F_{31} & F_{14} & F_{24} & F_{34} \\ +iE_z & +iE_x & iE_y & H_x & H_y & H_z \\ G_{34} & G_{14} & G_{24} & G_{23} & G_{31} & G_{12} \end{array}$$

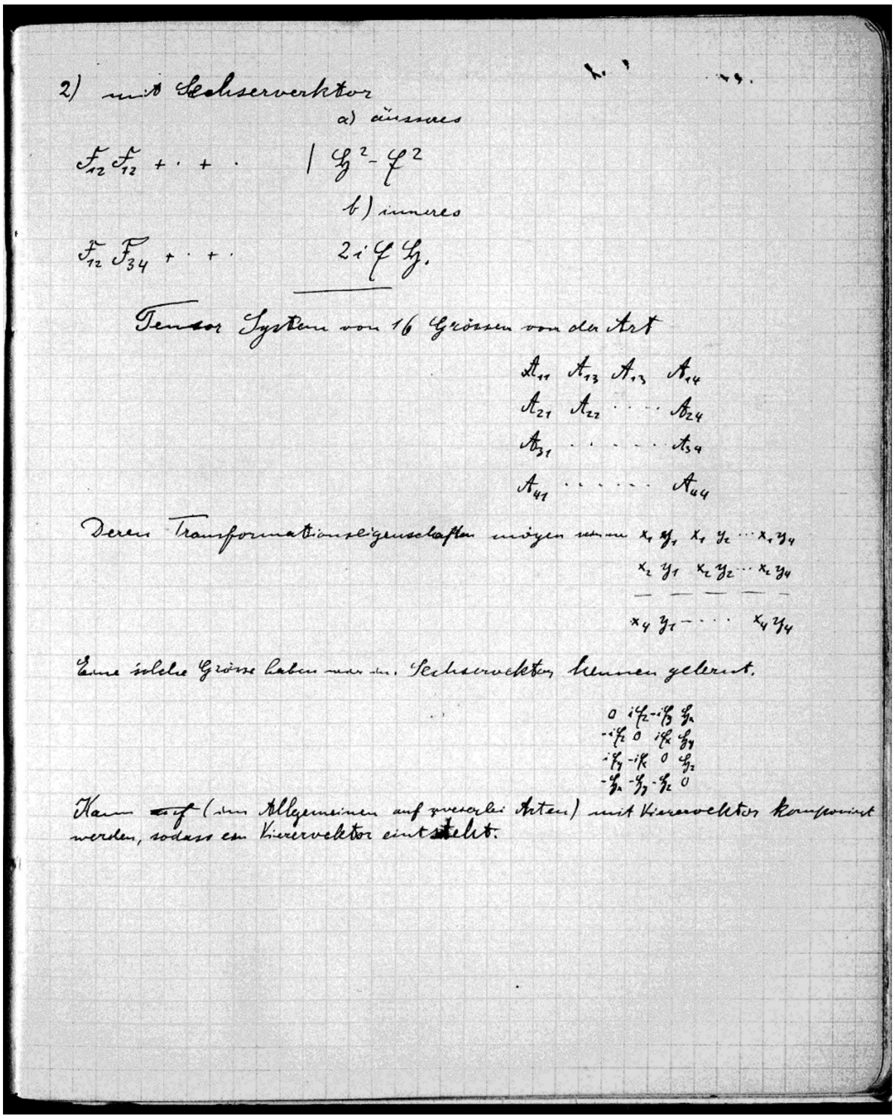
$H_x \; H_y \; H_z \; iE_x \; iE_y \; iE_z$

zugeordneter Vektor

[p. 33 L]


Zu jedem Sechservektor gibt es einen zugeordneten.
in dem gesetzt wird $\mathfrak{F}_{ij} = \mathfrak{G}_{kl}$ etc.
Produkte des Sechservektors.
1) mit Vierervektor
a) inneres
$\left(\mathfrak{F}_{12}\ \mathfrak{F}_{23}\ \mathfrak{F}_{31}\ \mathfrak{F}_{14}\ \mathfrak{F}_{24}\ \mathfrak{F}_{34}\right) \cdot (x_1\ x_2\ x_3\ x_4)$
Inneres Produkt Vierervektor
$+\mathfrak{F}_{12}x_2 + \mathfrak{F}_{31}x_3 + \mathfrak{F}_{14}x_4$ | i
$-\mathfrak{F}_{12}x_1 + \mathfrak{F}_{31}x_3 + \mathfrak{F}_{14}x_4$ | j
$- - - - - -$ | k
$-\mathfrak{F}_{14}x_1 + \cdot + \cdot$ | l
Maxwells Gleichungen sind solche innere Produkte
des Feldvektors u. seines zugeordneten mit Differentiations-
symbol.
b) äusseres
$\mathfrak{F}_{23}x_4\ \ \mathfrak{F}_{24}x_3\ \ \mathfrak{F}_{34}x_2$ | jkl
So kann man als inneres u. äusseres Produkt des Differentiations-
vektors mit Feldvektor Systeme der Maxwell'schen Gl. erhalten.

[p. 33R]



2) mit Sechservektor

α) äusseres

$\mathfrak{F}_{12}\mathfrak{F}_{12} + \cdot + \cdot \quad | \quad \mathfrak{H}^2 - \mathfrak{E}^2$

b) inneres

$\mathfrak{F}_{12}\mathfrak{F}_{34} + \cdot + \cdot \qquad 2i\mathfrak{E}\mathfrak{H}.$

Tensor System von 16 Grössen von der Art

$$\begin{matrix} A_{11} & A_{12} & A_{13} & A_{14} \\ A_{21} & A_{22} & \cdots & A_{24} \\ A_{31} & \cdots & \cdots & A_{34} \\ A_{41} & \cdots & \cdots & A_{44} \end{matrix}$$

Deren Transformationseigenschaften mögen wie

$$\begin{matrix} x_1 y_1 & x_1 y_2 & \cdots & x_1 y_4 \\ x_2 y_1 & x_2 y_2 & \cdots & x_2 y_4 \\ - & - & - & - \\ x_4 y_1 & \cdots & & x_4 y_4 \end{matrix}$$

Eine solche Grösse haben wir im Sechservektor kennen gelernt.

$$\begin{matrix} 0 & i\mathfrak{H}_z & -i\mathfrak{H}_y & \mathfrak{E}_x \\ -i\mathfrak{H}_z & 0 & i\mathfrak{H}_x & \mathfrak{E}_y \\ i\mathfrak{H}_y & -i\mathfrak{H}_x & 0 & \mathfrak{E}_z \\ -\mathfrak{E}_x & -\mathfrak{E}_y & -\mathfrak{E}_z & 0 \end{matrix}$$

Kann ~~auf~~ (im Allgemeinen auf zweierlei Arten) mit Vierervektor kombiniert werden, sodass ein Vierervektor entsteht.

[p. 33 L]

Zu jedem Sechservektor gibt es einen zugeordneten.
in dem gesetzt wird $F_{ij} = G_{kl}$ etc.

Produkte des Sechservektors.

1) mit Vierervektor a) inneres

$$(F_{\underline{12}}\ F_{23}\ F_{\underline{31}}\ F_{\underline{14}}\ F_{24}\ F_{34}) \cdot (x_1 x_2 x_3 x_4)$$

Inneres Produkt Vierervektor

$$\begin{array}{r|l}
+F_{12}x_2 + F_{31}x_3 + F_{14}x_4 & i \\
-F_{12}x_1 + F_{31}x_3 \qquad + F_{14}x_4 & j \\
------------- & \\
-F_{14}x_1 + \ \cdot\ + \ \cdot \qquad\qquad & k \\
 & l
\end{array}$$

Maxwells Gleichungen sind solche innere Produkte
des Feldvektors u seines zugeordeneten mit Differentiations-
symbol.

b) äusseres

$$\begin{array}{l|l}
F_{23}x_4 \quad F_{24}x_3 \quad F_{34}x_2 & jkl \\
------------- & \\
 &
\end{array}$$

So kann man al<l>s inneres u äusseres Produkt des Differentiations-
vektors mit Feldvektor Systeme der Maxwell'schen Gl. erhalten.

[p. 33 R]

2) mit Sechservektor

a) äusseres

$F_{12}F_{12} + \cdot + \cdot \quad \Big| \quad H^2 - E^2$

b) inneres

$F_{12}F_{34} + \cdot + \cdot \quad 2iEH.$

Tensor System von 16 Grössen von der Art

$$\begin{matrix} A_{11} & A_{13} & A_{13} & A_{14} \\ A_{11} & A_{22} & \cdots & A_{24} \\ A_{31} & \cdots & \cdots & A_{34} \\ A_{41} & \cdots & \cdots & A_{44} \end{matrix}$$

Deren Transformationseigenschaften mögen sein w

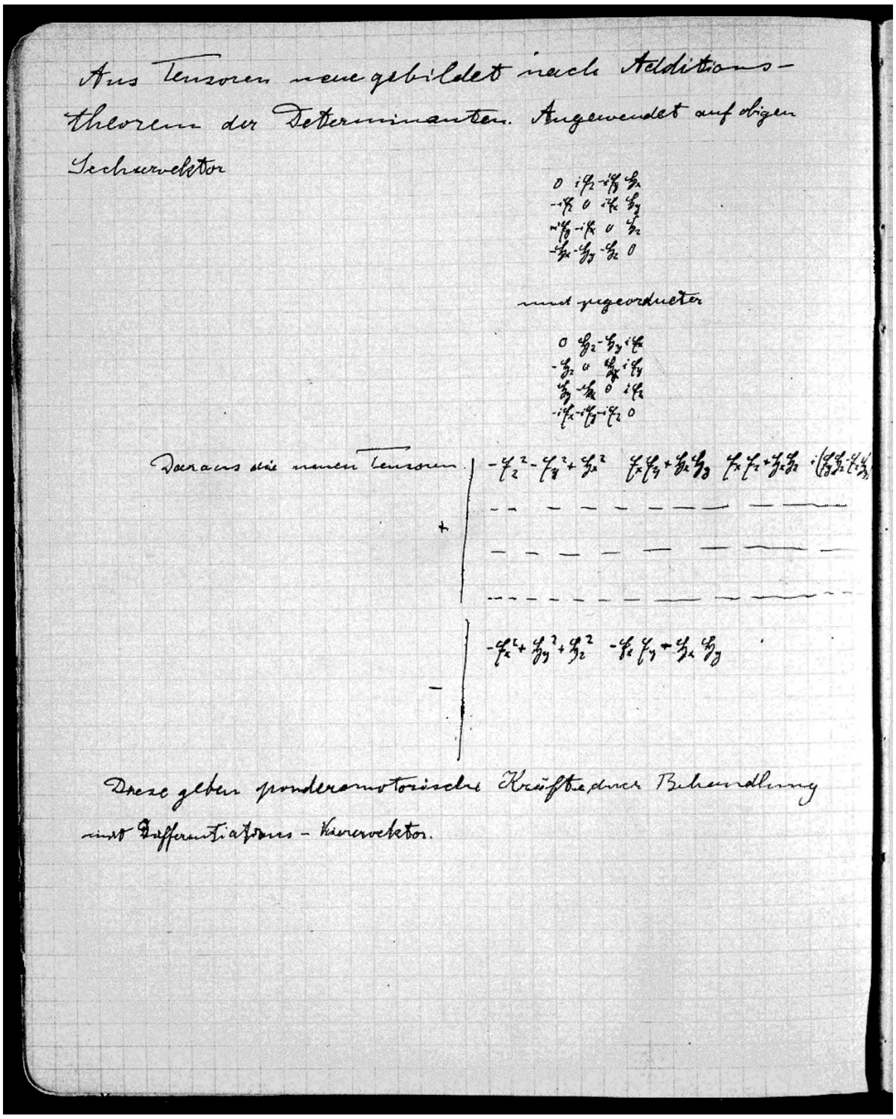
$$\begin{matrix} x_1y_1 & x_1y_2 & \cdots & x_1y_4 \\ x_2y_1 & x_2y_2 & \cdots & x_2y_4 \\ - & - & - & - \\ x_4y_1 & \cdots & \cdots & x_4y_4 \end{matrix}$$

Eine solche Grösse haben wir im Sechservektor kennen gelernt.

$$\begin{matrix} 0 & iE_z & -iE_y & H_x \\ -iE_z & 0 & iE_x & H_y \\ iE_y & -iE_x & 0 & H_z \\ -H_x & -H_y & -H_z & 0 \end{matrix}$$

Kann ~~auf~~ (im Allgemeinen auf zweierlei Arten) mit Vierervektor komponiert werden, sodass ein Vierervektor eintsteht.

[p. 34 L]



Aus Tensoren neue gebildet nach Additionstheorem der Determinanten. Angewendet auf obigen Sechservektor

$$\begin{matrix} 0 & i\mathfrak{E}_z & -i\mathfrak{E}_y & \mathfrak{H}_x \\ -i\mathfrak{E}_z & 0 & i\mathfrak{E}_x & \mathfrak{H}_y \\ i\mathfrak{E}_y & -i\mathfrak{E}_x & 0 & \mathfrak{H}_z \\ -\mathfrak{H}_x & -\mathfrak{H}_y & -\mathfrak{H}_z & 0 \end{matrix}$$

und zugeordneten

$$\begin{matrix} 0 & \mathfrak{H}_z & -\mathfrak{H}_y & i\mathfrak{E}_x \\ -\mathfrak{H}_z & 0 & \mathfrak{H}_x & i\mathfrak{E}_y \\ \mathfrak{H}_y & -\mathfrak{H}_x & 0 & i\mathfrak{E}_z \\ -i\mathfrak{E}_x & -i\mathfrak{E}_y & -i\mathfrak{E}_z & 0 \end{matrix}$$

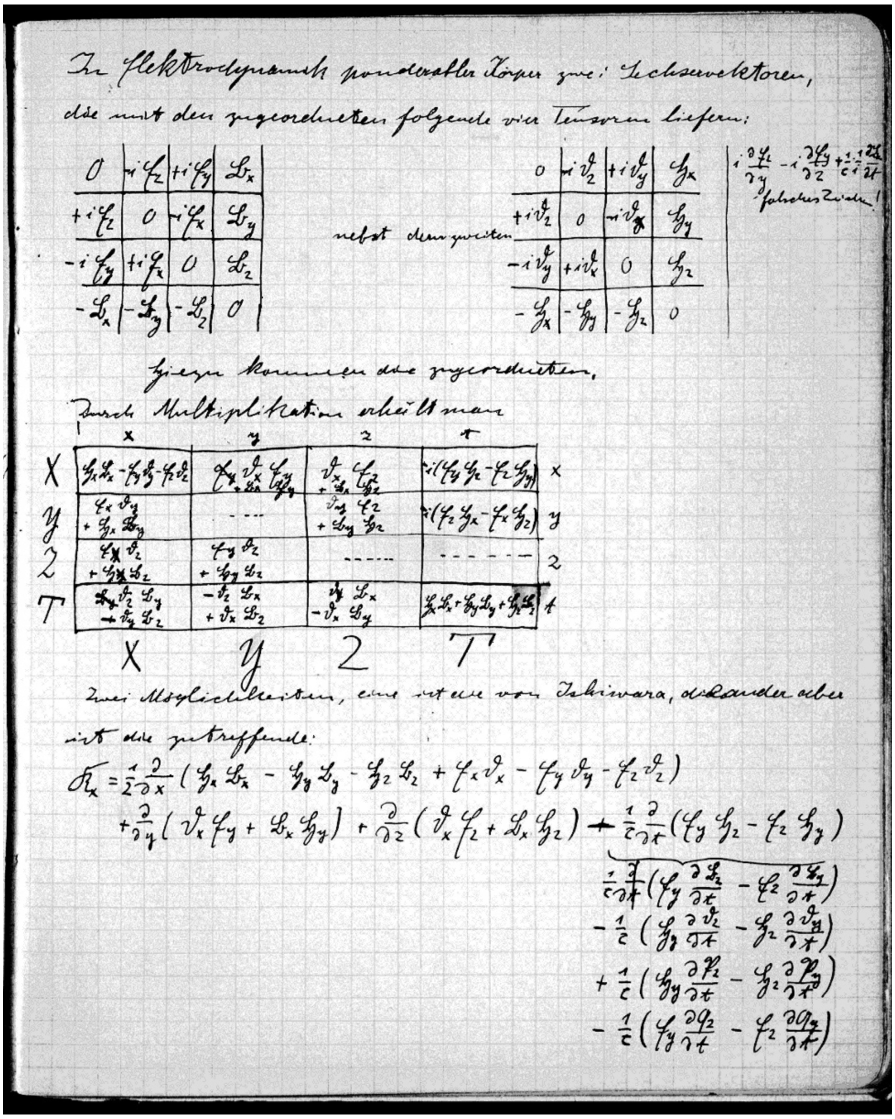
Daraus die neuen Tensoren

$$+\left|\begin{matrix} -\mathfrak{E}_z^2-\mathfrak{E}_y^2+\mathfrak{H}_x^2 & \mathfrak{E}_x\mathfrak{E}_y+\mathfrak{H}_x\mathfrak{H}_y & \mathfrak{E}_x\mathfrak{E}_z+\mathfrak{H}_x\mathfrak{H}_z & i(\mathfrak{E}_y\mathfrak{H}_z-\mathfrak{E}_z\mathfrak{H}_y) \\ - & - & - & - \\ - & - & - & - \\ - & - & - & - \end{matrix}\right.$$

$$-\left|\begin{matrix} -\mathfrak{E}_x^2+\mathfrak{H}_y^2+\mathfrak{H}_z^2 & -\mathfrak{E}_x\mathfrak{E}_y-\mathfrak{H}_x\mathfrak{H}_y & \cdot \end{matrix}\right.$$

Diese geben ponderomotorische Kräfte durch Behandlung mit Differentiations-Vierervektor.

[p. 34 R]



In Elektrodynamik ponderabler Körper zwei Sechservektoren, die mit den zugeordneten folgende vier Tensoren liefern:

$$\begin{array}{cccc} 0 & -i\mathfrak{E}_z & +i\mathfrak{E}_y & \mathfrak{B}_x \\ +i\mathfrak{E}_z & 0 & -i\mathfrak{E}_x & \mathfrak{B}_y \\ -i\mathfrak{E}_y & +i\mathfrak{E}_x & 0 & \mathfrak{B}_z \\ -\mathfrak{B}_x & -\mathfrak{B}_y & -\mathfrak{B}_z & 0 \end{array} \qquad \text{nebst dem zweiten} \qquad \begin{array}{cccc} 0 & -i\mathfrak{D}_z & +i\mathfrak{D}_y & \mathfrak{H}_x \\ +i\mathfrak{D}_z & 0 & -i\mathfrak{D}_x & \mathfrak{H}_y \\ -i\mathfrak{D}_y & +i\mathfrak{D}_x & 0 & \mathfrak{H}_z \\ -\mathfrak{H}_x & -\mathfrak{H}_y & -\mathfrak{H}_z & 0 \end{array}$$

$i\frac{\partial \mathfrak{E}_z}{\partial y} - i\frac{\partial \mathfrak{E}_y}{\partial z} + \frac{1}{c}\frac{\partial \mathfrak{B}_x}{\partial t}$ falsches Zeichen!

Hiezu kommen die zugeordneten.

Durch Multiplikation erhält man

	x	y	z	t	
X	$\mathfrak{H}_x\mathfrak{B}_x - \mathfrak{E}_y\mathfrak{D}_y - \mathfrak{E}_z\mathfrak{D}_z$	$\mathfrak{D}_x\mathfrak{E}_y + \mathfrak{B}_x\mathfrak{H}_y$	$\mathfrak{D}_x\mathfrak{E}_z + \mathfrak{B}_x\mathfrak{H}_z$	$i(\mathfrak{E}_y\mathfrak{H}_z - \mathfrak{E}_z\mathfrak{H}_y)$	x
Y	$\mathfrak{E}_x\mathfrak{D}_y + \mathfrak{H}_x\mathfrak{B}_y$	----	$\mathfrak{D}_y\mathfrak{E}_z + \mathfrak{B}_y\mathfrak{H}_z$	$i(\mathfrak{E}_z\mathfrak{H}_x - \mathfrak{E}_x\mathfrak{H}_z)$	y
Z	$\mathfrak{E}_x\mathfrak{D}_z + \mathfrak{H}_x\mathfrak{B}_z$	$\mathfrak{E}_y\mathfrak{D}_z + \mathfrak{H}_y\mathfrak{B}_z$	------	------	z
T	$\mathfrak{D}_z\mathfrak{B}_y - \mathfrak{D}_y\mathfrak{B}_z$	$-\mathfrak{D}_z\mathfrak{B}_x + \mathfrak{D}_x\mathfrak{B}_z$	$\mathfrak{D}_y\mathfrak{B}_x - \mathfrak{D}_x\mathfrak{B}_y$	$\mathfrak{H}_x\mathfrak{B}_x + \mathfrak{H}_y\mathfrak{B}_y + \mathfrak{H}_z\mathfrak{B}_z$	t
	X	Y	Z	T	

Zwei Möglichkeiten, eine ist die von Ishiwara, die andere aber ist die zutreffende:

$$\mathfrak{K}_x = \frac{1}{2}\frac{\partial}{\partial x}(\mathfrak{H}_x\mathfrak{B}_x - \mathfrak{H}_y\mathfrak{B}_y - \mathfrak{H}_z\mathfrak{B}_z + \mathfrak{E}_x\mathfrak{D}_x - \mathfrak{E}_y\mathfrak{D}_y - \mathfrak{E}_z\mathfrak{D}_z)$$

$$+\frac{\partial}{\partial y}(\mathfrak{D}_x\mathfrak{E}_y + \mathfrak{B}_x\mathfrak{H}_y) + \frac{\partial}{\partial z}(\mathfrak{D}_x\mathfrak{E}_z + \mathfrak{B}_x\mathfrak{H}_z) + \underbrace{\frac{1}{c}\frac{\partial}{\partial t}(\mathfrak{E}_y\mathfrak{H}_z - \mathfrak{E}_z\mathfrak{H}_y)}$$

$$\frac{1}{c}\frac{\partial}{\partial t}\left(\mathfrak{E}_y\frac{\partial \mathfrak{B}_z}{\partial t} - \mathfrak{E}_z\frac{\partial \mathfrak{B}_y}{\partial t}\right)$$

$$-\frac{1}{c}\left(\mathfrak{H}_y\frac{\partial \mathfrak{D}_z}{\partial t} - \mathfrak{H}_z\frac{\partial \mathfrak{D}_y}{\partial t}\right)$$

$$+\frac{1}{c}\left(\mathfrak{H}_y\frac{\partial \mathfrak{P}_z}{\partial t} - \mathfrak{H}_z\frac{\partial \mathfrak{P}_y}{\partial t}\right)$$

$$-\frac{1}{c}\left(\mathfrak{E}_y\frac{\partial \mathfrak{Q}_z}{\partial t} - \mathfrak{E}_z\frac{\partial \mathfrak{Q}_y}{\partial t}\right)$$

[p. 34 L]

Aus Tensoren neue gebildet nach Additionstheorem der Determinaten.
Angewendet auf obigen Sechservektor

$$\begin{array}{cccc} 0 & iE_z & -iE_y & H_x \\ -iE_z & 0 & iE_x & H_y \\ +iE_y & -iE_x & 0 & H_z \\ -H_x & -H_y & -H_z & 0 \end{array}$$

und zugeordneter

$$\begin{array}{cccc} 0 & H_z & -H_y & iE_x \\ -H_z & 0 & H_x & iE_y \\ H_y & -H_x & 0 & iE_z \\ -iE_x & -i\mathrm{E}_y & -iE_z & 0 \end{array}$$

Daraus die neuen Tensoren.

$$+\left|\begin{array}{cccc} -E_z^2 - E_y^2 + H_x^2 & E_xE_y + H_xH_y & E_xE_z + H_xH_z & i(E_yH_z - E_zH_y) \\ - - - & - - - & - - - & - - - \\ - - - & - - - & - - - & - - - \\ - - - & - - - & - - - & - - - \end{array}\right.$$

$$-\left|\begin{array}{cccc} -E_x^2 + H_y^2 + H_z^2 & -E_xE_y + H_xH_y & \cdot & \cdot \\ & & & \\ & & & \end{array}\right.$$

Diese geben ponderomotorsche Kräfte, durch Behandlung mit Differentiations-Vierervektor.

[p. 34 R]

In Elektrodynamik ponderabler Körper zwei Sechservektoren, die mit den zugeordneten folgende vier Tensoren liefern:

$$\begin{array}{c|c|c|c|}
0 & -iE_z & +iE_y & B_x \\ \hline
+iE_z & 0 & -iE_x & B_y \\ \hline
-iE_y & +iE_x & 0 & B_z \\ \hline
-B_x & -B_y & -B_z & 0
\end{array}
\quad \text{nebst dem zweiten} \quad
\begin{array}{c|c|c|c|}
0 & -iD_z & +iD_y & H_x \\ \hline
+iD_z & 0 & -iD_x & H_y \\ \hline
-iD_y & +iD_x & 0 & H_z \\ \hline
-H_x & -H_y & -H_z & 0
\end{array}$$

$$i\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} + \frac{1}{c}\cdot\frac{1}{i}\frac{\partial B_x}{\partial t}$$

falsches Zeichen!

Hie[r]zu kommen die zugeordneten.

Durch Multiplikation erhält man

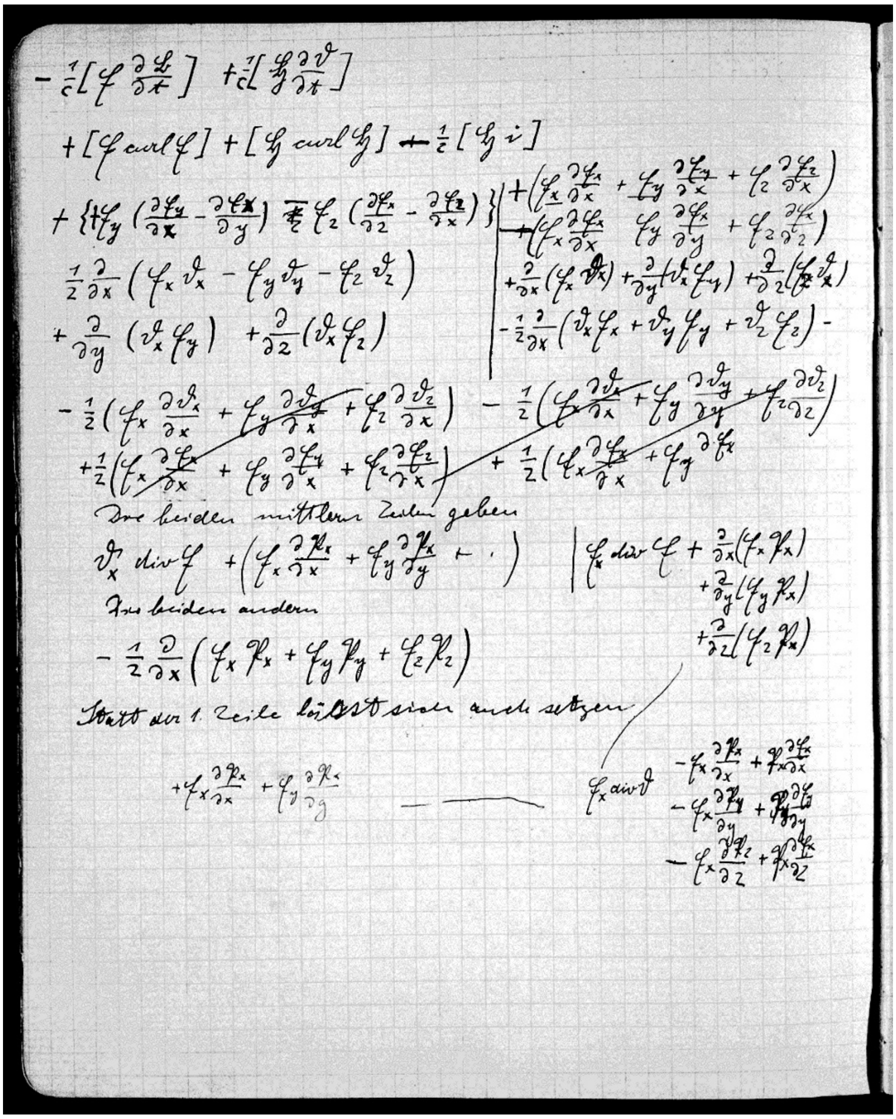
	x	y	z	t	
X	$H_xB_x - E_yD_y - E_zD_z$	$\cancel{E_y}D_xE_y$ $+B_xH_y$	D_xE_z $+B_xH_z$	$-i(E_yH_z - E_zH_y)$	x
Y	E_xD_y $+H_xB_y$	$- - - -$	D_yE_z $+B_yH_z$	$-i(E_zH_x - E_xH_z)$	y
Z	E_xD_z $+H_xB_z$	E_yD_z $+H_yB_z$	$- - - -$	$- - - - - - - - -$	z
T	$\cancel{B_y}D_zB_y$ $\cancel{+}-D_yB_z$	$-D_zB_x$ $+D_xB_z$	D_yB_x $-D_xB_y$	$H_xB_x + H_yB_y + H_zB_z$	t
	X	Y	Z	T	

Zwei Möglichkeiten, eine ist die von Ishiwara, die andere aber ist die zutreffende:

$$K_x = \frac{1}{2}\frac{\partial}{\partial x}(H_xB_x - H_yB_y - H_zB_z + E_xD_x - E_yD_y - E_zD_z)$$

$$+\frac{\partial}{\partial y}(D_xE_y + B_xH_y) + \frac{\partial}{\partial z}(D_xE_z + B_xH_z)\langle + \rangle \underbrace{-\frac{1}{c}\frac{\partial}{\partial t}(E_yH_z - E_zH_y)}_{\begin{array}{l}
\frac{1}{c}\cancel{\frac{\partial}{\partial t}}\left(E_y\frac{\partial B_z}{\partial t} - E_z\frac{\partial B_y}{\partial t}\right) \\
-\frac{1}{c}\left(H_y\frac{\partial D_z}{\partial t} - H_z\frac{\partial D_y}{\partial t}\right) \\
+\frac{1}{c}\left(H_y\frac{\partial P_z}{\partial t} - H_z\frac{\partial P_y}{\partial t}\right) \\
-\frac{1}{c}\left(E_y\frac{\partial Q_z}{\partial t} - E_z\frac{\partial Q_y}{\partial t}\right)
\end{array}}$$

[p. 35 L]



$$-\frac{1}{c}\left[\mathfrak{E}\frac{\partial \mathfrak{L}}{\partial t}\right] + \frac{1}{c}\left[\mathfrak{H}\frac{\partial \vartheta}{\partial t}\right]$$

$$+[\mathfrak{E}\,\mathrm{curl}\,\mathfrak{E}] + [\mathfrak{H}\,\mathrm{curl}\,\mathfrak{H}] \cancel{+} \frac{1}{c}[\mathfrak{H}\,i]$$

$$+\left\{\mathfrak{E}_y\left(\frac{\partial \mathfrak{E}_y}{\partial x}-\frac{\partial \mathfrak{E}_x}{\partial y}\right) - \mathfrak{E}_z\left(\frac{\partial \mathfrak{E}_x}{\partial z}-\frac{\partial \mathfrak{E}_z}{\partial x}\right)\right\}$$

$$\frac{1}{2}\frac{\partial}{\partial x}\left(\mathfrak{E}_x\vartheta_x - \mathfrak{E}_y\vartheta_y - \mathfrak{E}_z\vartheta_z\right)$$

$$+\frac{\partial}{\partial y}(\vartheta_x\mathfrak{E}_y) + \frac{\partial}{\partial z}(\vartheta_x\mathfrak{E}_z)$$

$$+\left(\mathfrak{E}_x\frac{\partial \mathfrak{E}_x}{\partial x} + \mathfrak{E}_y\frac{\partial \mathfrak{E}_y}{\partial x} + \mathfrak{E}_z\frac{\partial \mathfrak{E}_z}{\partial x}\right)$$

$$-\left(\mathfrak{E}_x\frac{\partial \mathfrak{E}_x}{\partial x} + \mathfrak{E}_y\frac{\partial \mathfrak{E}_x}{\partial y} + \mathfrak{E}_z\frac{\partial \mathfrak{E}_x}{\partial z}\right)$$

$$+\frac{\partial}{\partial x}(\mathfrak{E}_x\vartheta_x) + \frac{\partial}{\partial y}(\vartheta_x\mathfrak{E}_y) + \frac{\partial}{\partial z}(\mathfrak{E}_z\vartheta_x)$$

$$-\frac{1}{2}\frac{\partial}{\partial x}\left(\vartheta_x\mathfrak{E}_x + \vartheta_y\mathfrak{E}_y + \vartheta_z\mathfrak{E}_z\right) -$$

$$-\frac{1}{2}\left(\mathfrak{E}_x\frac{\partial \vartheta_x}{\partial x} + \mathfrak{E}_y\frac{\partial \vartheta_y}{\partial x} + \mathfrak{E}_z\frac{\partial \vartheta_z}{\partial x}\right) - \frac{1}{2}\left(\mathfrak{E}_x\frac{\partial \vartheta_x}{\partial x} + \mathfrak{E}_y\frac{\partial \vartheta_y}{\partial y} + \mathfrak{E}_z\frac{\partial \vartheta_z}{\partial z}\right)$$

$$+\frac{1}{2}\left(\mathfrak{E}_x\frac{\partial \mathfrak{E}_x}{\partial x} + \mathfrak{E}_y\frac{\partial \mathfrak{E}_y}{\partial x} + \mathfrak{E}_z\frac{\partial \mathfrak{E}_z}{\partial x}\right) + \frac{1}{2}\left(\mathfrak{E}_x\frac{\partial \mathfrak{E}_x}{\partial x} + \mathfrak{E}_y\frac{\partial \mathfrak{E}_x}{}\right.$$

Die beiden mittleren Zeilen geben

$$\vartheta_x\,\mathrm{div}\,\mathfrak{E} + \left(\mathfrak{E}_x\frac{\partial \mathfrak{K}_x}{\partial x} + \mathfrak{E}_y\frac{\partial \mathfrak{K}_x}{\partial y} + \cdot\right) \quad \Big|\ \mathfrak{E}_x\,\mathrm{div}\,\mathfrak{E} + \frac{\partial}{\partial x}(\mathfrak{E}_x\mathfrak{K}_x) + \frac{\partial}{\partial y}(\mathfrak{E}_y\mathfrak{K}_x) + \frac{\partial}{\partial z}(\mathfrak{E}_z\mathfrak{K}_x)$$

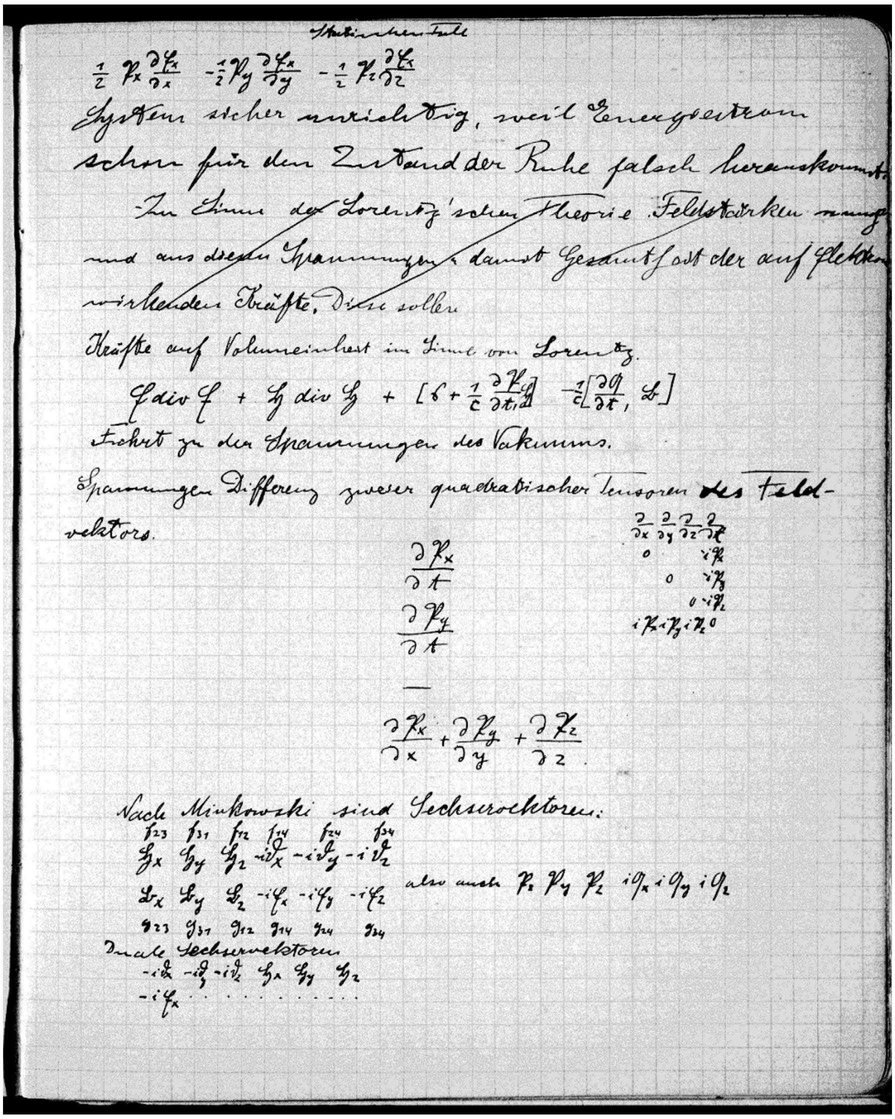
Die beiden andern

$$-\frac{1}{2}\frac{\partial}{\partial x}\left(\mathfrak{E}_x\mathfrak{K}_x + \mathfrak{E}_y\mathfrak{K}_y + \mathfrak{E}_z\mathfrak{K}_z\right)$$

Statt der 1. Zeile lässt sich auch setzen

$$+\mathfrak{E}_x\frac{\partial \mathfrak{K}_x}{\partial x} + \mathfrak{E}_y\frac{\partial \mathfrak{K}_x}{\partial y} \quad \text{———} \quad \mathfrak{E}_x\,\mathrm{div}\,\vartheta \quad \begin{array}{l} -\mathfrak{E}_x\frac{\partial \mathfrak{K}_x}{\partial x} + \mathfrak{K}_x\frac{\partial \mathfrak{E}_x}{\partial x} \\ -\mathfrak{E}_x\frac{\partial \mathfrak{K}_y}{\partial y} + \mathfrak{K}_y\frac{\partial \mathfrak{E}_x}{\partial y} \\ -\mathfrak{E}_x\frac{\partial \mathfrak{K}_z}{\partial z} + \mathfrak{K}_z\frac{\partial \mathfrak{E}_x}{\partial z} \end{array}$$

[p. 35 R]


Stationären Fall

$\frac{1}{2}\mathfrak{P}_x\frac{\partial \mathfrak{E}_x}{\partial x} \quad -\frac{1}{2}\mathfrak{P}_y\frac{\partial \mathfrak{E}_x}{\partial y} \quad -\frac{1}{2}\mathfrak{P}_z\frac{\partial \mathfrak{E}_x}{\partial z}$

System sicher unrichtig, weil Energiestrom schon für den Zustand der Ruhe falsch herauskommt.

~~Im Sinne der Lorentz'schen Theorie Feldstärken maßg. und aus diesen Spannungen, damit Gesamtheit der auf Elektron wirkenden Kräfte. Diese sollen~~

Kräfte auf Volumeinheit im Sinne von Lorentz.

$$\mathfrak{E}\,\mathrm{div}\,\mathfrak{E} + \mathfrak{H}\,\mathrm{div}\,\mathfrak{H} + \left[\mathfrak{s} + \frac{1}{c}\frac{\partial \mathfrak{E}}{\partial t}, \mathfrak{H}\right] - \frac{1}{c}\left[\frac{\partial \mathfrak{H}}{\partial t}, \mathfrak{E}\right]$$

Führt zu den Spannungen des Vakuums.

Spannungen Differenz zweier quadratischer Tensoren des Feldvektors.

$$\frac{\partial \mathfrak{P}_x}{\partial t} \qquad \frac{\partial \mathfrak{P}_y}{\partial t}$$

$$\frac{\partial \mathfrak{P}_x}{\partial x} + \frac{\partial \mathfrak{P}_y}{\partial y} + \frac{\partial \mathfrak{P}_z}{\partial z}.$$

$$\begin{array}{cccc} \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} & \frac{\partial}{\partial t} \\ 0 & & & i\mathfrak{P}_x \\ & 0 & & i\mathfrak{P}_y \\ & & 0 & i\mathfrak{P}_z \\ i\mathfrak{P}_x & i\mathfrak{P}_y & i\mathfrak{P}_z & 0 \end{array}$$

Nach Minkowski sind Sechservektoren:

$$\begin{array}{cccccc} f_{23} & f_{31} & f_{12} & f_{14} & f_{24} & f_{34} \\ \mathfrak{H}_x & \mathfrak{H}_y & \mathfrak{H}_z & -i\mathfrak{E}_x & -i\mathfrak{E}_y & -i\mathfrak{E}_z \end{array}$$

also auch $\mathfrak{P}_x\ \mathfrak{P}_y\ \mathfrak{P}_z\ i\mathfrak{Q}_x\ i\mathfrak{Q}_y\ i\mathfrak{Q}_z$

$$\begin{array}{cccccc} \mathfrak{L}_x & \mathfrak{L}_y & \mathfrak{L}_z & -i\mathfrak{E}_x & -i\mathfrak{E}_y & -i\mathfrak{E}_z \\ g_{23} & g_{31} & g_{12} & g_{14} & g_{24} & g_{34} \end{array}$$

Duale Sechservektoren

$$\begin{array}{cccccc} -i\mathfrak{E}_x & -i\mathfrak{E}_y & -i\mathfrak{E}_z & \mathfrak{H}_x & \mathfrak{H}_y & \mathfrak{H}_z \\ -i\mathfrak{E}_x & \dots & & & & \end{array}$$

[p. 35 L]

$$-\frac{1}{c}\left[E\frac{\partial B}{\partial t}\right] \quad +\frac{1}{c}\left[H\frac{\partial D}{\partial t}\right]$$

$$+[E\ \mathrm{curl}\ E]+[H\ \mathrm{curl}\ H]-\frac{1}{c}[Hi]$$

$$+\left\{+E_y\left(\frac{\partial E_y}{\partial x}-\frac{\partial E_x}{\partial y}\right)-E_z\left(\frac{\partial E_x}{\partial z}-\frac{\partial E_z}{\partial x}\right)\right\}$$

$$\frac{1}{2}\frac{\partial}{\partial x}(E_xD_x-E_yD_y-E_zD_z)$$

$$+\frac{\partial}{\partial y}(D_xE_y)+\frac{\partial}{\partial z}(D_xE_z)$$

$$+\left(\underline{E_x}\frac{\partial E_x}{\partial x}+\underline{E_y}\frac{\partial E_y}{\partial x}+E_z\frac{\partial E_z}{\partial x}\right)$$

$$-\left(E_x\frac{\partial E_x}{\partial x}\quad E_y\frac{\partial E_x}{\partial y}+E_z\frac{\partial E_x}{\partial z}\right)$$

$$+\frac{\partial}{\partial x}(E_xD_x)+\frac{\partial}{\partial y}(D_xE_y)+\frac{\partial}{\partial z}(E_zD_x)$$

$$-\frac{1}{2}\frac{\partial}{\partial x}(D_xE_x+D_yE_y+D_zE_z)-$$

$$\cancel{-\frac{1}{2}\left(E_x\frac{\partial D_x}{\partial x}+E_y\frac{\partial D_y}{\partial x}+E_z\frac{\partial D_z}{\partial x}\right)} \quad \cancel{-\frac{1}{2}\left(E_x\frac{\partial D_x}{\partial x}+E_y\frac{\partial D_y}{\partial y}+E_z\frac{\partial D_z}{\partial z}\right)}$$

$$\cancel{+\frac{1}{2}\left(E_x\frac{\partial E_x}{\partial x}+E_y\frac{\partial E_y}{\partial x}+E_z\frac{\partial E_z}{\partial x}\right)} \quad \cancel{+\frac{1}{2}\left(E_x\frac{\partial E_x}{\partial x}+E_y\frac{\partial E_x}{}\right)}$$

Die beiden mittlern Zeilen geben

$$D_x\ \mathrm{div}\ E \quad +\left(E_x\frac{\partial P_x}{\partial x}+E_y\frac{\partial P_x}{\partial y}+\quad\cdot\right)$$

$$E_x\mathrm{div}E\ +\frac{\partial}{\partial x}(E_xP_x)+\frac{\partial}{\partial y}(E_yP_x)+\frac{\partial}{\partial z}(E_zP_x)$$

Die beiden andern

$$-\frac{1}{2}\frac{\partial}{\partial x}(E_xP_x+E_yP_y+E_zP_z)$$

Statt der 1. Zeile lässt sich auch setzen

$$+E_x\frac{\partial P_x}{\partial x}+E_y\frac{\partial P_x}{\partial y}$$

$$E_x\mathrm{div}D \quad \begin{aligned}&-E_x\frac{\partial P_x}{\partial x}+P_x\frac{\partial E_x}{\partial x}\\ &-E_x\frac{\partial P_y}{\partial y}+P_x\frac{\partial E_y}{\partial y}\\ &-E_x\frac{\partial P_z}{\partial z}+P_x\frac{\partial E_x}{\partial z}\end{aligned}$$

[p. 35 R]

Statischer Fall

$$\frac{1}{2}P_x\frac{\partial E_x}{\partial x} \quad -\frac{1}{2}P_y\frac{\partial E_x}{\partial y} \quad -\frac{1}{2}P_z\frac{\partial E_x}{\partial z}$$

System sicher unrichtig, weil Energiestrom schon für den Zustand der Ruhe falsch herauskommt.

~~- Im Sinne der Lorentz'schen Theorie: Feldstärken $\mathfrak{e}$ und $\mathfrak{h}$ und aus diesen Spannungen u damit Gesamtheit der auf Elektron wirkenden Kräfte. Diese sollen~~

Kräfte auf Volumeneinheit im Sinne von Lorentz.

$$E\operatorname{div}E + H\operatorname{div}H + \left[s + \frac{1}{c}\frac{\partial P}{\partial t}, H\right] \quad -\frac{1}{c}\left[\frac{\partial Q}{\partial t}, B\right]$$

Führt zu den Spannungen des Vakuums.
Spannungen Differenz zweier quadratischer Tensoren des Feldvektors.

$$\begin{array}{cccc} \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} & \frac{\partial}{\partial t} \\ 0 & & & -iP_x \\ & 0 & & -iP_y \\ & & 0 & -iP_z \\ iP_x & iP_y & iP_z & 0 \end{array}$$

$$\frac{\partial P_x}{\partial t}$$

$$\frac{\partial P_y}{\partial t}$$

$$\frac{\partial \mathrm{P}_x}{\partial x} + \frac{\partial \mathrm{P}_y}{\partial y} + \frac{\partial \mathrm{P}_z}{\partial z}\,.$$

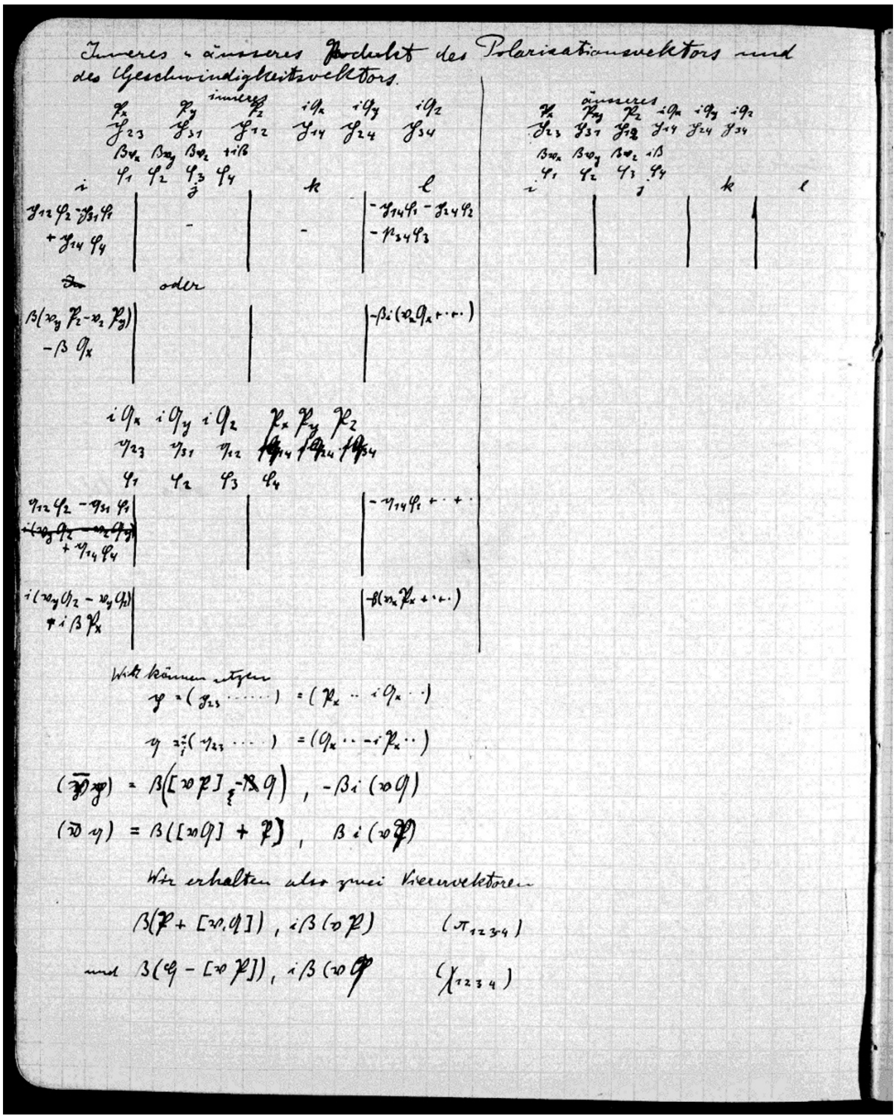
Nach Minkowski sind Sechservektoren:

$$\begin{array}{cccccc} f_{23} & f_{31} & f_{12} & f_{14} & f_{24} & f_{34} \\ H_x & H_y & H_z & -iD_x & -iD_y & -iD_z \\ B_x & B_y & B_z & -iE_x & -iE_y & -iE_z \\ g_{23} & g_{31} & g_{12} & g_{14} & g_{24} & g_{34} \end{array} \qquad \text{also auch} \qquad P_x \;\; P_y \;\; P_z \;\; iQ_x \;\; iQ_y \;\; iQ_z$$

Duale Sechservektoren

$$-iD_x \; -iD_y \; -iD_z \; H_x \; H_y \; H_z$$
$$-iE_x \cdots\cdots\cdots\cdots$$

[p. 36 L]


Inneres u. äusseres Produkt des Polarisationsvektors und des Geschwindigkeitsvektors.

inneres

$\mathfrak{P}_x \quad \mathfrak{P}_y \quad \mathfrak{P}_z \quad i\mathfrak{Q}_x \quad i\mathfrak{Q}_y \quad i\mathfrak{Q}_z$

$\mathfrak{f}_{23} \quad \mathfrak{f}_{31} \quad \mathfrak{f}_{12} \quad \mathfrak{f}_{14} \quad \mathfrak{f}_{24} \quad \mathfrak{f}_{34}$

$\beta\mathfrak{v}_x \quad \beta\mathfrak{v}_y \quad \beta\mathfrak{v}_z \quad +i\beta$

$\varphi_1 \quad \varphi_2 \quad \varphi_3 \quad \varphi_4$

i	j	k	l
$\mathfrak{f}_{12}\varphi_2 - \mathfrak{f}_{31}\varphi_1 + \mathfrak{f}_{14}\varphi_4$	–	–	$-\mathfrak{f}_{14}\varphi_1 - \mathfrak{f}_{24}\varphi_2 - \mathfrak{f}_{34}\varphi_3$

oder

$\beta(\mathfrak{v}_y\mathfrak{P}_z - \mathfrak{v}_z\mathfrak{P}_y) - \beta\mathfrak{Q}_x$			$-\beta i(\mathfrak{v}_x\mathfrak{Q}_x + \cdot + \cdot)$

$i\mathfrak{Q}_x \quad i\mathfrak{Q}_y \quad i\mathfrak{Q}_z \quad \mathfrak{P}_x \quad \mathfrak{P}_y \quad \mathfrak{P}_z$

$\gamma_{23} \quad \gamma_{31} \quad \gamma_{12} \quad \gamma_{14} \quad \gamma_{24} \quad \gamma_{34}$

$\varphi_1 \quad \varphi_2 \quad \varphi_3 \quad \varphi_4$

$\gamma_{12}\varphi_2 - \gamma_{31}\varphi_1$ ~~$i(\mathfrak{v}_y\mathfrak{Q}_z - \mathfrak{v}_z\mathfrak{Q}_y)$~~ $+ \gamma_{14}\varphi_4$			$-\gamma_{14}\varphi_1 + \cdot + \cdot$
$i(\mathfrak{v}_y\mathfrak{Q}_z - \mathfrak{v}_z\mathfrak{Q}_y) + i\beta\mathfrak{P}_x$			$-\beta(\mathfrak{v}_x\mathfrak{P}_x + \cdot + \cdot)$

äusseres

$\mathfrak{P}_x \quad \mathfrak{P}_y \quad \mathfrak{P}_z \quad i\mathfrak{Q}_x \quad i\mathfrak{Q}_y \quad i\mathfrak{Q}_z$

$\mathfrak{f}_{23} \quad \mathfrak{f}_{31} \quad \mathfrak{f}_{12} \quad \mathfrak{f}_{14} \quad \mathfrak{f}_{24} \quad \mathfrak{f}_{34}$

$\beta\mathfrak{v}_x \quad \beta\mathfrak{v}_y \quad \beta\mathfrak{v}_z \quad i\beta$

$\varphi_1 \quad \varphi_2 \quad \varphi_3 \quad \varphi_4$

$i \quad j \quad k \quad l$

Wir können setzen

$\mathfrak{f} = (\mathfrak{f}_{23} \cdots\cdots) = (\mathfrak{P}_x \cdots i\mathfrak{Q}_x \cdots)$

$\gamma = \frac{1}{i}(\gamma_{23} \cdots) = (\mathfrak{Q}_x \cdots -i\mathfrak{P}_x \cdots)$

$(\mathfrak{v}\mathfrak{f}) = \beta([\mathfrak{v}\mathfrak{P}] - \mathfrak{Q}), \quad -\beta i(\mathfrak{v}\mathfrak{Q})$

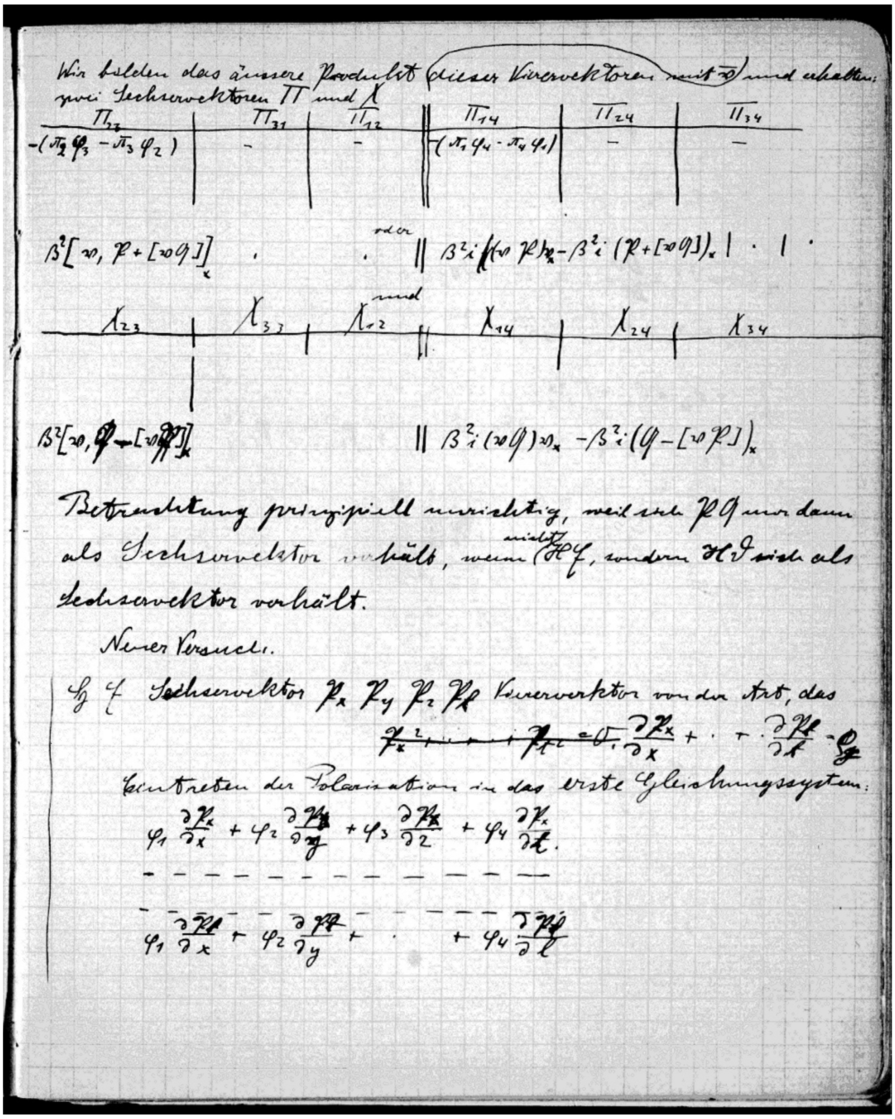
$(\mathfrak{v}\gamma) = \beta([\mathfrak{v}\mathfrak{Q}] + \mathfrak{P}), \quad \beta i(\mathfrak{v}\mathfrak{P})$

Wir erhalten also zwei Vierervektoren

$\beta(\mathfrak{P} + [\mathfrak{v}, \mathfrak{Q}]), \; i\beta(\mathfrak{v}\mathfrak{P}) \qquad (\pi_{1234})$

und $\beta(\mathfrak{Q} - [\mathfrak{v}\mathfrak{P}]), \; i\beta(\mathfrak{v}\mathfrak{Q}) \qquad (\chi_{1234})$

[p. 36 R]



Wir bilden das äussere Produkt dieser Vierervektoren mit $\mathfrak{w}$ und erhalten zwei Sechservektoren Π und χ

Π_{23}	Π_{31}	Π_{12}	Π_{14}	Π_{24}	Π_{34}
$-(\pi_2\varphi_3 - \pi_3\varphi_2)$	–	–	$-(\pi_1\varphi_4 - \pi_4\varphi_1)$	–	–
$\beta^2[\mathfrak{w}, \mathfrak{P} + [\mathfrak{w}\mathfrak{Q}]]_x$	.	. oder	$\beta^2 i(\mathfrak{w}\mathfrak{P})\mathfrak{w}_x - \beta^2 i(\mathfrak{P} + [\mathfrak{w}\mathfrak{Q}])_x$	.	.

und

χ_{23}	χ_{31}	χ_{12}	χ_{14}	χ_{24}	χ_{34}
$\beta^2[\mathfrak{w}, \mathfrak{Q} - [\mathfrak{w}\mathfrak{P}]]_x$			$\beta^2 i(\mathfrak{w}\mathfrak{Q})\mathfrak{w}_x - \beta^2 i(\mathfrak{Q} - [\mathfrak{w}\mathfrak{P}])_x$		

Betrachtung prinzipiell unrichtig, weil sich $\mathfrak{P}\mathfrak{Q}$ nur dann als Sechservektor verhält, wenn nicht $\mathfrak{E}\mathfrak{h}$, sondern $\mathfrak{E}\mathfrak{d}$ sich als Sechservektor verhält.

Neuer Versuch.

$\mathfrak{G}$ $\mathfrak{f}$ Sechservektor $\mathfrak{P}_x$ $\mathfrak{P}_y$ $\mathfrak{P}_z$ $\mathfrak{P}_t$ Vierervektor von der Art, dass

$$\cancel{\mathfrak{P}_x^2 + \cdots + \mathfrak{P}_t^2 = 0} \quad \frac{\partial \mathfrak{P}_x}{\partial x} + \cdot + \frac{\partial \mathfrak{P}_t}{\partial t} = \mathfrak{Q}$$

Eintreten der Polarisation in das erste Gleichungssystem:

$$\varphi_1 \frac{\partial \mathfrak{P}_x}{\partial x} + \varphi_2 \frac{\partial \mathfrak{P}_x}{\partial y} + \varphi_3 \frac{\partial \mathfrak{P}_x}{\partial z} + \varphi_4 \frac{\partial \mathfrak{P}_x}{\partial t}$$

– – – – – – – – – – – –

$$\varphi_1 \frac{\partial \mathfrak{P}_t}{\partial x} + \varphi_2 \frac{\partial \mathfrak{P}_t}{\partial y} + \cdot + \varphi_4 \frac{\partial \mathfrak{P}_t}{\partial t}$$

[p. 36 L]

Inneres u äusseres Produkt des Polarisationsvektors und des Geschwindigkeitsvektors.

inneres

$$\begin{array}{cccccc} P_x & P_y & P_z & iQ_x & iQ_y & iQ_z \\ g_{23} & g_{31} & g_{12} & g_{14} & g_{24} & g_{34} \\ \beta v_x & \beta v_y & \beta v_z & +i\beta & & \\ \varphi_1 & \varphi_2 & \varphi_3 & \varphi_4 & & \end{array}$$

$$\begin{array}{c|c|c|c} i & j & k & l \\ g_{12}\varphi_2 - g_{31}\varphi_1 & - & - & -g_{14}\varphi_1 - g_{24}\varphi_2 \\ +g_{14}\varphi_4 & & & -g_{34}\varphi_3 \end{array}$$

~~In~~ oder

$$\begin{array}{c|c|c|c} \beta(v_y P_z - v_z P_y) & & & -\beta_i(v_x Q_x + \cdot + \cdot) \\ -\beta Q_x & & & \end{array}$$

$$\begin{array}{cccccc} iQ_x & iQ_y & iQ_z & P_x & P_y & P_z \\ q_{23} & q_{31} & q_{12} & q_{14} & q_{24} & q_{34} \\ \varphi_1 & \varphi_2 & \varphi_3 & \varphi_4 & & \end{array}$$

$$\begin{array}{c|c|c|c} q_{12}\varphi_2 - q_{31}\varphi_1 & & & -q_{14}\varphi_1 + \cdot + \cdot \\ i\cancel{(v_y Q_z - v_z Q_y)} & & & \\ +q_{14}\varphi_4 & & & \end{array}$$

$$\begin{array}{c|c|c|c} i(v_y Q_z - v_y Q_z) & & & -\beta(v_x P_x + \cdot + \cdot) \\ +\beta Px & & & \end{array}$$

äusseres

$$\begin{array}{cccccc} P_x & P_y & P_z & iQ_x & iQ_y & iQ_z \\ g_{23} & g_{31} & g_{12} & g_{14} & g_{24} & g_{34} \\ \beta v_x & \beta v_y & \beta v_z & i\beta & & \\ \varphi_1 & \varphi_2 & \varphi_3 & \varphi_4 & & \end{array}$$

$$\begin{array}{c|c|c|c} i & j & k & l \\ & & & \end{array}$$

Wir können setzen

$$g = (g_{23} \cdot\cdot\cdot\cdot\cdot\cdot) = (P_x \cdot\cdot \ iQ_x \cdot\cdot)$$

$$q = \frac{1}{i}(q_{23} \cdot\cdot\cdot\cdot\cdot\cdot) = (Q_x \cdot\cdot -iP_x \cdot\cdot)$$

$$(\bar{v}p) = \beta([vP] - \cancel{\beta}Q), \ -\beta i(vQ)$$

$$(\bar{v}q) = \beta([vQ] + P), \ \beta i(vP)$$

Wir erhalten also zwei Vierervektoren

$$\beta(P + [v,Q]), \ i\beta(vP) \qquad (\pi_{1\ 2\ 3\ 4})$$

und $\beta(G - [v,P]), \ i\beta(vQ \qquad (\chi_{1\ 2\ 3\ 4})$

[p. 36 R]

Wir bilden das äussere Produkt dieser Vierervektoren mit $\bar{v}$ und erhalten: zwei Sechservektoren Π und χ

Π_{23}	Π_{31}	Π_{12}	Π_{14}	Π_{24}	Π_{34}
$-(\pi_2\varphi_3 - \pi_3\varphi_2)$	—	—	$-(\pi_1\varphi_4 - \pi_4\varphi_1)$	—	—

oder

$\beta^2[v, P + [vQ]]_x$	$\cdot$	$\cdot$	$\beta^2 i(vP)v_x - \beta^2 i(P + [vQ])_x$	$\cdot$	$\cdot$

und

χ_{23}	χ_{33}	χ_{12}	χ_{14}	χ_{24}	χ_{34}
$\beta^2[v, Q - [vP]]_x$			$\beta^2 i(vQ)v_x - \beta^2 i(Q - [vP])_x$		

Betrachtung prinzipiell unrichtig, weil sich $\mathfrak{P}\mathfrak{Q}$ nur dann als Sechservektor verhält, wenn nicht $\mathfrak{H}\mathfrak{E}$, sondern $\mathfrak{H}\mathfrak{D}$ sich als Sechservektor verhält.

Neuer Versuch.

$\mathfrak{H}\mathfrak{E}$ Sechservektor $\quad P_x \; P_y \; P_z \; P_l$ Vierervektor von der Art, das

$$\cancel{P_x{}^2 + \cdot + \cdot + P_l{}^2 = 0.} \quad \frac{\partial P_x}{\partial x} + \cdot + \cdot \frac{\partial P_l}{\partial l} = \rho_p$$

Eintreten der Polarisation in das erste Gleichungssystem:

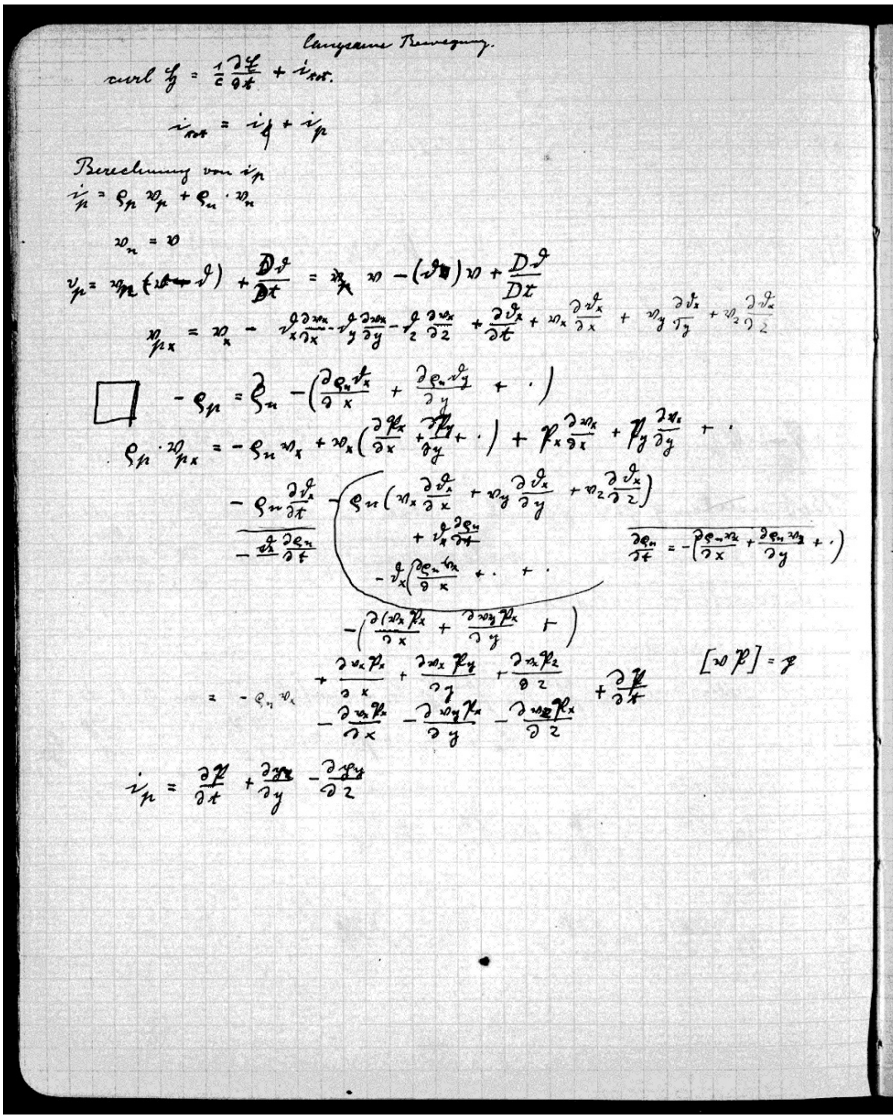
$$\varphi_1 \frac{\partial P_x}{\partial x} + \varphi_2 \frac{\partial P_x}{\partial y} + \varphi_3 \frac{\partial P_x}{\partial z} + \varphi_4 \frac{\partial P_x}{\partial l}.$$

– – – – – – – – – – – – – – – –

– – – – – – – – – – – – – – – –

$$\varphi_1 \frac{\partial P_l}{\partial x} + \varphi_2 \frac{\partial P_l}{\partial y} + \quad \cdot \quad + \varphi_4 \frac{\partial P_l}{\partial l}$$

[p. 37 L]



Langsame Bewegung.

$$\mathrm{curl}\,\mathfrak{H} = \frac{1}{c}\frac{\partial \mathfrak{E}}{\partial t} + i_{rot.}$$

$$i_{rot} = i_q + i_p$$

Berechnung von i_p

$$i_p = \varrho_p \mathfrak{w}_p + \varrho_n \cdot \mathfrak{w}_n$$

$$\mathfrak{w}_n = \mathfrak{w}$$

$$v_p = \mathfrak{w}_p \cancel{(\mathfrak{w} - \vartheta)} + \frac{D\vartheta}{Dt} = \mathfrak{w}\,\mathfrak{w} - (\vartheta\nabla)\mathfrak{w} + \frac{D\vartheta}{Dt}$$

$$\mathfrak{w}_{px} = \mathfrak{w}_x - \vartheta_x\frac{\partial \mathfrak{w}_x}{\partial x} - \vartheta_y\frac{\partial \mathfrak{w}_x}{\partial y} - \vartheta_z\frac{\partial \mathfrak{w}_x}{\partial z} + \frac{\partial \vartheta_x}{\partial t} + \mathfrak{w}_x\frac{\partial \vartheta_x}{\partial x} + \mathfrak{w}_y\frac{\partial \vartheta_x}{\partial y} + \mathfrak{w}_z\frac{\partial \vartheta_x}{\partial z}$$

$$\boxed{}\quad -\varrho_p = \varrho_n - \left(\frac{\partial \varrho_n \vartheta_x}{\partial x} + \frac{\partial \varrho_n \vartheta_y}{\partial y} + \cdot\right)$$

$$\varrho_p \cdot \mathfrak{w}_{px} = -\varrho_n \mathfrak{w}_x + \mathfrak{w}_x\left(\frac{\partial \mathfrak{P}_x}{\partial x} + \frac{\partial \mathfrak{P}_y}{\partial y} + \cdot\right) + \mathfrak{P}_x\frac{\partial \mathfrak{w}_x}{\partial x} + \mathfrak{P}_y\frac{\partial \mathfrak{w}_x}{\partial y} + \cdot$$

$$-\varrho_n\frac{\partial \vartheta_x}{\partial t} - \varrho_n\left(\mathfrak{w}_x\frac{\partial \vartheta_x}{\partial x} + \mathfrak{w}_y\frac{\partial \vartheta_x}{\partial y} + \mathfrak{w}_z\frac{\partial \vartheta_x}{\partial z}\right)$$

$$-\vartheta_x\frac{\partial \varrho_n}{\partial t} \qquad + \vartheta_x\frac{\partial \varrho_n}{\partial t}$$

$$-\vartheta_x\left(\frac{\partial \varrho_n \mathfrak{w}_x}{\partial x} + \cdot + \cdot\right.$$

$$\frac{\partial \varrho_n}{\partial t} = -\left(\frac{\partial \varrho_n \mathfrak{w}_x}{\partial x} + \frac{\partial \varrho_n \mathfrak{w}_y}{\partial y} + \cdot\right)$$

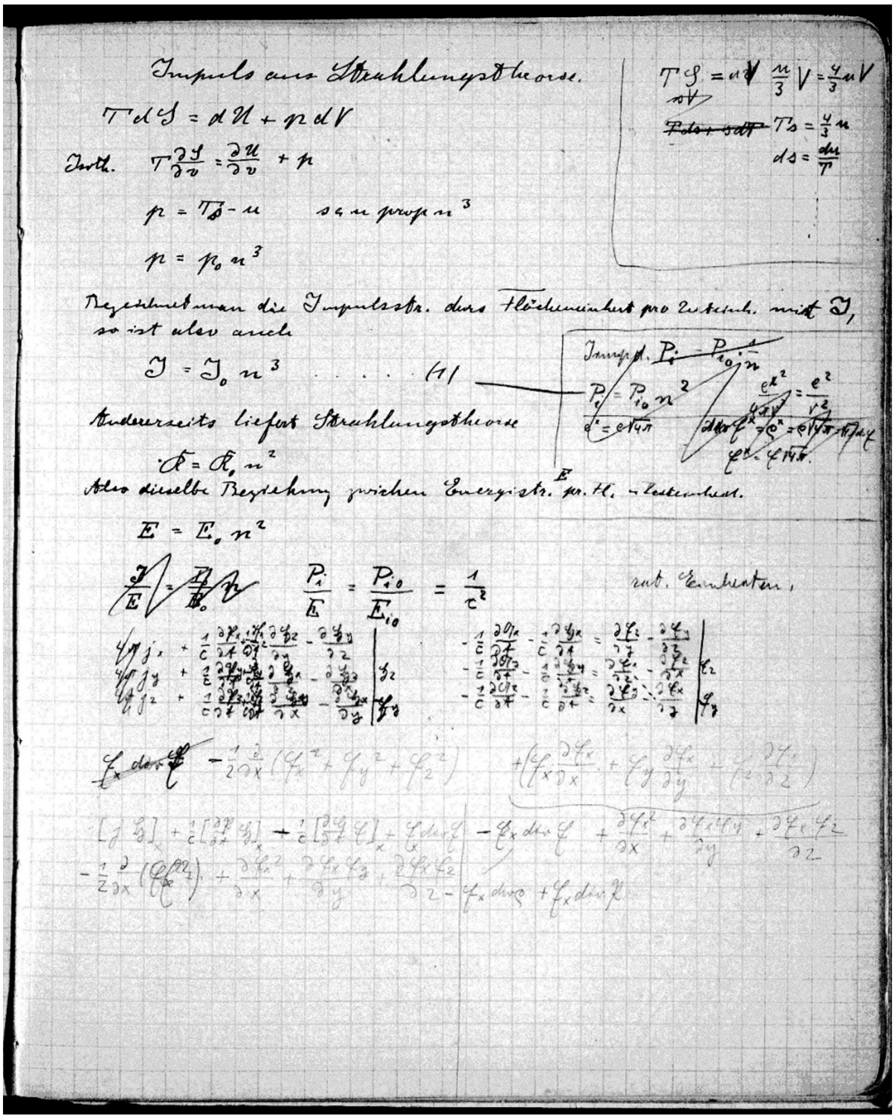
$$-\left(\frac{\partial(\mathfrak{w}_x \mathfrak{P}_x)}{\partial x} + \frac{\partial \mathfrak{w}_y \mathfrak{P}_x}{\partial y} + \right)$$

$$= -\varrho_n \mathfrak{w}_x + \frac{\partial \mathfrak{w}_x \mathfrak{P}_x}{\partial x} + \frac{\partial \mathfrak{w}_x \mathfrak{P}_y}{\partial y} + \frac{\partial \mathfrak{w}_x \mathfrak{P}_z}{\partial z} - \frac{\partial \mathfrak{w}_x \mathfrak{P}_x}{\partial x} - \frac{\partial \mathfrak{w}_y \mathfrak{P}_x}{\partial y} - \frac{\partial \mathfrak{w}_z \mathfrak{P}_x}{\partial z} + \frac{\partial \mathfrak{P}}{\partial t}$$

$$[\mathfrak{w}\mathfrak{P}] = \mathfrak{q}$$

$$i_p = \frac{\partial \mathfrak{P}}{\partial t} + \frac{\partial \mathfrak{q}_z}{\partial y} - \frac{\partial \mathfrak{q}_y}{\partial z}$$

[p. 37 R]


Impuls aus Strahlungstheorie.

$T\,dS = dU + p\,dV$

Isoth. $T\frac{\partial S}{\partial v} = \frac{\partial u}{\partial v} + p$

$p = T\mathfrak{s} - u \qquad s, u \text{ prop } n^3$

$p = p_0 n^3$

Bezeichnet man die Impulsstr. durch Flächeneinheit pro Zeiteinh. mit $\mathfrak{J}$, so ist also auch

$\mathfrak{J} = \mathfrak{J}_0 n^3 \quad \ldots \quad (1)$

Andererseits liefert Strahlungstheorie

$\mathfrak{E} = \mathfrak{E}_0 n^2$

Also dieselbe Beziehung zwischen Energiestr. pr. Fl. u. Zeiteinheit.

$E = E_0 n^2$

$\frac{P_1}{E} = \frac{P_{10}}{E_{10}} = \frac{1}{c^2}$ rat. Einheiten.

$TS = \frac{u}{3}V = \frac{4}{3}uV$

$Ts = \frac{4}{3}u$

$ds = \frac{du}{T}$

Impuls. $P_1 = P_{10} n^2$

[p. 37 L]

langsame Bewegung.

$$\text{curl } H = \frac{1}{c}\frac{\partial E}{\partial t} + i_{\text{tot.}}$$

$$i_{\text{tot}} = i_{\cancel{l}} + i_p$$

Berechnung von i_p

$$i_p = \rho_p v_p + \rho_n \cdot v_n$$

$$v_n = v$$

$$v_p = v_n(v - D) + \frac{DD}{Dt} = \cancel{v_p} v - (D)v + \frac{DD}{Dt}$$

$$v_{px} = v_x - D_x\frac{\partial v_x}{\partial x} - D_y\frac{\partial v_x}{\partial y} - D_z\frac{\partial v_x}{\partial z} + \frac{\partial D_x}{\partial t} + v_x\frac{\partial D_x}{\partial x} + v_y\frac{\partial D_x}{\partial y} + v_z\frac{\partial D_x}{\partial z}$$

$$-\rho_p = \rho_n - \left(\frac{\partial \rho_n D_x}{\partial x} + \frac{\partial \rho_n D_y}{\partial y} + \cdot\right)$$

$$\rho_p \cdot v_{px} = -\rho_n v_x + v_x\left(\frac{\partial P_x}{\partial x} + \frac{\partial P_y}{\partial y} + \cdot\right) + P_x\frac{\partial v_x}{\partial x} + P_y\frac{\partial v_x}{\partial y} + \cdot$$

$$-\rho_n\frac{\partial D_x}{\partial t} - \rho_n\left(v_x\frac{\partial D_x}{\partial x} + v_y\frac{\partial D_x}{\partial y} + v_z\frac{\partial D_x}{\partial z}\right)$$

$$-\underline{D_x}\frac{\partial \rho_n}{\partial t} \qquad + D_x\frac{\partial \rho_n}{\partial t} \qquad \frac{\partial \rho_n}{\partial t} = -\left(\frac{\partial \rho_n v_x}{\partial x} + \frac{\partial \rho_n v_y}{\partial y} + \cdot\right)$$

$$-D_x\left(\frac{\partial \rho_n v_x}{\partial x} + \cdot + \cdot\right.$$

$$-\left(\frac{\partial (v_x P_x}{\partial x} + \frac{\partial v_y P_x}{\partial y} + \right)$$

$$+\frac{\partial v_x P_x}{\partial x} + \frac{\partial v_x P_y}{\partial y} + \frac{\partial v_x P_z}{\partial z} \qquad [vP] = g$$

$$= -\rho_n v_x \qquad + \frac{\partial P}{\partial t}$$

$$-\frac{\partial v_x P_x}{\partial x} - \frac{\partial v_y P_x}{\partial y} - \frac{\partial v_z P_x}{\partial z}$$

$$i_p = \frac{\partial P}{\partial t} + \frac{\partial g_z}{\partial y} - \frac{\partial g_y}{\partial z}$$

[p. 37 R]

Impuls aus Strahlungstheorie.

$$TdS = dU + pdV$$

Isoth. $T\frac{\partial S}{\partial v} = \frac{\partial U}{\partial v} + p$

$p = Ts - u$ s u u prop n^3

$p = p_o n^3$

$TS = uV \quad \frac{u}{3}V = \frac{4}{3}uV$

$\cancel{sV}$

$\cancel{Tds + sdT} Ts = \frac{4}{3}u$

$ds = \frac{du}{T}$

Bezeichnet man die Impulsstr. durs Flächeninhalt pro Zeiteinh. mit $\mathfrak{I}$, so ist also auch

$I = I_o n^3 \quad \cdots\cdots \quad (1)$

Imp.d. $\cancel{P_i = P_{i_o} \cdot \frac{1}{n}}$ $\cancel{\frac{e^{x^2}}{4xr^2} = \frac{e^2}{r^2}}$

$\cancel{P_i = P_{i_o} n^2}$

$\cancel{e^x = e\sqrt{4\pi}\ \text{div}\ E^x = \rho^x = \rho\sqrt{4\pi} = \sqrt{4\pi}\,\text{div}E}$

$\cancel{E^x = E\sqrt{4\pi}.}$

Andererseits liefert Strahlungstheorie

$K = K_o n^2$

E

Also dieselbe Beziehung zwischen Energistr. pr. Fl. u Zeiteinheit.

$E = E_o n^2$

$\cancel{\frac{I}{E} = \frac{I}{E_o}n}$ $\qquad \frac{P_i}{E} = \frac{P_{io}}{E_{io}} = \frac{1}{c^2}$ rat. Einheiten.

$$\cancel{4\pi}\; j_x + \frac{1}{c}\frac{\partial P_x}{\partial t} + \frac{1}{c}\frac{\partial E_x}{\partial t} = \frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} \qquad -\frac{1}{c}\frac{\partial Q_x}{\partial t} - \frac{1}{c}\frac{\partial H_x}{\partial t} = \frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z}$$

$$\cancel{4\pi}\; j_y + \frac{1}{c}\frac{\partial P_y}{\partial t} + \frac{1}{c}\frac{\partial E_y}{\partial t} [=] \frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} \quad H_z \qquad -\frac{1}{c}\frac{\partial Q_y}{\partial t} - \frac{1}{c}\frac{\partial H_y}{\partial t} = \frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} \quad E_z$$

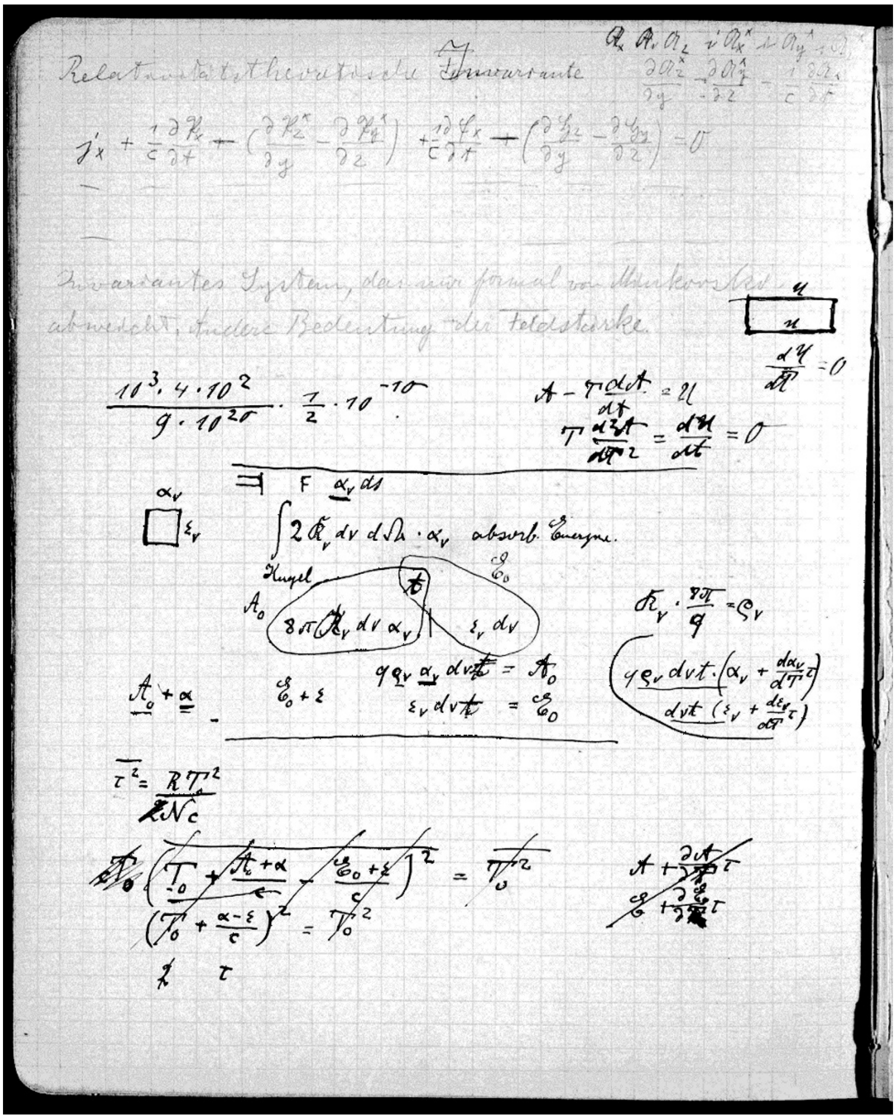
$$\cancel{4\pi}\; j_z + \frac{1}{c}\frac{\partial P_z}{\partial t} + \frac{1}{c}\frac{\partial E_z}{\partial t} [=] \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} \quad \text{-H}_y \qquad -\frac{1}{c}\frac{\partial Q_z}{\partial t} - \frac{1}{c}\frac{\partial H_z}{\partial t} = \frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} \quad -E_y$$

$$\cancel{E_x \text{div} E} \qquad -\frac{1}{2}\frac{\partial}{\partial x}(E_x^2 + E_y^2 + E_z^2) \qquad + \underbrace{\left(E_x\frac{\partial E_x}{\partial x} + E_y\frac{\partial E_x}{\partial y} + E_z\frac{\partial E_x}{\partial z}\right)}$$

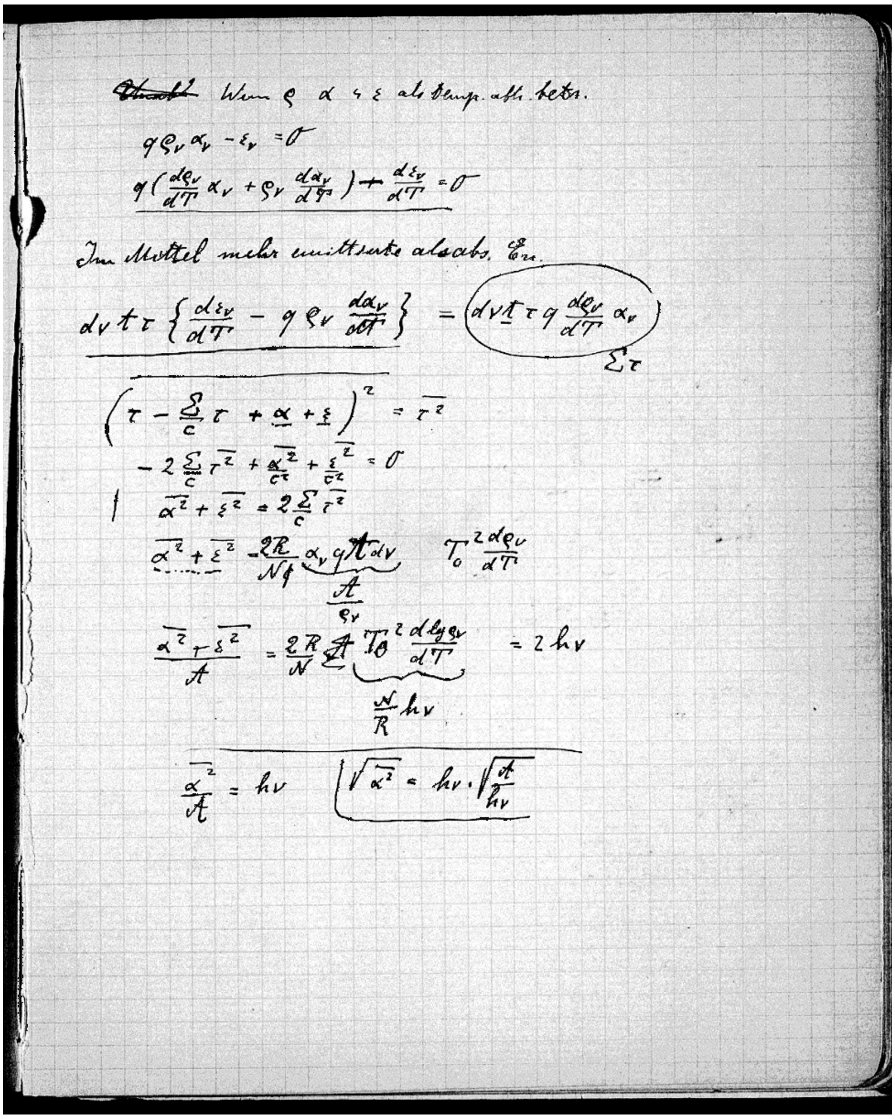
$$[jH]_x + \frac{1}{c}\left[\frac{\partial P}{\partial t}H\right]_x - \frac{1}{c}\left[\frac{\partial Q}{\partial t}E\right]_x - E_x\ \text{div}\ E \qquad - E_x\ \text{div}\ E + \frac{\partial E_x^2}{\partial x} + \frac{\partial E_x E_y}{\partial y} + \frac{\partial E_x E_z}{\partial z}$$

$$-\frac{1}{2}\frac{\partial}{\partial x}(E^2 +)\cdot + \frac{\partial E_x^2}{\partial x} + \frac{\partial E_x E_y}{\partial y} + \frac{\partial E_x E_z}{\partial z} - \qquad E_x \text{div}\rho + E_x \text{div} P$$

[p. 38 L]


Relativitätstheoretische Invariante
Invariantes System, das nur formal von Minkowski's abweicht. Andere Bedeutung der Feldstärke.
absorb. Energie
Kugel

[p. 38 R]


~~Überl.~~ Wenn ϱ α u. ε als Temp. abh. betr.
$q\varrho_\nu\alpha_\nu - \varepsilon_\nu = 0$
$q\left(\frac{d\varrho_\nu}{dT}\alpha_\nu + \varrho_\nu\frac{d\alpha_\nu}{dT}\right) + \frac{d\varepsilon_\nu}{dT} = 0$
Im Mittel mehr emittierte als abs. En.
$d\nu\, t\tau\left\{\frac{d\varepsilon_\nu}{dT} - q\varrho_\nu\frac{d\alpha_\nu}{dT}\right\} = d\nu\, t\,\tau\, q\frac{d\varrho_\nu}{dT}\alpha_\nu$
$\Sigma\tau$
$\left(\tau - \frac{\Sigma}{c}\tau + \alpha + \varepsilon\right)^2 = \overline{\tau^2}$
$-2\frac{\Sigma}{c}\overline{\tau^2} + \frac{\overline{\alpha^2}}{c^2} + \frac{\overline{\varepsilon^2}}{c^2} = 0$
$\overline{\alpha^2} + \overline{\varepsilon^2} = 2\frac{\Sigma}{c}\overline{\tau^2}$
$\overline{\alpha^2} + \overline{\varepsilon^2} = \frac{2R}{Nf}\underbrace{\alpha_\nu q A\, d\nu}_{\frac{A}{\varrho_\nu}} \quad T_0^2\frac{d\varrho_\nu}{dT}$
$\frac{\overline{\alpha^2} + \overline{\varepsilon^2}}{A} = \frac{2R}{N}\Sigma A\underbrace{T_0^2\frac{d\lg\varrho_\nu}{dT}}_{\frac{N}{R}h\nu} = 2h\nu$
$\frac{\overline{\alpha^2}}{A} = h\nu \quad \sqrt{\overline{\alpha^2}} = h\nu\cdot\sqrt{\frac{A}{h\nu}}$

[p. 38 L]

Relativitätstheoretische Invariante $\quad A_x A_y A_z \quad iA_x^x \quad iA_y^x \quad iA_z^x$

$$\frac{\partial A_z^x}{\partial y} - \frac{\partial A_y^x}{\partial z} - \frac{1}{c}\frac{\partial A_x}{\partial t}$$

$$j_x + \frac{1}{c}\frac{P_x}{\partial t} - \left(\frac{\partial P_z^x}{\partial y} - \frac{\partial P_y^x}{\partial z}\right) + \frac{1}{c}\frac{\partial E_x}{\partial t} - \left(\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z}\right) = 0$$

Invariantes System, das nur formal von Minkowski abweicht. Andere Bedeutung der Feldstärke.

$$\boxed{\begin{matrix} u \\ u \end{matrix}}$$

$$\frac{10^3 \cdot 4 \cdot 10^2}{9 \cdot 10^{20}} \cdot \frac{1}{2} \cdot 10^{-10}.$$

$$A - T\frac{dA}{dt} = U \qquad \frac{dU}{dt} = 0$$

$$T\frac{d^2A}{dt^2} = \frac{dU}{dt} = 0$$

$\alpha_\nu \qquad \underline{\alpha_\nu} dt$

$\boxed{}\ \varepsilon_\nu \qquad \int_{\text{Kugel}} 2K_\nu d\nu d\Omega \cdot \alpha_\nu$ absorb. Energie.

$A_o \qquad E_0$

$8\pi A_\nu d\nu\, \alpha_\nu \qquad t \qquad \varepsilon_\nu d_\nu$

$$K_\nu \cdot \frac{8\pi}{q} = \rho_\nu$$

$$q\underline{\rho_\nu}\underline{\alpha_\nu} d\nu t = A_0$$

$$\underline{A_o} + \underline{\underline{\alpha}} \qquad E_0 + \varepsilon \qquad \varepsilon_\nu d\nu t = E_0$$

$$q\underline{\rho\nu d\nu t} \cdot \left(\alpha_\nu + \frac{d\alpha_\nu}{dT}\tau\right)$$

$$\underline{d\nu t} \cdot \left(\varepsilon_\nu + \frac{d\varepsilon_\nu}{dT}\tau\right)$$

$$\overline{\tau^2} = \frac{RT_0^2}{\cancel{2}Nc}$$

$$\cancel{cT_0}\overline{\left(\cancel{T_0} + \frac{A_0 + \alpha}{c} - \frac{E_0 + \varepsilon}{c}\right)^2} = \overline{\cancel{T_0}^2}$$

$$\cancel{\left(T_0 + \frac{\alpha - \varepsilon}{c}\right)^2 = T_0^2}$$

$$\cancel{A + \frac{\partial A}{\partial T}\tau}$$

$$\cancel{E + \frac{\partial E_0}{\partial T}\tau}$$

$\cancel{2} \qquad \tau$

[p. 38 R]

⟨Unabh⟩ Wenn ρ α u ε als temp. abh. betr.

$$q\rho_\nu\alpha_\nu - \varepsilon_\nu = 0$$

$$q\left(\frac{d\rho_\nu}{dT}\alpha_\nu + \rho_\nu\frac{d\alpha_\nu}{dT}\right) \langle + \rangle - \frac{d\varepsilon_\nu}{dT} = 0$$

Im Mittel mehr emittierte als abs. En.

$$\underline{dv t\tau\left\{\frac{d\varepsilon_\nu}{dT} - q\rho_\nu\frac{d\alpha_\nu}{dT}\right\}} = dv t\tau q\frac{d\rho_\nu}{dT}\alpha_\nu$$

$$\Sigma\tau$$

$$\left(\tau - \frac{\Sigma}{c}\tau + \underline{\alpha} + \underline{\varepsilon}\right)^2 = \overline{\tau^2}$$

$$-2\frac{\Sigma}{c}\overline{\tau^2} + \frac{\overline{\alpha^2}}{c^2} + \frac{\overline{\varepsilon^2}}{c^2} = 0$$

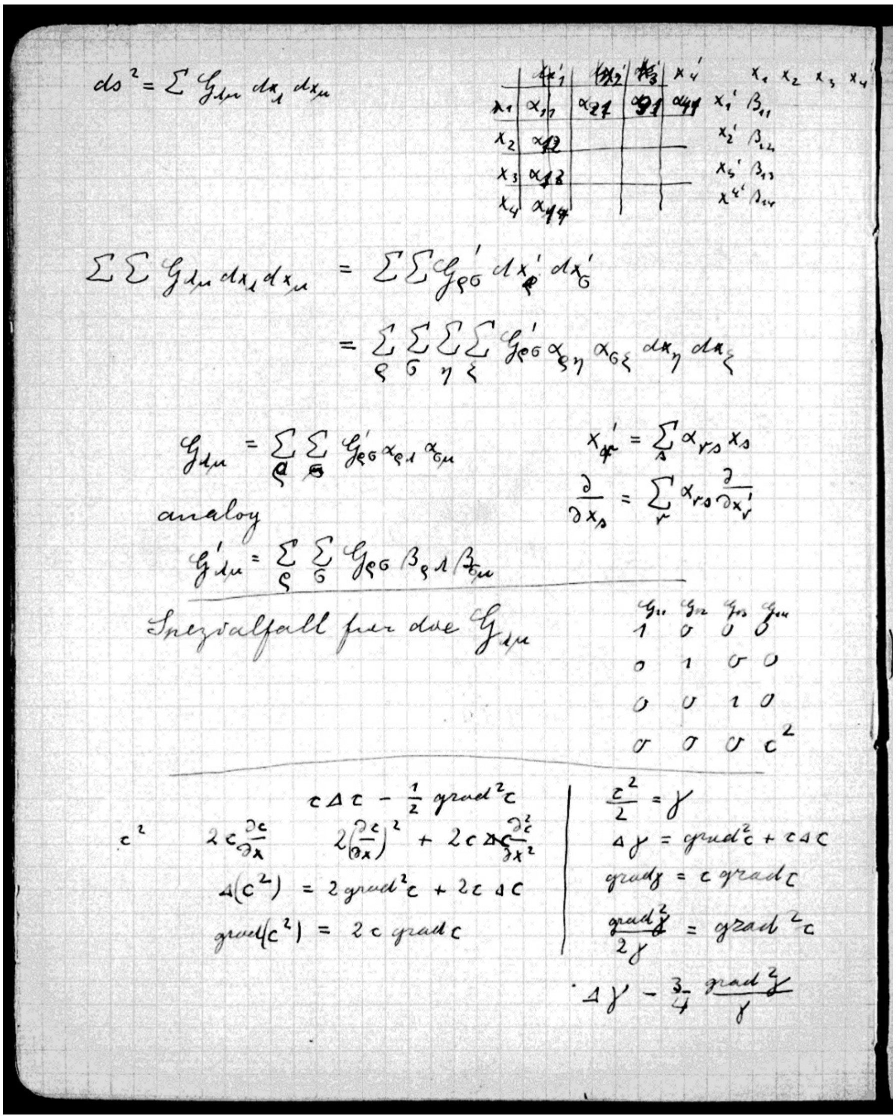
$$\overline{\alpha^2} + \overline{\varepsilon^2} = 2\frac{\Sigma}{c}\overline{\tau^2}$$

$$\overline{\alpha^2} + \overline{\varepsilon^2} = \frac{2R}{N\cancel{c}}\underbrace{\alpha_\nu qA dv}_{\frac{A}{\rho_\nu}} \qquad T_0^2\frac{d\rho_\nu}{dT}$$

$$\frac{\overline{\alpha^2} + \overline{\varepsilon^2}}{A} = \frac{2R}{N}\cancel{A}\underbrace{T_0^2\frac{d\lg\rho_\nu}{dT}}_{\frac{N}{R}hv} = 2hv$$

$$\frac{\overline{\alpha^2}}{A} = hv \qquad \boxed{\sqrt{\overline{\alpha^2}} = hv\cdot\sqrt{\frac{A}{hv}}}$$

[p. 39 L]



$$ds^2 = \sum G_{\lambda\mu}\, dx_\lambda\, dx_\mu$$

	x_1'	x_2'	x_3'	x_4'
x_1	α_{11}	α_{21}	α_{31}	α_{41}
x_2	α_{12}			
x_3	α_{13}			
x_4	α_{14}			

	x_1	x_2	x_3	x_4
x_1'	β_{11}			
x_2'	β_{12}			
x_3'	β_{13}			
x_4'	β_{14}			

$$\sum\sum G_{\lambda\mu}\, dx_\lambda\, dx_\mu = \sum\sum G'_{\rho\sigma}\, dx'_\rho\, dx'_\sigma$$

$$= \sum_\rho \sum_\sigma \sum_\eta \sum_\zeta G'_{\rho\sigma}\, \alpha_{\rho\eta}\, \alpha_{\sigma\zeta}\, dx_\eta\, dx_\zeta$$

$$G_{\lambda\mu} = \sum_\rho \sum_\sigma G'_{\rho\sigma}\, \alpha_{\rho\lambda}\, \alpha_{\sigma\mu}$$

$$x_r' = \sum_s \alpha_{rs} x_s$$

$$\frac{\partial}{\partial x_s} = \sum_r \alpha_{rs} \frac{\partial}{\partial x_r'}$$

analog

$$G'_{\lambda\mu} = \sum_\rho \sum_\sigma G_{\rho\sigma}\, \beta_{\rho\lambda}\, \beta_{\sigma\mu}$$

Spezialfall fuer die $G_{\lambda\mu}$

G_{11}	G_{22}	G_{33}	G_{44}
1	0	0	0
0	1	0	0
0	0	1	0
0	0	0	c^2

$$c\Delta c - \frac{1}{2}\,\mathrm{grad}^2 c$$

$$c^2 \qquad 2c\frac{\partial c}{\partial x} \qquad 2\left(\frac{\partial c}{\partial x}\right)^2 + 2c\frac{\partial^2 c}{\partial x^2}$$

$$\Delta(c^2) = 2\,\mathrm{grad}^2 c + 2c\,\Delta c$$

$$\mathrm{grad}(c^2) = 2c\,\mathrm{grad}\, c$$

$$\frac{c^2}{2} = \gamma$$

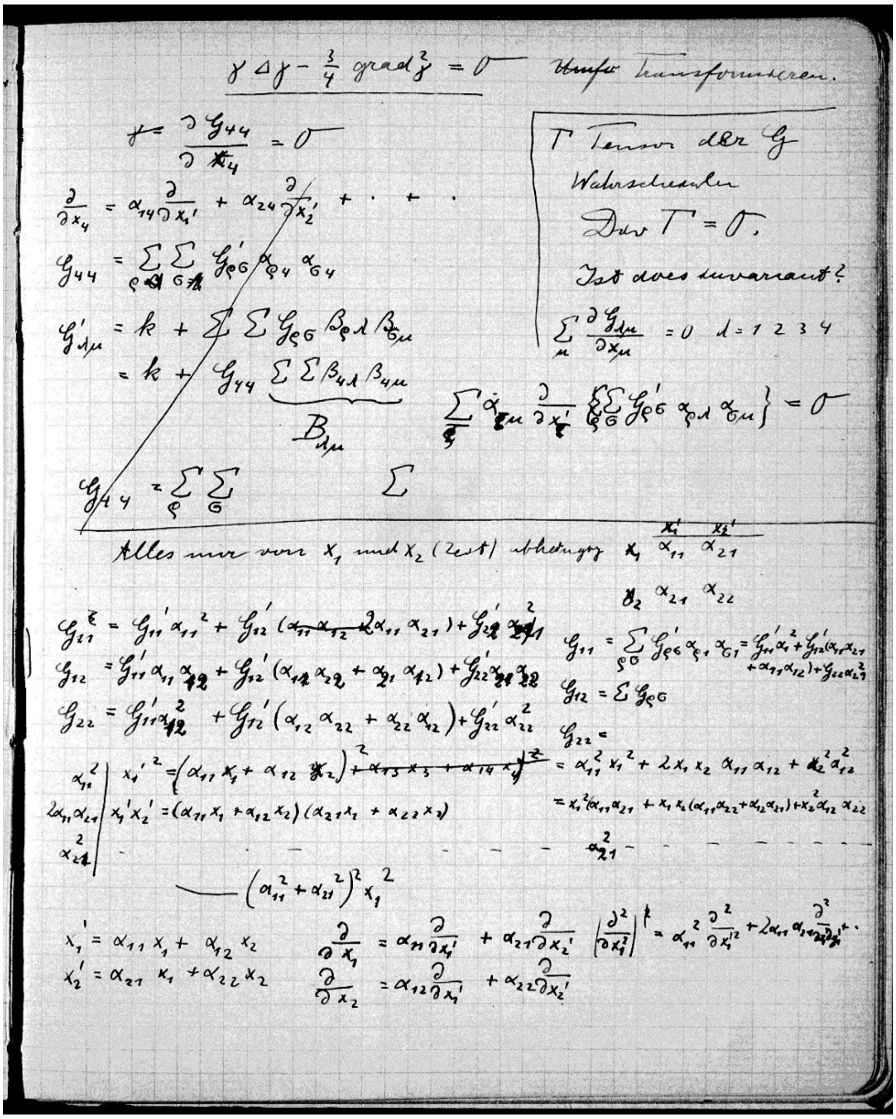
$$\Delta\gamma = \mathrm{grad}^2 c + c\Delta c$$

$$\mathrm{grad}\,\gamma = c\,\mathrm{grad}\, c$$

$$\frac{\mathrm{grad}^2 \gamma}{2\gamma} = \mathrm{grad}^2 c$$

$$\Delta\gamma - \frac{3}{4}\frac{\mathrm{grad}^2\gamma}{\gamma}$$

[p. 39 R]



$\gamma \Delta \gamma - \frac{3}{4} \mathrm{grad}^2 \gamma = 0$ Umfo Transformieren.

$\gamma = \frac{\partial G_{44}}{\partial x_4} = 0$

$\frac{\partial}{\partial x_4} = \alpha_{14} \frac{\partial}{\partial x_1'} + \alpha_{24} \frac{\partial}{\partial x_2'} + \cdot + \cdot$

$G_{44} = \sum_\rho \sum_\sigma G'_{\rho\sigma} \alpha_{\rho 4} \alpha_{\sigma 4}$

$G'_{\lambda\mu} = k + \sum \sum G_{\rho\sigma} \beta_{\rho\lambda} \beta_{\sigma\mu}$

$= k + G_{44} \underbrace{\sum \sum \beta_{4\lambda} \beta_{4\mu}}_{B_{\lambda\mu}}$

$G_{44} = \sum_\rho \sum_\sigma \qquad \sum$

Γ Tensor der G
Wahrscheinlich
Div Γ = 0.
Ist dies invariant?

$\sum_\mu \frac{\partial G_{\lambda\mu}}{\partial x_\mu} = 0 \quad \lambda = 1\ 2\ 3\ 4$

$\sum \alpha_{\xi\mu} \frac{\partial}{\partial x'_\xi} \left\{ \sum_{\rho\sigma} G'_{\rho\sigma} \alpha_{\rho\lambda} \alpha_{\sigma\mu} \right\} = 0$

Alles nur von x_1 und x_2 (Zeit) abhängig

$$\begin{array}{c|cc} & x_1' & x_2' \\ \hline x_1 & \alpha_{11} & \alpha_{21} \\ x_2 & \alpha_{21} & \alpha_{22} \end{array}$$

$G_{11} = G'_{11}\alpha_{11}^2 + G'_{12}(2\alpha_{11}\alpha_{21}) + G'_{22}\alpha_{21}^2$

$G_{12} = G'_{11}\alpha_{11}\alpha_{12} + G'_{12}(\alpha_{11}\alpha_{22} + \alpha_{21}\alpha_{12}) + G'_{22}\alpha_{21}\alpha_{22}$

$G_{22} = G'_{11}\alpha_{12}^2 + G'_{12}(\alpha_{12}\alpha_{22} + \alpha_{22}\alpha_{12}) + G'_{22}\alpha_{22}^2$

$G_{11} = \sum_{\rho\sigma} G'_{\rho\sigma}\alpha_{\rho 1}\alpha_{\sigma 1} = G'_{11}\alpha_1^2 + G'_{12}(\alpha_{11}\alpha_{21} + \alpha_{11}\alpha_{12}) + G_{22}\alpha_{21}^2$

$G_{12} = \sum G_{\rho\sigma}$

$G_{22} =$

$\alpha_{11}^2 \mid x_1'^2 = (\alpha_{11}x_1 + \alpha_{12}x_2)^2 = \alpha_{11}^2 x_1^2 + 2x_1x_2\alpha_{11}\alpha_{12} + x_2^2\alpha_{12}^2$

$2\alpha_{11}\alpha_{21} \mid x_1'x_2' = (\alpha_{11}x_1 + \alpha_{12}x_2)(\alpha_{21}x_1 + \alpha_{22}x_2)$

$= x_1^2\alpha_{11}\alpha_{21} + x_1x_2(\alpha_{11}\alpha_{22} + \alpha_{12}\alpha_{21}) + x_2^2\alpha_{12}\alpha_{22}$

$\alpha_{21}^2 \mid - \quad - \quad - \quad \alpha_{21}^2 - - - -$

$(\alpha_{11}^2 + \alpha_{21}^2)^2 x_1^2$

$x_1' = \alpha_{11}x_1 + \alpha_{12}x_2$

$x_2' = \alpha_{21}x_1 + \alpha_{22}x_2$

$\frac{\partial}{\partial x_1} = \alpha_{11}\frac{\partial}{\partial x_1'} + \alpha_{21}\frac{\partial}{\partial x_2'}$

$\frac{\partial}{\partial x_2} = \alpha_{12}\frac{\partial}{\partial x_1'} + \alpha_{22}\frac{\partial}{\partial x_2'}$

$\left(\frac{\partial^2}{\partial x_1^2}\right) = \alpha_{11}^2 \frac{\partial^2}{\partial x_1'^2} + 2\alpha_{11}\alpha_{21}\frac{\partial^2}{\partial x_1'\partial x_2'} + \cdot$

[p. 39 L]

$$ds^2 = \sum G_{\lambda\mu} dx_\lambda dx_\mu$$

	x'_1	x'_2	x'_3	x'_4
x_1	α_{11}	α_{21}	α_{31}	α_{41}
x_2	α_{12}			
x_3	α_{13}			
x_4	α_{14}			

	x_1	x_2	x_3	x_4
x'_1	β_{11}			
x'_2	β_{12}			
x'_3	β_{13}			
x'_4	β_{14}			

$$\sum\sum G_{\lambda\mu} dx_\lambda dx_\mu = \sum\sum G'_{\rho\sigma} dx'_\rho dx'_\sigma$$

$$= \sum_\rho \sum_\sigma \sum_\eta \sum_\zeta G'_{\rho\sigma} \alpha_{\rho\eta} \alpha_{\sigma\zeta} dx_\eta dx_\zeta$$

$$G_{\lambda\mu} = \sum_\rho \sum_\sigma G'_{\rho\sigma} \alpha_{\rho\lambda} \alpha_{\sigma\mu} \qquad x'_r = \sum_s \alpha_{rs} x_s$$

analog $\qquad \dfrac{\partial}{\partial x_s} = \sum_r \alpha_{rs} \dfrac{\partial}{\partial x'_r}$

$$G'_{\lambda\mu} = \sum_\rho \sum_\sigma G_{\rho\sigma} \beta_{\rho\lambda} \beta_{\sigma\mu}$$

Spezialfall für die $G_{\lambda\mu}$

G_{11}	G_{12}	G_{13}	G_{14}
1	0	0	0
0	1	0	0
0	0	1	0
0	0	0	c^2

$$c\Delta c - \frac{1}{2}\mathrm{grad}^2 c$$

$$c^2 \qquad 2c\frac{\partial c}{\partial x} \qquad 2c\left(\frac{\partial c}{\partial x}\right)^2 + 2c\cancel{\Delta} c\frac{\partial^2 c}{\partial x^2}$$

$$\Delta(c^2) = 2\,\mathrm{grad}^2 c + 2c\Delta c$$

$$\mathrm{grad}(c^2) = 2c\ \mathrm{grad}\ c$$

$$\frac{c^2}{2} = \gamma$$

$$\Delta\gamma = \mathrm{grad}^2 c + c\Delta c$$

$$\mathrm{grad}\ \gamma = c\ \mathrm{grad}\ c$$

$$\frac{\mathrm{grad}^2\gamma}{2\gamma} = \mathrm{grad}\ ^2 c$$

$$\Delta\gamma - \frac{3}{4}\frac{\mathrm{grad}^2\gamma}{\gamma}$$

[p. 39 R]

$$\underline{\gamma\Delta\gamma - \frac{3}{4}\operatorname{grad}^2\gamma = 0}$$

<Umfo> Transformieren.

$$\cancel{\gamma =} \frac{\partial G_{44}}{\partial x_4} = 0$$

$$\frac{\partial}{\partial x_4} = \alpha_{14}\frac{\partial}{\partial x_1'} + \alpha_{24}\frac{\partial}{\partial x_2'} + \cdot + \cdot$$

$$G_{44} = \sum_\rho\sum_\sigma G'_{\rho\sigma}\alpha_{\rho 4}\alpha_{\sigma 4}$$

$$G'_{\lambda\mu} = k + \sum\sum G_{\rho\sigma}\beta_{\rho\lambda}\beta_{\sigma\mu}$$

$$= k + G_{44}\underbrace{\sum\sum\beta_{4\lambda}\beta_{4\mu}}_{B_{\lambda\mu}}$$

$$G_{44} = \sum_\rho\sum_\sigma \qquad\qquad \sum$$

Γ Tensor der G

Wahrscheinln

Div $\Gamma = 0$.

Ist dies invariant?

$$\sum_\mu \frac{\partial G_{\lambda\mu}}{\partial x_\mu} = 0 \quad \lambda = 1\ 2\ 3\ 4$$

$$\sum_\tau \alpha_{\tau u}\frac{\partial}{\partial x_\tau'}\left\{\sum_\rho\sum_\sigma G'_{\rho\sigma}\alpha_{\rho\lambda}\alpha_{\sigma\mu}\right\} = 0$$

Alles nur von x_1 und x_2 (Zeit) abhängig

	x_1'	x_2'
x_1	α_{11}	α_{21}
x_2	α_{21}	α_{22}

$$G_{11} = G'_{11}\,\alpha_{11}^2 + G'_{12}\,(\cancel{\alpha_{11}\alpha_{12} +} 2\alpha_{11}\alpha_{21}) + G'_{22}\,\alpha_{21}^2$$

$$G_{12} = G'_{11}\,\alpha_{11}\alpha_{12} + G'_{12}\,(\alpha_{11}\alpha_{22} + \alpha_{21}\alpha_{12}) + G'_{22}\,\alpha_{21}\alpha_{22}$$

$$G_{22} = G'_{11}\,\alpha_{12}^2 + G'_{12}\,(\alpha_{12}\alpha_{22} + \alpha_{22}\alpha_{12}) + G'_{22}\,\alpha_{22}^2$$

$$G_{11} = \sum_{\rho\sigma} G'_{\rho\sigma}\,\alpha_{\rho 1}\alpha_{\sigma 1} = G'_{11}\,\alpha_1^2 + G'_{12}\,(\alpha_{11}\alpha_{21} + \alpha_{11}\alpha_{12}) + G_{22}\alpha_{21}^2$$

$$G_{12} = \sum G_{\rho\sigma}$$

$$G_{22} =$$

$$\begin{array}{r|l}
\alpha_{11}^2 & x_1'^2 = (\alpha_{11}x_1 + \alpha_{12}x_2)^2\langle+\alpha_{13}x_3 + \alpha_{14}x_4)^2\rangle = \alpha_{11}^2x_1^2 + 2x_1x_2\alpha_{11}\alpha_{12} + x_2^2\alpha_{12}^2 \\
2\alpha_{11}\alpha_{21} & x_1'x_2' = (\alpha_{11}x_1 + \alpha_{12}x_2)(\alpha_{21}x_1 + \alpha_{22}x_2) \quad = x_1^2(\alpha_{11}\alpha_{21} + x_1x_2(\alpha_{11}\alpha_{22} + \alpha_{12}\alpha_{21}) + x_2^2\alpha_{12}\alpha_{22} \\
\alpha_{21}^2 & - \quad - \quad - \quad - \quad - \quad \alpha_{21}^2 \quad - \quad - \quad -
\end{array}$$

$$(\alpha_{11}^2 + \alpha_{21}^2)^2x_1^2$$

$$x_1' = \alpha_{11}x_1 + \alpha_{12}x_2 \qquad \frac{\partial}{\partial x_1} = \alpha_{11}\frac{\partial}{\partial x_1'} + \alpha_{21}\frac{\partial}{\partial x_2'} \qquad \left(\frac{\partial^2}{\partial x_1^2}\right)^{\langle 2\rangle} = \alpha_{11}^2\frac{\partial^2}{\partial x_1'^2} + 2\alpha_{11}\alpha_{21}\frac{\partial^2}{\partial x_1'\,\partial y_1'} + \cdot$$

$$x_2' = \alpha_{21}x_1 + \alpha_{22}x_2 \qquad \frac{\partial}{\partial x_2} = \alpha_{12}\frac{\partial}{\partial x_1'} + \alpha_{22}\frac{\partial}{\partial x_2'}$$

[p. 40 L]

$\cancel{\varphi \Delta \Delta \Delta \varphi}$

$\cancel{\mathrm{grad}^2\varphi \; \mathrm{grad}^2 \, \mathrm{grad}^2\varphi}$

$\Delta\varphi \quad \varphi\Delta\varphi \quad \mathrm{grad}^2\varphi$ 2. Ordnung

$\Delta\Delta\varphi \quad \Delta(\varphi\Delta\varphi) \quad \Delta(\mathrm{grad}^2\varphi) \quad \varphi(\Delta\Delta\varphi) \quad \mathrm{grad}^2\Delta\varphi$ 4. Ordnung

$$\frac{\partial^2}{\partial x^2}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right) + \frac{\partial^2}{\partial y^2}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)$$

$$\frac{\partial^4}{\partial x^4} + 2\frac{\partial^4}{\partial x^2 \partial y^2} + \frac{\partial^4}{\partial y^4} = \underline{\Delta\Delta\varphi}$$

$$\underline{\Delta\varphi \cdot \Delta\varphi} + \varphi\Delta\Delta\varphi + \underbrace{\left(\frac{\partial\varphi}{\partial x}\frac{\partial\Delta\varphi}{\partial x} + \cdot + \cdot\right)}_{\mathrm{grad}\,\varphi \; \mathrm{grad}\,\Delta\varphi}$$

Die ersten 2 Schritte

2 Dimensionen

System der G äquivalent dem System $\frac{\partial^2\varphi}{\partial x_\mu \partial x_\nu}$

Gleichung soll so sein, dass in jedem Glied nach allen x gleich oft diff. wird.

Linear unmöglich von 8. Ordnung in φ

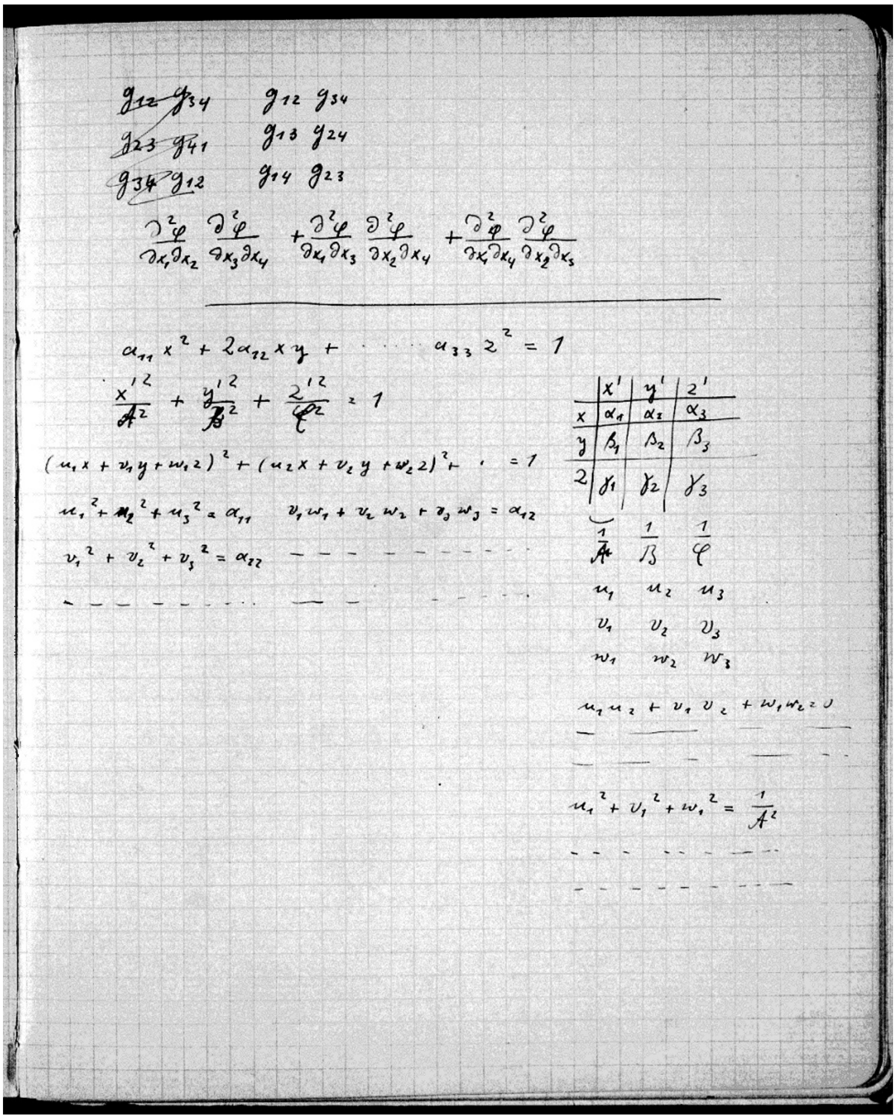
Quadratisch

$\frac{\partial^2\varphi}{\partial x_1^2} \; \frac{\partial^6\varphi}{\partial x_2^2 \partial x_3^2 \partial x_4^2}$ etc. wird notwendig 4. Ordnung.

dritten Grades in φ wird 2. Ordnung, wie es sein muss.

$\frac{\partial^2\varphi}{\partial x_1^2} \; \frac{\partial^2\varphi}{\partial x_2^2} \; \frac{\partial^4\varphi}{\partial x_3^2 \partial x_4^2}$

[p. 40 R]



$\cancel{g_{12}\,g_{34}}$ $\qquad g_{12}\,g_{34}$

$\cancel{g_{23}\,g_{41}}$ $\qquad g_{13}\,g_{24}$

$\cancel{g_{34}\,g_{12}}$ $\qquad g_{14}\,g_{23}$

$$\frac{\partial^2\varphi}{\partial x_1\partial x_2}\frac{\partial^2\varphi}{\partial x_3\partial x_4}+\frac{\partial^2\varphi}{\partial x_1\partial x_3}\frac{\partial^2\varphi}{\partial x_2\partial x_4}+\frac{\partial^2\varphi}{\partial x_1\partial x_4}\frac{\partial^2\varphi}{\partial x_2\partial x_3}$$

$$\alpha_{11}x^2+2\alpha_{12}xy+\cdots\alpha_{33}z^2=1$$

$$\frac{x'^2}{A^2}+\frac{y'^2}{B^2}+\frac{z'^2}{C^2}=1$$

$$(u_1x+v_1y+w_1z)^2+(u_2x+v_2y+w_2z)^2+\cdot=1$$

$u_1^2+u_2^2+u_3^2=\alpha_{11}$ $\qquad v_1w_1+v_2w_2+v_3w_3=\alpha_{12}$

$v_1^2+v_2^2+v_3^2=\alpha_{22}$ $\qquad$ — — — — — —

— — — — — — — — — —

	x'	y'	z'
x	α_1	α_2	α_3
y	β_1	β_2	β_3
z	γ_1	γ_2	γ_3

$\frac{1}{A}$ $\quad \frac{1}{B}$ $\quad \frac{1}{C}$

$u_1 \quad u_2 \quad u_3$

$v_1 \quad v_2 \quad v_3$

$w_1 \quad w_2 \quad w_3$

$u_1u_2+v_1v_2+w_1w_2=0$

— — — — —

— — — — —

$u_1^2+v_1^2+w_1^2=\frac{1}{A^2}$

— — — — —

— — — — —

[p. 40 L]

$$\langle \varphi\Delta\Delta\varphi \rangle$$

$$\langle \mathrm{grad}^2\varphi \quad \mathrm{grad}^2 \quad \mathrm{grad}^2\varphi \rangle$$

$$\Delta\varphi \qquad \varphi\Delta\varphi \qquad \mathrm{grad}^2\varphi \qquad \text{2. Ordnung}$$

$$\Delta\Delta\varphi \qquad \Delta(\varphi\Delta\varphi) \qquad \Delta(\mathrm{grad}^2\varphi) \qquad \varphi(\Delta\Delta\varphi) \qquad \mathrm{grad}^2\Delta\varphi \qquad \text{4. Ordnung}$$

$$\frac{\partial^2}{\partial x^2}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \langle + \frac{\partial^2}{\partial z^2} \rangle\right) + \frac{\partial^2}{\partial y^2}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)$$

$$\frac{\partial^4}{\partial x^4} + 2\frac{\partial^4}{\partial x^2 \partial y^2} + \frac{\partial^4}{\partial y^4} = \underline{\Delta\Delta\varphi}$$

$$\underline{\Delta\varphi \cdot \Delta\varphi} + \varphi\Delta\Delta\varphi + \underbrace{\left(\frac{\partial\varphi}{\partial x}\frac{\partial\Delta\varphi}{\partial x} + \cdot + \cdot\right)}_{\mathrm{grad}\varphi \ \mathrm{grad}\Delta\varphi}$$

Die ersten 2 Schritte

---

2 Dimensionen

System der G äquivalent dem System $\dfrac{\partial^2\varphi}{\partial x_\mu \partial x_\nu}$

Gleichung soll so sein, dass in jedem Glied nach allen x gleich oft diff. wird.

Linear unmöglich von 8. Ordnung in φ

Quadratisch

$$\frac{\partial^2\varphi}{\partial x_1^2}\frac{\partial^6\varphi}{\partial x_2^2 \partial x_3^2 \partial x_4^2} \quad \text{etc. wird notwendig 4. Ordnung.}$$

dritten Grades in φ wird 2. Ordnung, wie es sein muss.

$$\frac{\partial^2\varphi}{\partial x_1^2}\frac{\partial^2\varphi}{\partial x_2^2}\frac{\partial^4\varphi}{\partial x_3^2 \partial x_4^2}$$

[p. 40 R]

$$\begin{array}{ll} g_{12}\, \cancel{g_{34}} & g_{12}\, g_{34} \\ \cancel{g_{23}}\, g_{41} & g_{13}\, g_{24} \\ \cancel{g_{34}}\, g_{12} & g_{14}\, g_{23} \end{array}$$

$$\frac{\partial^2\varphi}{\partial x_1 \partial x_2}\frac{\partial^2\varphi}{\partial x_3 \partial x_4} + \frac{\partial^2\varphi}{\partial x_1 \partial x_3}\frac{\partial^2\varphi}{\partial x_2 \partial x_4} + \frac{\partial^2\varphi}{\partial x_1 \partial x_4}\frac{\partial^2\varphi}{\partial x_2 \partial x_3}$$

$$\alpha_{11}x^2 + 2\alpha_{12}xy + \cdot\cdot\cdot\cdot\cdot\ \alpha_{33}z^2 = 1$$

$$\frac{x'^2}{A^2} + \frac{y'^2}{B^2} + \frac{z'^2}{C^2} = 1$$

	x'	y'	z'
x	α_1	α_2	α_3
y	β_1	β_2	β_3
z	γ_1	γ_2	γ_3

$$(u_1x + v_1y + w_1z)^2 + (u_2x + v_2y + w_2z)^2 + \cdot = 1$$

$$\begin{array}{ccc} \frac{1}{A} & \frac{1}{B} & \frac{1}{C} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{array}$$

$$u_1^2 + u_2^2 + u_3^2 = \alpha_{11} \qquad v_1w_1 + v_2w_2 + v_3w_3 = \alpha_{12}$$

$$v_1^2 + v_2^2 + v_3^2 = \alpha_{22} \qquad \text{— — — — — — — —}$$

$$\text{— — — — — —} \qquad \text{— — — — — — — —}$$

$$u_1u_2 + v_1v_2 + w_1w_2 = 0$$

— — — — — — — —

— — — — — — — —

$$u_1^2 + v_1^2 + w_1^2 = \frac{1}{A^2}$$

— — — — — —

— — — — — —

[p. 41 L]

Die u v w bestimmen Lage und Grösse des Ellipsoids.

Beliebige (lineare) Transformation der x y z in x'y'z'

Bei invarianter Ellipsoidfunktion

$$\sum(u_1 x + v_1 y + w_1 z)^2 = \sum(u_1' x' + v_1' y' + w_1' z')^2 \cancel{+ \cdots}$$

$$\cancel{u_1'^2 + v_1'^2 + w_1'^2 = \sum(u_1}$$

$$u_1'^2 + u_2'^2 + u_3'^2 = \sum(u_1\beta_{11} + v_1\beta_{21} + w_1\beta_{31})^2$$

$$v_1'^2 + v_2'^2 + v_3'^2 = \sum u_1\beta_{12} + v_1\beta_{22} + w_1\beta_{32})^2$$

— — — — — — — — — —

$$\begin{array}{cccc} & x' & y' & z' \\ x & \beta_{11} & \beta_{21} & \beta_{31} \\ y & \beta_{12} & \beta_{22} & \beta_{32} \\ z & \beta_{13} & \beta_{23} & \beta_{33}. \end{array}$$

$$\begin{array}{cccc} & x & y & z \\ x' & \alpha_{11} & \alpha_{21} & \alpha_{31} \\ y' & \alpha_{12} & \alpha_{22} & \alpha_{32} \\ z' & \alpha_{13} & \alpha_{23} & \alpha_{33} \end{array}$$

Einfachste Substitutionen, deren Determinante = 1.

$$dx_\nu' = dx_\nu + \sum p_{\nu\sigma}^{x} dx_\sigma$$

X_ν sind homogen u. zweiten Grades in den Koordinaten.

Es werden nur zwei Koordinaten transformiert

$$\frac{\partial X}{\partial x} + \frac{\partial Y}{\partial y} = 0$$

$$\begin{array}{cc} 1 & y \\ \cdot & 1 \\ y & x \\ \cancel{x} & -y \end{array}$$

$$X = \frac{\partial \psi}{\partial y} \qquad p_{11}^{x} = \frac{\partial^2 \psi}{\partial x \partial y} \qquad p_{12}^{x} = \frac{\partial^2 \psi}{\partial y^2}$$

$$Y = -\frac{\partial \psi}{\partial x} \qquad p_{21}^{x} = -\frac{\partial^2 \psi}{\partial x^2} \qquad p_{22}^{x} = -\frac{\partial^2 \psi}{\partial x \partial y}$$

$$\psi = r^3$$

$$\frac{\partial \psi}{\partial x} = rx \qquad \frac{\partial^2 \psi}{\partial x^2} = r + \frac{x^2}{r}$$

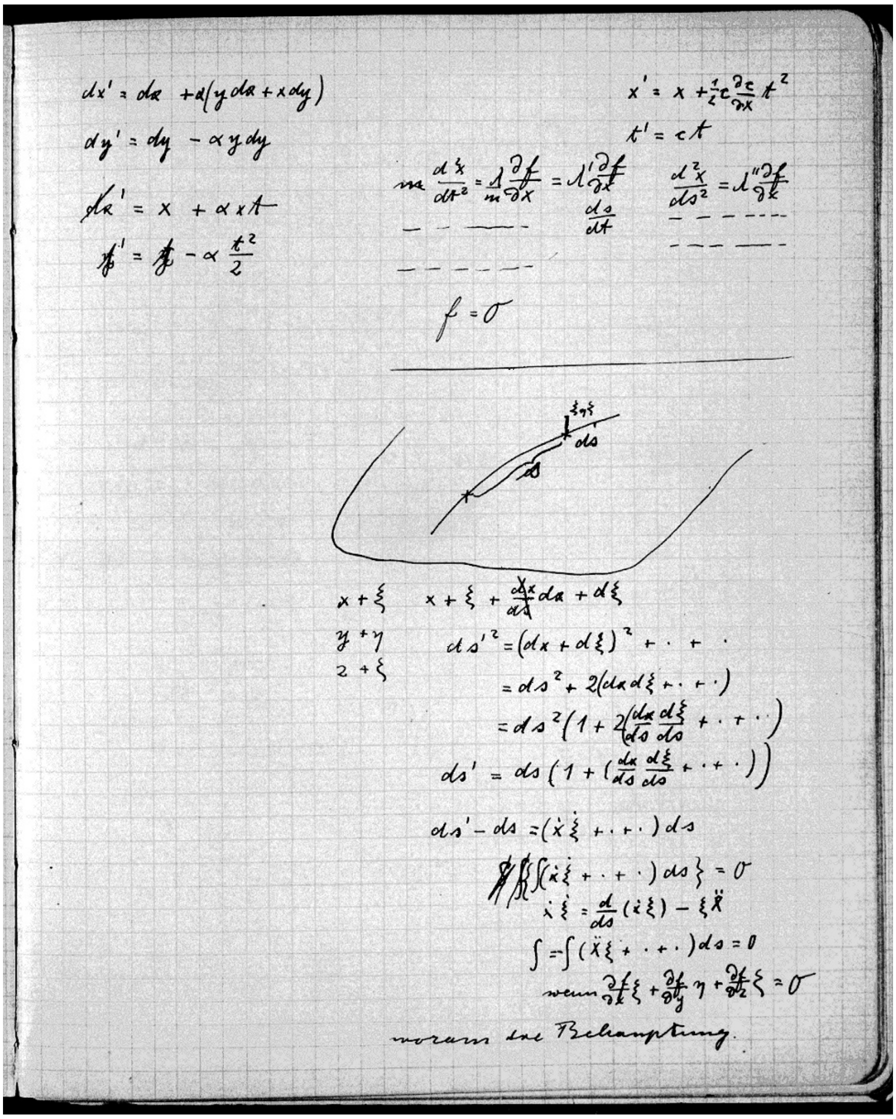
$$\frac{\partial \psi}{\partial y} = ry \qquad \frac{\partial^2 \psi}{\partial x \partial y} = \frac{xy}{r} \qquad \frac{\partial^2 \psi}{\partial y^2} = r + \frac{y^2}{r}$$

$$\psi = r^2 x$$

$$\frac{\partial \psi}{\partial x} = r^2 + 2x^2 \qquad \frac{\partial^2 \psi}{\partial x^2} = 6x$$

$$\frac{\partial \psi}{\partial y} = 2xy \qquad \frac{\partial^2 \psi}{\partial x \partial y} = 2y \qquad \frac{\partial^2 \psi}{\partial y^2} = 2x$$

[p. 41 R]



$dx' = dx + \alpha(y\,dx + x\,dy)$

$dy' = dy - \alpha y\,dy$

$\cancel{dx}' = x + \alpha x t$

$\cancel{y}' = \cancel{y} - \alpha \frac{t^2}{2}$

$x' = x + \frac{1}{2}c\frac{\partial c}{\partial x}t^2$

$t' = ct$

$m\frac{d^2x}{dt^2} = \frac{1}{m}\frac{\partial f}{\partial x} = \lambda'\frac{\partial f}{\partial x}$ $\qquad \frac{d^2x}{ds^2} = \lambda''\frac{\partial f}{\partial x}$

$\frac{ds}{dt}$

$f = 0$

$x + \xi \qquad x + \xi + \frac{dx}{ds}dx + d\xi$

$y + \eta \qquad ds'^2 = (dx + d\xi)^2 + \cdot + \cdot$

$z + \zeta$

$= ds^2 + 2(dx\,d\xi + \cdot + \cdot)$

$= ds^2\left(1 + 2\left(\frac{dx}{ds}\frac{d\xi}{ds} + \cdot + \cdot\right)\right)$

$ds' = ds\left(1 + \left(\frac{dx}{ds}\frac{d\xi}{ds} + \cdot + \cdot\right)\right)$

$ds' - ds = (\dot{x}\dot{\xi} + \cdot + \cdot)\,ds$

$\delta\left\{\int(\dot{x}\dot{\xi} + \cdot + \cdot)\,ds\right\} = 0$

$\dot{x}\dot{\xi} = \frac{d}{ds}(\dot{x}\xi) - \xi\ddot{x}$

$\int = -\int(\ddot{x}\xi + \cdot + \cdot)\,ds = 0$

wenn $\frac{\partial f}{\partial x}\xi + \frac{\partial f}{\partial y}\eta + \frac{\partial f}{\partial z}\zeta = 0$

woraus die Behauptung.

[p. 41 L]

Die uvw bestimmen Lage und Grösse des Ellipsoids.
Beliebige lineare Transformation der xyz in $x'y'z'$
bei invarianter Ellipsoidfunktion

$$\sum(u_1x+v_1y+w_1z)^2 \quad = \sum(u_1'x'+v_1'y'+w_1'z')^2 \quad \cancel{+ \quad \cdot \quad + \quad \cdot}$$

$$\langle u_1'^2+v_1'^2+w_1'^2 = \sum(u_1\rangle$$

$$u_1'^2+u_2'^2+u_3'^2 = \sum(u_1\beta_{11}+v_1\beta_{21}+w_1\beta_{31})^2$$

$$v_1'^2+v_2'^2+v_3'^2 = \sum(u_1\beta_{12}+v_1\beta_{22}+w_1\beta_{32})^2$$

— — — — — — — — — — — — —

$$\begin{array}{c|ccc} & x' & y' & z' \\ \hline x & \beta_{11} & \beta_{21} & \beta_{31} \\ y & \beta_{12} & \beta_{22} & \beta_{32} \\ z & \beta_{13} & \beta_{23} & \beta_{33} \end{array}$$

$$\begin{array}{c|ccc} & x & y & z \\ \hline x' & \alpha_{11} & \alpha_{21} & \alpha_{31} \\ y' & \alpha_{12} & \alpha_{22} & \alpha_{32} \\ z' & \alpha_{13} & \alpha_{23} & \alpha_{33} \end{array}$$

Einfachste Substitutionen, deren Determinate = 1.

$$dx_\nu' = dx_\nu + \sum p_{\nu\sigma}^x dx_\sigma$$

X_ν sind homogen u zweiten Grades in den Koordinaten.

Es werden nur zwei Koordinaten transformiert

$$\frac{\partial X}{\partial x}+\frac{\partial Y}{\partial y} = 0$$

$$\begin{array}{cc} 1 & y \\ \cdot & 1 \\ y & x \\ \langle -x\rangle & -y \end{array}$$

$$X = \frac{\partial\psi}{\partial y} \qquad \rho_{11}^x = \frac{\partial^2\psi}{\partial x\partial y} \qquad \rho_{12}^x = \frac{\partial^2\psi}{\partial y^2}$$

$$Y = -\frac{\partial\psi}{\partial x} \qquad \rho_{21}^x = -\frac{\partial^2\psi}{\partial x^2} \qquad \rho_{22}^x = -\frac{\partial^2\psi}{\partial x\partial y}$$

$$\psi = r^3 \qquad \frac{\partial^2\psi}{\partial x^2} = r+\frac{x^2}{r}$$

$$\frac{\partial\psi}{\partial x} = rx \qquad \frac{\partial^2\psi}{\partial x\partial y} = \frac{xy}{r}$$

$$\frac{\partial\psi}{\partial y} = ry \qquad \frac{\partial^2\psi}{\partial y^2} = r+\frac{y^2}{r}$$

$$\psi = r^2x \qquad \frac{\partial^2\psi}{\partial x^2} = 6x$$

$$\frac{\partial\psi}{\partial x} = r^2+2x^2 \qquad \frac{\partial^2\psi}{\partial x\partial y} = 2y$$

$$\frac{\partial\psi}{\partial y} = 2xy \qquad \frac{\partial^2\psi}{\partial y^2} = 2x$$

[p. 41 R]

$$dx' = dx + \alpha(ydx + xdy)$$

$$dy' = dy - \alpha ydy$$

$$\cancel{d}x' = x + \alpha xt$$

$$t' = t - \alpha\frac{t^2}{2}$$

$$x' = x + \frac{1}{2}c\frac{\partial c}{\partial x}t^2$$

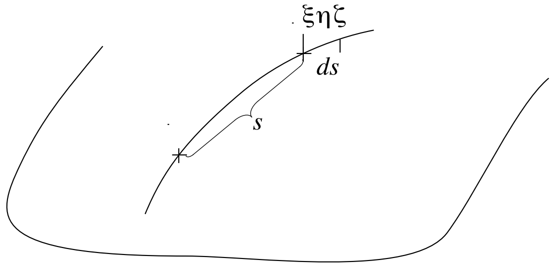
$$t' = ct$$

$$\cancel{m}\frac{d^2x}{dt^2} = \frac{\lambda}{m}\frac{\partial f}{\partial x} = \lambda'\frac{\partial f}{\partial x} \qquad \frac{d^2x}{ds^2} = \lambda''\frac{\partial f}{\partial x}$$

— — — — — $\frac{ds}{dt}$ — — — — —

— — — — — — — — — —

$$f = 0$$



$$x + \xi \qquad x + \xi + \frac{\cancel{dx}}{\cancel{ds}}dx + d\xi$$

$$y + \eta \qquad ds'^2 = (dx + d\xi)^2 + \cdot + \cdot$$

$$z + \zeta \qquad = ds^2 + 2(dxd\xi + \cdot + \cdot)$$

$$= ds^2\left(1 + 2\left(\frac{dx}{ds}\frac{d\xi}{ds} + \cdot + \cdot\right)\right.$$

$$ds' = ds\left(1 + \left(\frac{dx}{ds}\frac{d\xi}{ds} + \cdot + \cdot\right)\right)$$

$$ds' - ds = (\dot{x}\dot{\xi} + \cdot + \cdot)ds$$

$$\langle\delta\int\rangle\left\{\int(\dot{x}\dot{\xi} + \cdot + \cdot)ds\right\} = 0$$

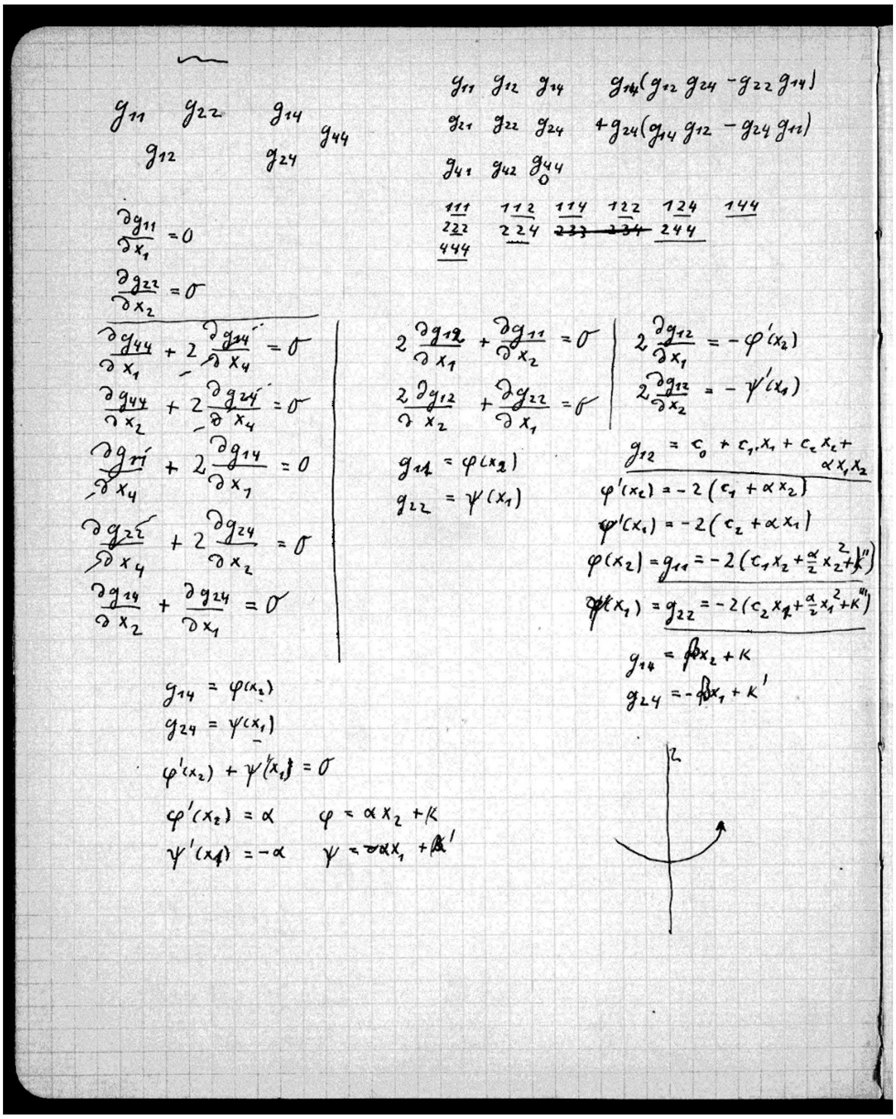
$$\dot{x}\dot{\xi}\frac{d}{ds}(\dot{x}\xi) - \xi\ddot{x}$$

$$\int = \int(\ddot{x}\xi + \cdot + \cdot)ds = 0$$

wenn $\frac{\partial f}{\partial x}\xi + \frac{\partial f}{\partial y}\eta + \frac{\partial f}{\partial z}\zeta = 0$

woraus die Behauptung.

[p. 42 L]



g_{11} g_{22} g_{14} g_{44}

g_{12} g_{24}

$$\frac{\partial g_{11}}{\partial x_1} = 0$$

$$\frac{\partial g_{22}}{\partial x_2} = 0$$

$$\frac{\partial g_{44}}{\partial x_1} + 2\frac{\partial g_{14}}{\partial x_4} = 0$$

$$\frac{\partial g_{44}}{\partial x_2} + 2\frac{\partial g_{24}}{\partial x_4} = 0$$

$$\frac{\partial g_{11}}{\partial x_4} + 2\frac{\partial g_{14}}{\partial x_1} = 0$$

$$\frac{\partial g_{22}}{\partial x_4} + 2\frac{\partial g_{24}}{\partial x_2} = 0$$

$$\frac{\partial g_{14}}{\partial x_2} + \frac{\partial g_{24}}{\partial x_1} = 0$$

$$2\frac{\partial g_{12}}{\partial x_1} + \frac{\partial g_{11}}{\partial x_2} = 0$$

$$2\frac{\partial g_{12}}{\partial x_2} + \frac{\partial g_{22}}{\partial x_1} = 0$$

$$g_{11} = \varphi(x_2)$$

$$g_{22} = \psi(x_1)$$

$$\begin{matrix} g_{11} & g_{12} & g_{14} \\ g_{21} & g_{22} & g_{24} \\ g_{41} & g_{42} & \underset{0}{g_{44}} \end{matrix} \qquad \begin{matrix} g_{14}(g_{12}g_{24} - g_{22}g_{14}) \\ +g_{24}(g_{14}g_{12} - g_{24}g_{11}) \end{matrix}$$

111 112 114 122 124 144

222 224 ~~233~~ ~~234~~ 244

444

$$2\frac{\partial g_{12}}{\partial x_1} = -\varphi'(x_2)$$

$$2\frac{\partial g_{12}}{\partial x_2} = -\psi'(x_1)$$

$$g_{12} = c_0 + c_1x_1 + c_2x_2 + \alpha x_1 x_2$$

$$\varphi'(x_2) = -2(c_1 + \alpha x_2)$$

$$\psi'(x_1) = -2(c_2 + \alpha x_1)$$

$$\varphi(x_2) = g_{11} = -2\left(c_1x_2 + \frac{\alpha}{2}x_2^2 + k''\right)$$

$$\psi(x_1) = g_{22} = -2\left(c_2x_1 + \frac{\alpha}{2}x_1^2 + k'''\right)$$

$$g_{14} = \beta x_2 + k$$

$$g_{24} = -\beta x_1 + k'$$

$$g_{14} = \varphi(x_2)$$

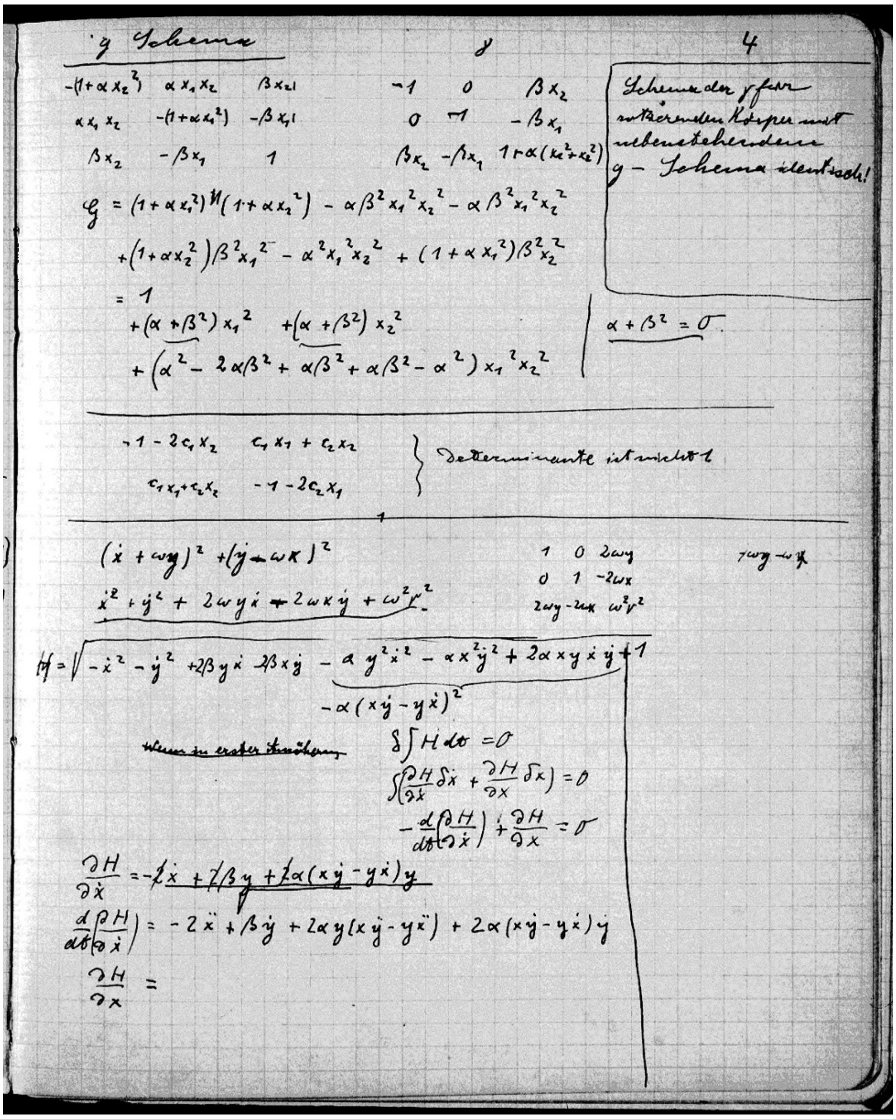
$$g_{24} = \psi(x_1)$$

$$\varphi'(x_2) + \psi'(x_1) = 0$$

$$\varphi'(x_2) = \alpha \qquad \varphi = \alpha x_2 + k$$

$$\psi'(x_1) = -\alpha \qquad \psi = -\alpha x_1 + k'$$

[p. 42 R]



g Schema $\qquad \gamma \qquad 4$

$$\begin{array}{ccc} -(1+\alpha x_2^2) & \alpha x_1 x_2 & \beta x_2 \\ \alpha x_1 x_2 & -(1+\alpha x_1^2) & -\beta x_1 \\ \beta x_2 & -\beta x_1 & 1 \end{array} \qquad \begin{array}{ccc} -1 & 0 & \beta x_2 \\ 0 & -1 & -\beta x_1 \\ \beta x_2 & -\beta x_1 & 1+\alpha(x_1^2+x_2^2) \end{array}$$

Schema der γ für rotierenden Körper mit nebenstehendem g – Schema identisch!

$$G = (1+\alpha x_1^2)(1+\alpha x_2^2) - \alpha\beta^2 x_1^2 x_2^2 - \alpha\beta^2 x_1^2 x_2^2$$
$$+(1+\alpha x_2^2)\beta^2 x_1^2 - \alpha^2 x_1^2 x_2^2 + (1+\alpha x_1^2)\beta^2 x_2^2$$
$$= 1$$
$$+(\alpha+\beta^2)x_1^2 \quad +(\alpha+\beta^2)x_2^2$$
$$+(\alpha^2 - 2\alpha\beta^2 + \alpha\beta^2 + \alpha\beta^2 - \alpha^2)x_1^2 x_2^2$$

$\alpha + \beta^2 = 0.$

$$\left.\begin{array}{cc} -1-2c_1x_2 & c_1x_1+c_2x_2 \\ c_1x_1+c_2x_2 & -1-2c_2x_1 \end{array}\right\} \text{Determinante ist nicht 1}$$

$$(\dot{x}+\omega y)^2 + (\dot{y}-\omega x)^2$$
$$\dot{x}^2 + \dot{y}^2 + 2\omega y\dot{x} - 2\omega x\dot{y} + \omega^2 r^2.$$

$$\begin{array}{ccc} 1 & 0 & 2\omega y \\ 0 & 1 & -2\omega x \\ 2\omega y & -2\omega x & \omega^2 r^2 \end{array} \qquad +\omega y - \omega x$$

$$H = \sqrt{-\dot{x}^2 - \dot{y}^2 + 2\beta y\dot{x} - 2\beta x\dot{y} \underbrace{- \alpha y^2\dot{x}^2 - \alpha x^2\dot{y}^2 + 2\alpha xy\dot{x}\dot{y}}_{-\alpha(x\dot{y}-y\dot{x})^2} + 1}$$

~~Wenn in erster Annäherung~~

$$\delta\int H\,dt = 0$$
$$\int\left(\frac{\partial H}{\partial\dot{x}}\delta\dot{x} + \frac{\partial H}{\partial x}\delta x\right) = 0$$
$$-\frac{d}{dt}\left(\frac{\partial H}{\partial\dot{x}}\right) + \frac{\partial H}{\partial x} = 0$$

$$\frac{\partial H}{\partial\dot{x}} = -2\dot{x} + 2\beta y + 2\alpha(x\dot{y}-y\dot{x})y$$
$$\frac{d}{dt}\left(\frac{\partial H}{\partial\dot{x}}\right) = -2\ddot{x} + \beta\dot{y} + 2\alpha y(x\ddot{y}-y\ddot{x}) + 2\alpha(x\dot{y}-y\dot{x})\dot{y}$$
$$\frac{\partial H}{\partial x} =$$

[p. 42 L]

g_{11} g_{22} g_{14} g_{44}
g_{12} g_{24}

$$\begin{array}{ccc} g_{11} & g_{12} & g_{14} \\ g_{21} & g_{22} & g_{24} \\ g_{41} & g_{42} & g_{44} \\ & & 0 \end{array}$$

$$g_{14}(g_{12}g_{24} - g_{22}g_{14}) + g_{24}(g_{14}g_{12} - g_{24}g_{12})$$

$$\frac{\partial g_{11}}{\partial x_1} = 0$$

$$\frac{\partial g_{22}}{\partial x_2} = 0$$

$\underline{111}$ $\underline{112}$ $\underline{114}$ $\underline{122}$ $\underline{124}$ $\underline{144}$
$\underline{222}$ $\underline{224}$ <233 234> $\underline{244}$
$\underline{444}$

$$\frac{\partial g_{44}}{\partial x_1} + 2\frac{\partial g_{\cancel{1}4}'}{\partial x_4'} = 0$$

$$\frac{\partial g_{44}}{\partial x_2} + 2\frac{\partial g_{\cancel{2}4}'}{\partial x_4'} = 0$$

$$\frac{\partial g_{1\cancel{1}}'}{\cancel{\partial x_4}} + 2\frac{\partial g_{14}}{\partial x_1} = 0$$

$$\frac{\partial g_{2\cancel{2}}'}{\cancel{\partial x_4}} + 2\frac{\partial g_{24}}{\partial x_2} = 0$$

$$\frac{\partial g_{14}}{\partial x_2} + \frac{\partial g_{24}}{\partial x_1} = 0$$

$$2\frac{\partial g_{12}}{\partial x_1} + \frac{\partial g_{11}}{\partial x_2} = 0 \qquad 2\frac{\partial g_{12}}{\partial x_1} = -\varphi'(x_2)$$

$$2\frac{\partial g_{12}}{\partial x_2} + \frac{\partial g_{22}}{\partial x_1} = 0 \qquad 2\frac{\partial g_{12}}{\partial x_2} = -\psi'(x_1)$$

$$g_{11} = \varphi(x_2) \qquad g_{12} = c_0 + c_1x_1 + c_2x_2 + \alpha x_1x_2$$

$$g_{22} = \psi(x_1)$$

$$\varphi'(x_2) = -2(c_1 + \alpha x_2)$$

$$\psi'(x_1) = -2(c_2 + \alpha x_1)$$

$$\varphi(x_2) = g_{11} = -2\left(c_1x_2 + \frac{\alpha}{2}x_2^2 + \kappa''\right)$$

$$\psi(x_1) = g_{22} = -2\left(c_2x_1 + \frac{\alpha}{2}x_1^2 + \kappa'''\right)$$

$$g_{14} = \beta x_2 + \kappa$$

$$g_{24} = -\beta x_1 + \kappa'$$

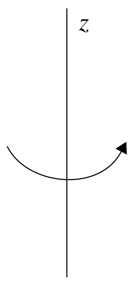
$$g_{14} = \varphi(x_2)$$

$$g_{24} = \psi(\underline{x_1})$$

$$\varphi'(x_2) + \psi'(x_1) = 0$$

$$\varphi'(x_2) = \alpha \qquad \varphi = \alpha x_2 + \kappa$$

$$\psi'(x_1) = -\alpha \qquad \psi = -\alpha x_1 + \kappa'$$



[p. 42 R]

<u>g Schema</u>

4

$$\begin{array}{ccc|ccc} & & & & \gamma & \\ -(1+\alpha x_2^2) & a x_1 x_2 & \beta x_2 & -1 & 0 & \beta x_2 \\ a x_1 x_2 & -(1+\alpha x_1^2) & -\beta x_1 & 0 & -1 & -\beta x_1 \\ \beta x_2 & -\beta x_1 & 1 & \beta x_2 & -\beta x_1 & 1+\alpha(x_1^2 x_2^2) \end{array}$$

Schema der γ für rotierenden Körper mit nebenstehendem g-Schema identisch!

$$G = (1+\alpha x_1^2)(1+\alpha x_2^2) - \alpha\beta^2 x_1^2 x_2^2 - \alpha\beta^2 x_1^2 x_2^2$$

$$+ (1+\alpha x_2^2)\beta^2 x_1^2 - \alpha^2 x_1^2 x_2^2 + (1+\alpha x_1^2)\beta^2 x_2^2$$

$$= 1$$

$$+ (\underline{\alpha+\beta^2})x_1^2 + (\underline{\alpha+\beta^2})x_2^2$$

$$+ (\alpha^2 - 2\alpha\beta^2 + \alpha\beta^2 + \alpha\beta^2 - \alpha^2)x_1^2 x_2^2$$

$$\underline{\alpha + \beta^2 = 0}$$

$$\left.\begin{array}{cc} -1 - 2c_1 x_2 & c_1 x_1 + c_2 x_2 \\ c_1 x_1 + c_2 x_2 & -1 - 2c_2 x_1 \end{array}\right\}$$ Determinante ist nicht 1.

$$(\dot{x} + \omega y)^2 + (\dot{y} - \omega x)^2$$

$$\underline{\dot{x}^2 + \dot{y}^2 + 2\omega y\dot{x} - 2\omega x\dot{y} + \omega^2 r^2}$$

$$\begin{array}{ccc} 1 & 0 & 2\omega y \\ 0 & 1 & -2\omega x \\ 2\omega y & -2\omega x & \omega^2 r^2 \end{array}$$

$+\omega y - \omega x$

$$H = \sqrt{-\dot{x}^2 - \dot{y}^2 + 2\beta y\dot{x} - 2\beta x\dot{y} \underbrace{- \alpha y^2\dot{x}^2 - \alpha x^2\dot{y}^2 + 2\alpha xy\dot{x}\dot{y}}_{-\alpha(x\dot{y} - y\dot{x})^2} + 1}$$

⟨Wenn in erster Annäherung⟩ $\delta\int H\,dt = 0$

$$\int\left(\frac{\partial H}{\partial \dot{x}}\delta\dot{x} + \frac{\partial H}{\partial x}\delta x\right) = 0$$

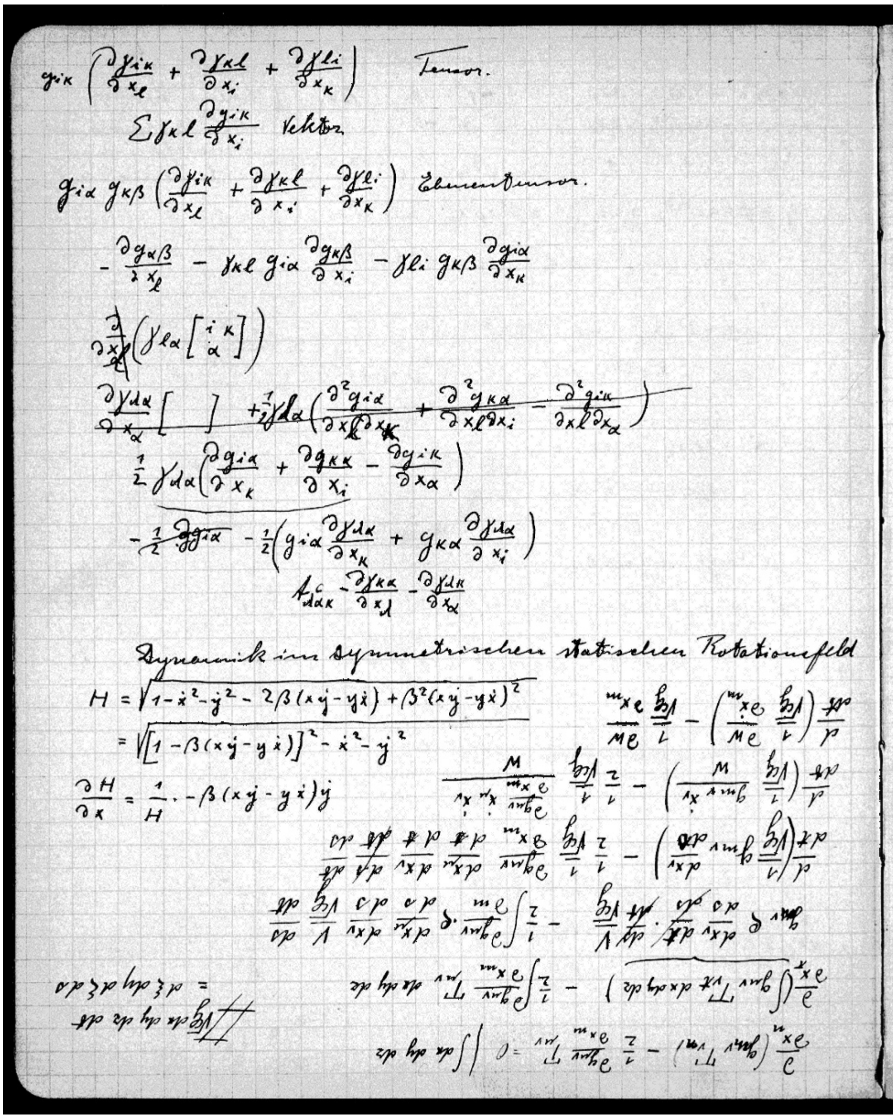
$$-\frac{d}{dt}\left(\frac{\partial H}{\partial \dot{x}}\right) + \frac{\partial H}{\partial x} = 0$$

$$\frac{\partial H}{\partial \dot{x}} = \frac{-\langle 2\rangle\dot{x} + \langle 2\rangle\beta y + \langle 2\rangle\alpha(x\dot{y} - y\dot{x})y}{\sqrt{}}$$

$$\frac{d}{dt}\left(\frac{\partial H}{\partial \dot{x}}\right) = -2\ddot{x} + \beta\dot{y} + 2\alpha y(x\ddot{y} - y\ddot{x}) + 2\alpha(x\dot{y} - y\dot{x})\dot{y}$$

$$\frac{\partial H}{\partial x} =$$

[p. 43 LA]



$$g_{ik}\left(\frac{\partial\gamma_{ik}}{\partial x_l}+\frac{\partial\gamma_{kl}}{\partial x_i}+\frac{\partial\gamma_{li}}{\partial x_k}\right)\quad \text{Tensor.}$$

$$\Sigma\gamma_{kl}\frac{\partial g_{ik}}{\partial x_i}\quad \text{Vektor}$$

$$g_{i\alpha}\,g_{k\beta}\left(\frac{\partial\gamma_{ik}}{\partial x_l}+\frac{\partial\gamma_{kl}}{\partial x_i}+\frac{\partial\gamma_{li}}{\partial x_k}\right)\quad \text{Ebenentensor.}$$

$$-\frac{\partial g_{\alpha\beta}}{\partial x_l}-\gamma_{kl}\,g_{i\alpha}\frac{\partial g_{k\beta}}{\partial x_i}-\gamma_{li}\,g_{k\beta}\frac{\partial g_{i\alpha}}{\partial x_k}$$

$$\frac{\partial}{\partial x_l}\left(\gamma_{l\alpha}\begin{bmatrix} i\,k\\ \alpha\end{bmatrix}\right)$$

$$\cancel{\frac{\partial\gamma_{\lambda\alpha}}{\partial x_\alpha}[\quad]+\frac{1}{2}\gamma_{\lambda\alpha}\left(\frac{\partial^2 g_{i\alpha}}{\partial x_l\partial x_k}+\frac{\partial^2 g_{k\alpha}}{\partial x_l\partial x_i}-\frac{\partial^2 g_{ik}}{\partial x_l\partial x_\alpha}\right)}$$

$$\frac{1}{2}\gamma_{\lambda\alpha}\left(\frac{\partial g_{i\alpha}}{\partial x_k}+\frac{\partial g_{k\alpha}}{\partial x_i}-\frac{\partial g_{ik}}{\partial x_\alpha}\right)$$

$$-\cancel{\frac{1}{2}\partial g_{i\alpha}}-\frac{1}{2}\left(g_{i\alpha}\frac{\partial\gamma_{\lambda\alpha}}{\partial x_k}+g_{k\alpha}\frac{\partial\gamma_{\lambda\alpha}}{\partial x_i}\right)$$

$$A^c_{\lambda\alpha k}-\frac{\partial\gamma_{k\alpha}}{\partial x_\lambda}-\frac{\partial\gamma_{\lambda k}}{\partial x_\alpha}$$

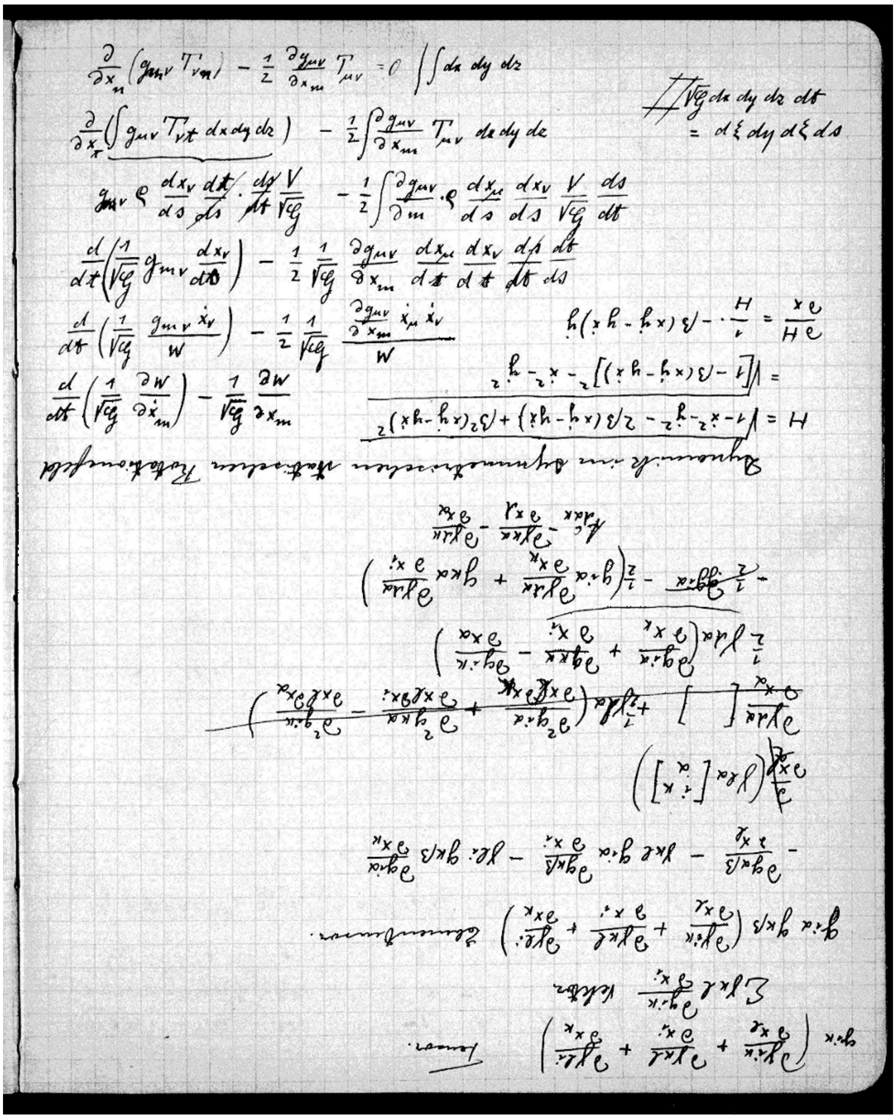
Dynamik im symmetrischen statischen Rotationsfeld

$$H=\sqrt{1-\dot x^2-\dot y^2-2\beta(x\dot y-y\dot x)+\beta^2(x\dot y-y\dot x)^2}$$

$$=\sqrt{[1-\beta(x\dot y-y\dot x)]^2-\dot x^2-\dot y^2}$$

$$\frac{\partial H}{\partial x}=\frac{1}{H}\cdot-\beta(x\dot y-y\dot x)\dot y$$

[p. 43 LB: same as p. 43LA but upside down]



$$\frac{\partial}{\partial x_n}(g_{m\nu}T_{\nu n}) - \frac{1}{2}\frac{\partial g_{\mu\nu}}{\partial x_m}T_{\mu\nu} = 0 \quad \iint dx\,dy\,dz$$

$$\frac{\partial}{\partial x_\tau}\left(\int g_{\mu\nu}T_{\nu t}\,dx\,dy\,dz\right) - \frac{1}{2}\int\frac{\partial g_{\mu\nu}}{\partial x_m}T_{\mu\nu}\,dx\,dy\,dz \qquad \int\sqrt{G}\,dx\,dy\,dz\,dt = d\xi\,d\eta\,d\zeta\,d\sigma$$

$$g_{m\nu}\rho\frac{dx_\nu}{ds}\frac{dt}{ds}\cdot\frac{d\sigma}{dt}\frac{V}{\sqrt{G}} - \frac{1}{2}\int\frac{\partial g_{\mu\nu}}{\partial x_m}\cdot\rho\frac{dx_\mu}{ds}\frac{dx_\nu}{ds}\frac{V}{\sqrt{G}}\frac{ds}{dt}$$

$$\frac{d}{dt}\left(\frac{1}{\sqrt{G}}g_{m\nu}\frac{dx_\nu}{dt}\right) - \frac{1}{2}\frac{1}{\sqrt{G}}\frac{\partial g_{\mu\nu}}{\partial x_m}\frac{dx_\mu}{dt}\frac{dx_\nu}{dt}\frac{d\rho}{dt}\frac{dt}{ds}$$

$$\frac{d}{dt}\left(\frac{1}{\sqrt{G}}\frac{g_{m\nu}\dot{x}_\nu}{W}\right) - \frac{1}{2}\frac{1}{\sqrt{G}}\frac{\frac{\partial g_{\mu\nu}}{\partial x_m}\dot{x}_\mu\dot{x}_\nu}{W}$$

$$\frac{d}{dt}\left(\frac{1}{\sqrt{G}}\frac{\partial W}{\partial \dot{x}_m}\right) - \frac{1}{\sqrt{G}}\frac{\partial W}{\partial x_m}$$

[p. 43 LA: the part of p. 43 written from the top of the page]

$$g_{i\kappa}\left(\frac{\partial\gamma_{i\kappa}}{\partial x_l}+\frac{\partial\gamma_{\kappa l}}{\partial x_i}+\frac{\partial\gamma_{li}}{\partial x_\kappa}\right) \qquad \text{Tensor.}$$

$$\sum\gamma_{\kappa l}\frac{\partial g_{i\kappa}}{\partial x_i} \qquad \text{Vektor.}$$

$$g_{i\alpha}g_{\kappa\beta}\left(\frac{\partial\gamma_{i\kappa}}{\partial x_l}+\frac{\partial\gamma_{\kappa l}}{\partial x_i}+\frac{\partial\gamma_{li}}{\partial x_\kappa}\right) \qquad \text{Ebenentensor.}$$

$$-\frac{\partial g_{\alpha\beta}}{\partial x_l}-\gamma_{\kappa l}g_{i\alpha}\frac{\partial g_{\kappa\beta}}{\partial x_i}-\gamma_{li}g_{\kappa\beta}\frac{\partial g_{i\alpha}}{\partial x_\kappa}$$

$$\cancel{\frac{\partial}{\partial x_l}}\left(\gamma_{l\alpha}\begin{bmatrix} i\kappa \\ \alpha \end{bmatrix}\right)$$

$$\cancel{\frac{\partial\gamma_{\lambda\alpha}}{\partial x_\alpha}\begin{bmatrix} \; \end{bmatrix}+\frac{1}{2}\gamma_{l\alpha}\left(\frac{\partial^2 g_{i\alpha}}{\partial x_l\partial x_\kappa}+\frac{\partial^2 g_{\kappa\alpha}}{\partial x_l\partial x_i}-\frac{\partial^2 g_{i\kappa}}{\partial x_l\partial x_\alpha}\right)}$$

$$\frac{1}{2}\underbrace{\gamma_{\lambda\alpha}\left(\frac{\partial g_{i\alpha}}{\partial x_\kappa}+\frac{\partial g_{\kappa\alpha}}{\partial x_i}\right.}-\left.\frac{\partial g_{i\kappa}}{\partial x_\alpha}\right)$$

$$\cancel{-\frac{1}{2}\partial g_{i\alpha}} \qquad -\frac{1}{2}\left(g_{i\alpha}\frac{\partial\gamma_{\lambda\alpha}}{\partial x_\kappa}+g_{\kappa\alpha}\frac{\partial\gamma_{\lambda\alpha}}{\partial x_i}\right)$$

$$t^c_{\lambda\alpha\kappa}-\frac{\partial\gamma_{\kappa\alpha}}{\partial x_\lambda}-\frac{\partial\gamma_{\lambda\kappa}}{\partial x_\alpha}$$

Dynamik im symmetrischen statischen Rotationsfeld

$$H = \sqrt{1-\dot{x}^2-\dot{y}^2-2\beta(x\dot{y}-y\dot{x})+\beta^2(x\dot{y}-y\dot{x})^2}$$

$$= \sqrt{[1-\beta(x\dot{y}-y\dot{x})]^2-\dot{x}^2-\dot{y}^2}$$

$$\frac{\partial H}{\partial x} = \frac{1}{H}\cdot\ -\beta(x\dot{y}-y\dot{x})\dot{y}$$

[ ]

[p. 43 LB: the part of p. 43L written from the bottom of the page]

$$\frac{\partial}{\partial x_n}(g_{m\nu}T_{\nu n}) - \frac{1}{2}\frac{\partial g_{\mu\nu}}{\partial x_m}T_{\mu\nu} = 0 \quad \Big| \int dx\, dy\, dz$$

$$\cancel{/\!/}\sqrt{G}dx\, dy\, dz\, dt = d\xi d\eta d\zeta ds$$

$$\frac{\partial}{\partial x_t}\Big(\underbrace{\int g_{\mu\nu}T_{\nu t}dx\, dy\, dz}_{}\Big) - \frac{1}{2}\int\frac{\partial g_{\mu\nu}}{\partial x_m}T_{\mu\nu}dx\, dy\, dz$$

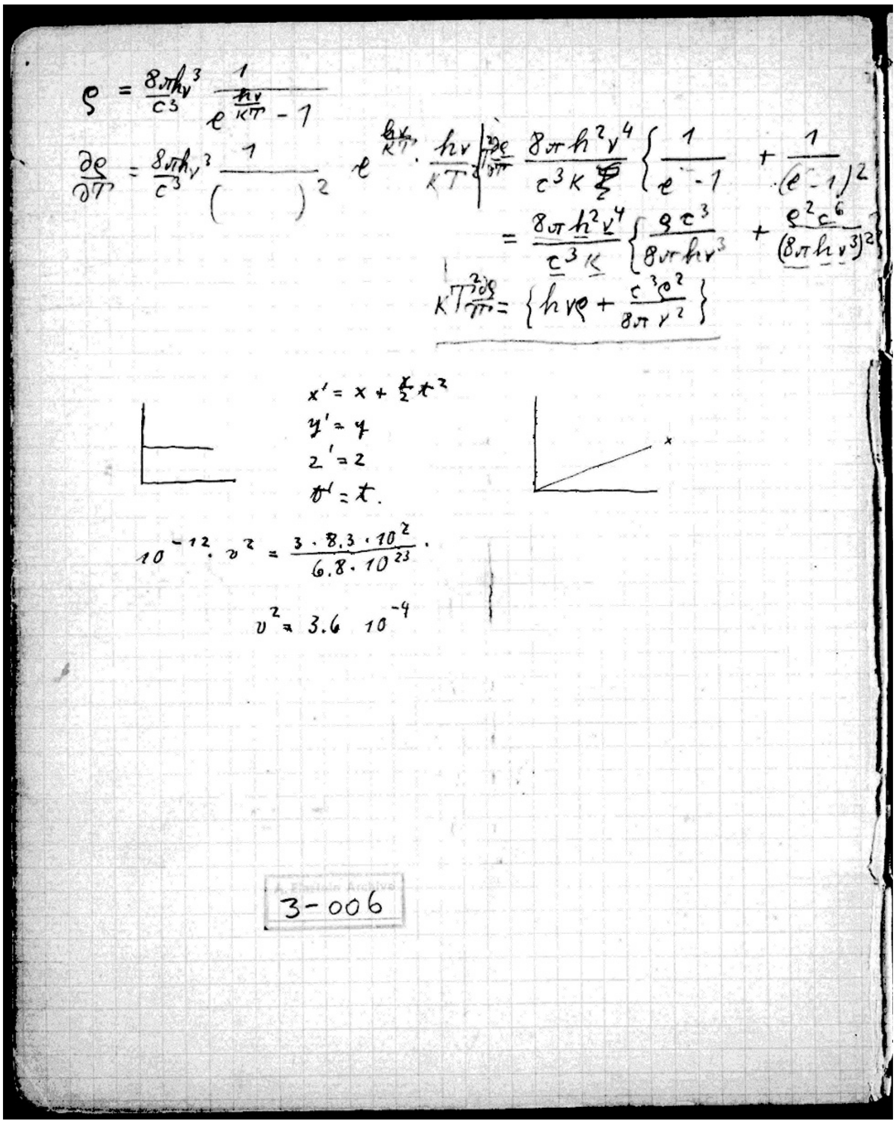
$$g_{\mu\nu}\rho\frac{dx_\nu}{ds}\frac{d\cancel{t}}{\cancel{d}s}\cdot\frac{d\cancel{s}}{\cancel{d}t}\frac{V}{\sqrt{G}} - \frac{1}{2}\int\frac{\partial g_{\mu\nu}}{\partial m}\cdot\rho\frac{dx_\mu}{ds}\frac{dx_\nu}{ds}\frac{V}{\sqrt{G}}\frac{ds}{dt}$$

$$\frac{d}{dt}\left(\frac{1}{\sqrt{G}}g_{m\nu}\frac{dx_\nu}{ds}\right) - \frac{1}{2}\frac{1}{\sqrt{G}}\frac{\partial g_{\mu\nu}}{\partial x_m}\frac{dx_\mu}{dt}\frac{dx_\nu}{dt}\frac{d\cancel{s}}{\cancel{d}t}\frac{dt}{ds}$$

$$\frac{d}{dt}\left(\frac{1}{\sqrt{G}}\frac{g_{m\nu}\dot{x}_\nu}{w}\right) - \frac{1}{2}\frac{1}{\sqrt{G}}\frac{\frac{\partial g_{\mu\nu}}{\partial x_m}\dot{x}_\mu\dot{x}_\nu}{w}$$

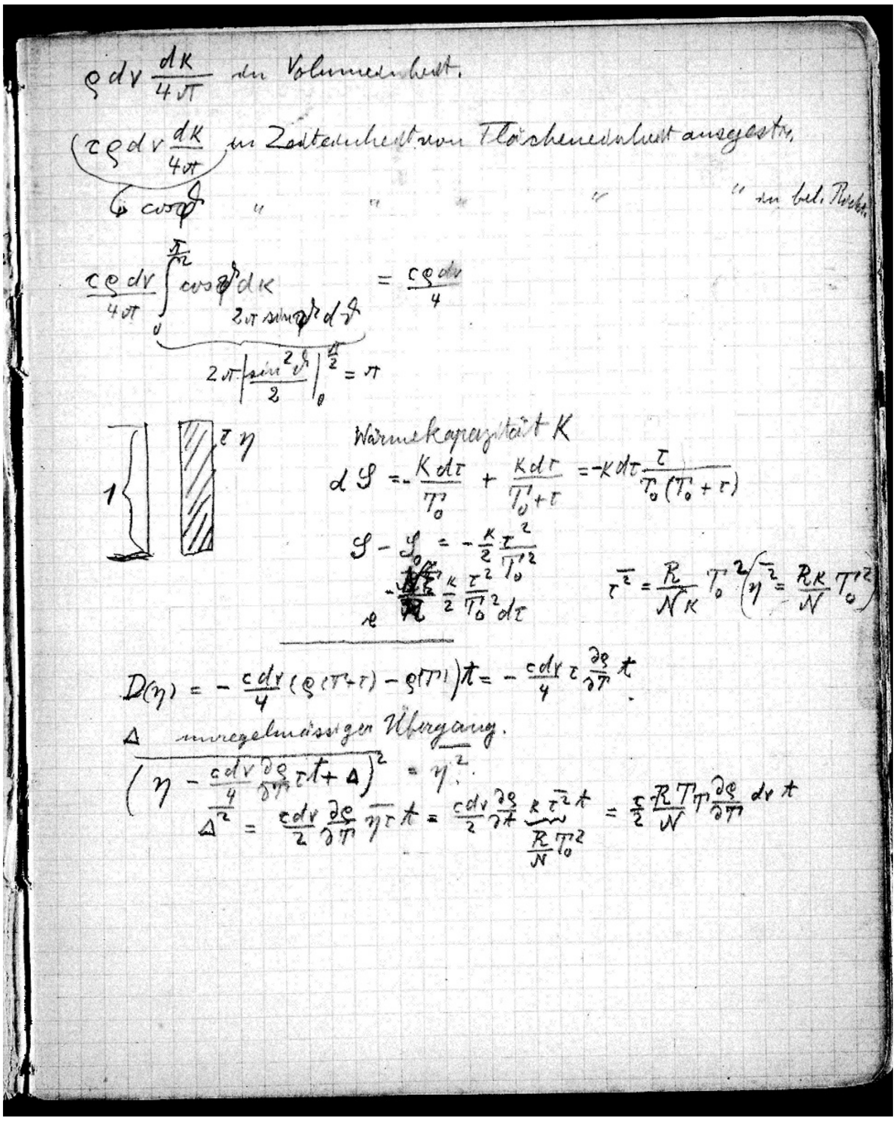
$$\frac{d}{dt}\left(\frac{1}{\sqrt{G}}\frac{\partial w}{\partial \dot{x}_m}\right) - \frac{1}{\sqrt{G}}\frac{\partial w}{\partial x_m}$$

[p. 01 L]



$\varrho = \frac{8\pi h\nu^3}{c^3}\frac{1}{e^{\frac{h\nu}{kT}} - 1}$
$\frac{\partial\varrho}{\partial T} = \frac{8\pi h\nu^3}{c^3}\frac{1}{(\quad)^2}\, e^{\frac{h\nu}{kT}} \cdot \frac{h\nu}{kT^2}$
$\frac{\partial\varrho}{\partial T} = \frac{8\pi h^2\nu^4}{c^3 k T^2}\left\{\frac{1}{e-1} + \frac{1}{(e-1)^2}\right\}$
$= \frac{8\pi h^2\nu^4}{c^3 k}\left\{\frac{\varrho c^3}{8\pi h\nu^3} + \frac{\varrho^2 c^6}{(8\pi h\nu^3)^2}\right\}$
$kT^2\frac{\partial\varrho}{\partial T} = \left\{h\nu\varrho + \frac{c^3\varrho^2}{8\pi\nu^2}\right\}$
$x' = x + \frac{\gamma}{2}t^2$
$y' = y$
$z' = z$
$t' = t.$
$10^{-12}\cdot v^2 = \frac{3\cdot 8.3\cdot 10^2}{6.8\cdot 10^{23}}.$
$v^2 = 3.6\ 10^{-4}$
A. Einstein Archive
3-006

[p. 01 R]



$\varrho \, dv \frac{dK}{4\pi}$ in Volumeneinheit.

$\left(c \varrho \, dv \frac{dK}{4\pi} \right)$ in Zeiteinheit von Flächeneinheit ausgestr.

$\cdot \cos\vartheta$ " " " " in bel. Richt.

$$\frac{c\varrho \, dv}{4\pi} \int_0^{\frac{\pi}{2}} \cos\vartheta \, \underbrace{dK}_{2\pi \sin\vartheta \, d\vartheta} = \frac{c\varrho \, dv}{4}$$

$$2\pi \left| \frac{\sin^2\vartheta}{2} \right|_0^{\frac{\pi}{2}} = \pi$$

1 τ η

Wärmekapazität K

$$dS = -\frac{K d\tau}{T_0'} + \frac{K d\tau}{T_0' + \tau} = -K d\tau \frac{\tau}{T_0(T_0 + \tau)}$$

$$S - S_0 = -\frac{K}{2} \frac{\tau^2}{T_0'^2}$$

$$e^{-\frac{K}{2}\frac{\tau^2}{T_0^2}} d\tau \qquad \overline{\tau^2} = \frac{R}{NK} T_0^2 \left(\overline{\eta^2} = \frac{RK}{N} T_0'^2 \right)$$

$$D(\eta) = -\frac{c\,dv}{4} (\varrho(T' + \tau) - \varrho(T')) t = -\frac{c\,dv}{4} \tau \frac{\partial \varrho}{\partial T'} t$$

Δ unregelmässiger Übergang.

$$\overline{\left(\eta - \frac{c\,dv}{4} \frac{\partial \varrho}{\partial T'} \tau t + \Delta \right)^2} = \overline{\eta^2}.$$

$$\overline{\Delta^2} = \frac{c\,dv}{2} \frac{\partial \varrho}{\partial T'} \overline{\eta \tau} t = \frac{c\,dv}{2} \frac{\partial \varrho}{\partial T'} \underbrace{K \overline{\tau^2}}_{\frac{R}{N} T_0^2} t = \frac{c}{2} \frac{R}{N} T'^2 \frac{\partial \varrho}{\partial T'} dv \, t$$

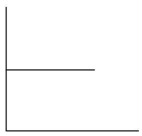
[p. 01 L]

$$\rho = \frac{8\pi}{c^3}h\nu^3 \; \frac{1}{e^{\frac{h\nu}{\kappa T}}-1}$$

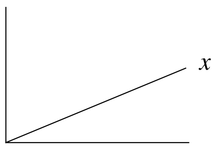
$$\frac{\partial\rho}{\partial T} = \frac{8\pi}{c^3}h\nu^3 \; \frac{1}{\left(\qquad\right)^2} \qquad e^{\frac{h\nu}{\kappa T}}\cdot\frac{h\nu}{\kappa T^2} \qquad \Bigg| \; T^2\frac{\partial\rho}{\partial T} \quad \frac{8\pi h^2\nu^4}{c^3\kappa\cancel{T}}\left\{\frac{1}{e^{\cdot}-1}+\frac{1}{(e^{\cdot}-1)^2}\right.$$

$$= \frac{\underline{8\pi h^2 v^2}}{\underline{c}^3\underline{\kappa}}\left\{\frac{\rho c^3}{8\pi h\nu^3}+\frac{\rho^2c^6}{(\underline{8\pi h\ v^3})^2}\right\}$$

$$\kappa T^2\frac{\partial\rho}{\partial T} = \left\{h\nu\rho+\frac{c^3\rho^2}{8\pi\nu^2}\right\}$$



$$\begin{aligned} x' &= x+\frac{\gamma}{2}t^2 \\ y' &= y \\ z' &= z \\ t' &= t. \end{aligned}$$



$$10^{-12}\cdot\upsilon^2 = \frac{3\cdot 8.3\cdot 10^7}{6.8\cdot 10^{23}}\cdot$$

$$\upsilon^2 = 3.6 \quad 10^{-4}$$

[p. 01 R]

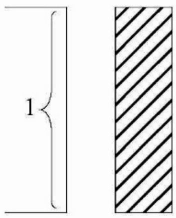
$\rho d\nu \frac{d\kappa}{4\pi}$ in Volumeinheit.

$\tau\rho d\nu \frac{d\kappa}{4\pi}$ in Zeiteinheit von Flächeneinheit ausgestr.

$\cos\vartheta$ " " " " " in bel. Richt.

$$\frac{c\rho d\nu}{4\pi}\underbrace{\int_0^{\frac{\pi}{2}} \cos\vartheta d\kappa}_{2\pi\sin\vartheta d\vartheta} = \frac{c\rho d\nu}{4}$$

$$2\pi\left|\frac{\sin^2\vartheta}{2}\right|_0^{\frac{\pi}{2}} = \pi$$



$\tau\,\eta$ Wärmekapazität κ

$$dS = -\frac{\kappa d\tau}{T_0} + \frac{\kappa d\tau}{T_0+\tau} = -\kappa d\tau \frac{\tau}{T_0(T_0+\tau)}$$

$$S - S_0 = -\frac{\kappa\tau^2}{2T_0^2}$$

$$e^{-\frac{N<T>\kappa}{R}\frac{\tau^2}{2T_0^2}} d\tau \qquad \overline{\tau^2} = \frac{R}{N\kappa}T_0^2 \left(\overline{\eta^2} = \frac{R\kappa}{N}T_0^2\right)$$

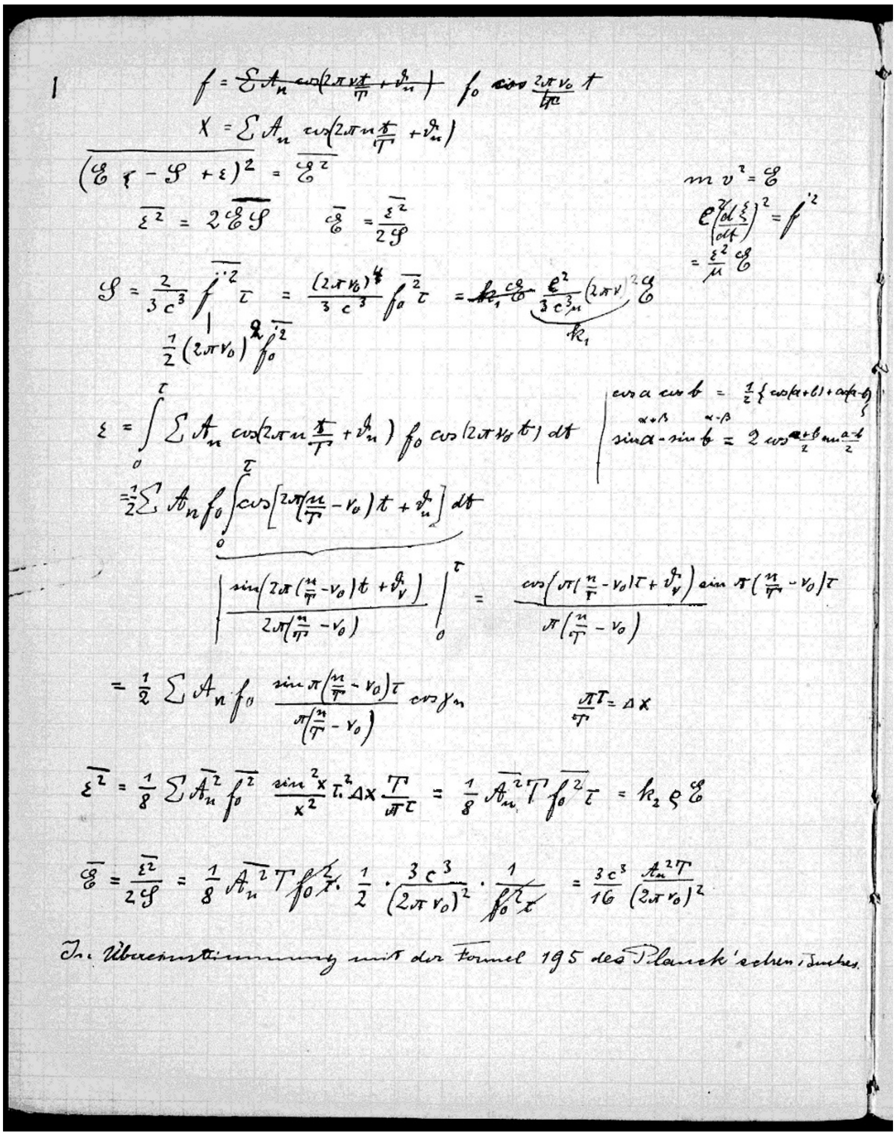
$$D(\eta) = -\frac{cd\nu}{4}(\rho(T+\tau)-\rho(T))t = -\frac{cd\nu}{4}\tau\frac{\partial\rho}{\partial T}t$$

Δ unregelmässiger Übergang.

$$\overline{\left(\eta - \frac{cd\nu}{4}\frac{\partial\rho}{\partial T}\tau t + \Delta\right)^2} = \overline{\eta^2}$$

$$\overline{\Delta^2} = \frac{cd\nu}{2}\frac{\partial\rho}{\partial T}\overline{\eta\tau}t = \frac{cd\nu}{2}\frac{\partial\rho}{\partial t}\kappa\underbrace{\overline{\tau^2}}_{\frac{R}{N}T_0^2}t = \frac{c}{2}\;\frac{-RT}{N}T\frac{\partial\rho}{\partial T}d\nu\, t$$

[p. 02 L]



1

$$f = \cancel{\sum A_n \cos\left(2\pi n \frac{t}{T} + \vartheta_n\right)} \quad f_0 \cos \frac{2\pi \nu_0}{\cancel{T}} t$$

$$X = \sum A_n \cos\left(2\pi n \frac{t}{T} + \vartheta_n\right)$$

$$\overline{(\mathfrak{E}\,\xi - \mathfrak{S} + \varepsilon)^2} = \overline{\mathfrak{E}^2}$$

$$m v^2 = \mathfrak{E}$$

$$\overline{\varepsilon^2} = 2\overline{\mathfrak{E}\mathfrak{S}} \qquad \overline{\mathfrak{E}} = \frac{\overline{\varepsilon^2}}{2\mathfrak{S}}$$

$$e^2 \left(\frac{d\xi}{dt}\right)^2 = \dot{f}^2 = \frac{\varepsilon^2}{\mu} \mathfrak{E}$$

$$\mathfrak{S} = \frac{2}{3c^3} \overline{\dot{f}^2}\,\tau = \frac{(2\pi\nu_0)^4}{3c^3} \overline{f_0^2}\,\tau = \cancel{k_1 \mathfrak{E}}\ \underbrace{\frac{e^2}{3c^3\mu}(2\pi\nu)^2}_{k_1} \mathfrak{E}$$

$$\overline{\dot{f}^2} = \frac{1}{2}(2\pi\nu_0)^2 \overline{f_0^2}$$

$$\varepsilon = \int_0^\tau \sum A_n \cos\left(2\pi n \frac{t}{T} + \vartheta_n\right) f_0 \cos(2\pi\nu_0 t)\, dt$$

$$\cos a \cos b = \frac{1}{2}\{\cos(a+b) + \cos(a-b)\}$$

$$\sin \alpha - \sin b = 2 \cos \frac{a+b}{2} \sin \frac{a-b}{2} \qquad (\alpha+\beta,\ \alpha-\beta)$$

$$= \frac{1}{2} \sum A_n f_0 \underbrace{\int_0^\tau \cos\left[2\pi\left(\frac{n}{T} - \nu_0\right) t + \vartheta_n\right] dt}$$

$$\left| \frac{\sin\left(2\pi\left(\frac{n}{T} - \nu_0\right)t + \vartheta_\nu\right)}{2\pi\left(\frac{n}{T} - \nu_0\right)} \right|_0^\tau = \frac{\cos\left(\pi\left(\frac{n}{T} - \nu_0\right)\tau + \vartheta_\nu\right) \sin \pi\left(\frac{n}{T} - \nu_0\right)\tau}{\pi\left(\frac{n}{T} - \nu_0\right)}$$

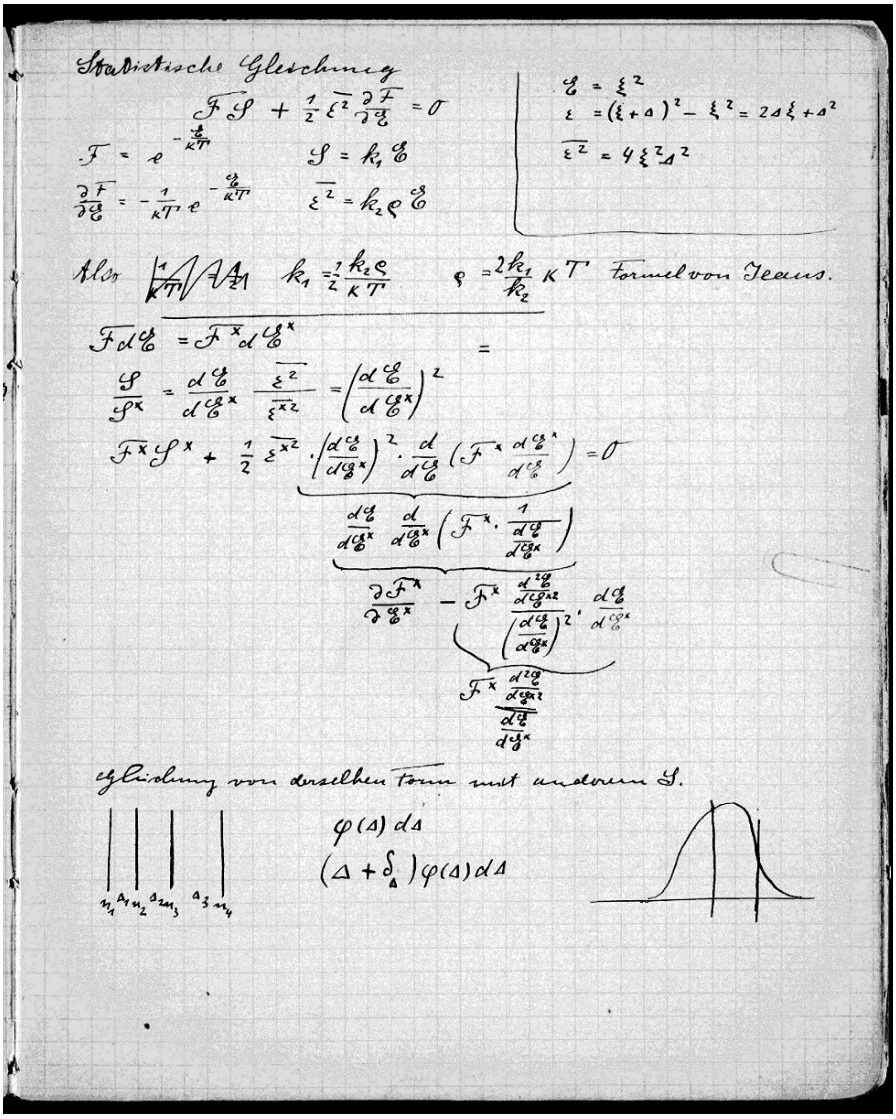
$$= \frac{1}{2} \sum A_n f_0 \frac{\sin \pi\left(\frac{n}{T} - \nu_0\right)\tau}{\pi\left(\frac{n}{T} - \nu_0\right)} \cos \gamma_n \qquad \frac{\pi\tau}{T} = \Delta x$$

$$\overline{\varepsilon^2} = \frac{1}{8} \sum \overline{A_n^2}\, \overline{f_0^2} \frac{\sin^2 x}{x^2} \tau^2 \Delta x \frac{T}{\pi\tau} = \frac{1}{8} \overline{A_n^2} T \overline{f_0^2} \tau = k_2 \varrho \mathfrak{E}$$

$$\overline{\mathfrak{E}} = \frac{\overline{\varepsilon^2}}{2\mathfrak{S}} = \frac{1}{8} \overline{A_n^2} T \cancel{f_0^2 \tau} \cdot \frac{1}{2} \cdot \frac{3c^3}{(2\pi\nu_0)^2} \cdot \frac{1}{\cancel{f_0^2 \tau}} = \frac{3c^3}{16} \frac{\overline{A_n^2} T}{(2\pi\nu_0)^2}$$

In Übereinstimmung mit der Formel 195 des Planck'schen Buches.

[p. 02 R]



Statistische Gleichung

$$\mathfrak{F}\mathfrak{S} + \frac{1}{2}\overline{\varepsilon^2}\frac{\partial \mathfrak{F}}{\partial \mathfrak{E}} = 0$$

$$\mathfrak{F} = e^{-\frac{\mathfrak{E}}{\kappa T}} \qquad \mathfrak{S} = k_1 \mathfrak{E}$$

$$\frac{\partial \mathfrak{F}}{\partial \mathfrak{E}} = -\frac{1}{\kappa T} e^{-\frac{\mathfrak{E}}{\kappa T}} \qquad \overline{\varepsilon^2} = k_2 \varrho \mathfrak{E}$$

$$\mathfrak{E} = \xi^2$$

$$\varepsilon = (\xi+\Delta)^2 - \xi^2 = 2\Delta\xi + \Delta^2$$

$$\overline{\varepsilon^2} = 4\xi^2\Delta^2$$

Also $k_1 = \frac{1}{2}\frac{k_2 \varrho}{\kappa T}$ $\qquad \varrho = \frac{2k_1}{k_2}\kappa T$ Formel von Jeans.

$$\mathfrak{F} d\mathfrak{E} = \mathfrak{F}^x d\mathfrak{E}^x \qquad =$$

$$\frac{\mathfrak{S}}{\mathfrak{S}^x} = \frac{d\mathfrak{E}}{d\mathfrak{E}^x} \qquad \frac{\overline{\varepsilon^2}}{\overline{\varepsilon^{x2}}} = \left(\frac{d\mathfrak{E}}{d\mathfrak{E}^x}\right)^2$$

$$\mathfrak{F}^x\mathfrak{S}^x + \frac{1}{2}\overline{\varepsilon^{x2}} \cdot \left(\frac{d\mathfrak{E}}{d\mathfrak{E}^x}\right)^2 \cdot \frac{d}{d\mathfrak{E}}\left(\mathfrak{F}^x \frac{d\mathfrak{E}^x}{d\mathfrak{E}}\right) = 0$$

$$\frac{d\mathfrak{E}}{d\mathfrak{E}^x}\,\frac{d}{d\mathfrak{E}^x}\left(\mathfrak{F}^x \cdot \frac{1}{\frac{d\mathfrak{E}}{d\mathfrak{E}^x}}\right)$$

$$\frac{\partial \mathfrak{F}^x}{\partial \mathfrak{E}^x} - \mathfrak{F}^x \frac{\frac{d^2\mathfrak{E}}{d\mathfrak{E}^{x2}}}{\left(\frac{d\mathfrak{E}}{d\mathfrak{E}^x}\right)^2} \cdot \frac{d\mathfrak{E}}{d\mathfrak{E}^x}$$

$$\mathfrak{F}^x \frac{\frac{d^2\mathfrak{E}}{d\mathfrak{E}^{x2}}}{\frac{d\mathfrak{E}}{d\mathfrak{E}^x}}$$

Gleichung von derselben Form mit anderem $\mathfrak{S}$.

$n_1\ \Delta_1\ n_2\ \Delta_2\ n_3\ \Delta_3\ n_4$

$$\varphi(\Delta)\,d\Delta$$

$$(\Delta + \delta_\Delta)\varphi(\Delta)\,d\Delta$$

[p. 02 L]

$$f = \cancel{\sum A_n \cos\left(2\pi\nu\frac{t}{T} + \vartheta_n\right)} \qquad f_0 \cos\frac{2\pi\nu_0\ t}{\cancel{T}}$$

$$x = \sum A_n \cos\left(2\pi n\frac{t}{T} + \vartheta_n\right)$$

$$\overline{(E - S + \varepsilon)^2} = \overline{E^2} \qquad\qquad m\upsilon^2 = E$$

$$\overline{\varepsilon^2} = 2\overline{E}\overline{S} \qquad \overline{E} = \frac{\overline{\varepsilon^2}}{2S} \qquad\qquad e^2\left(\frac{d\xi}{dt}\right)^2 = f^2$$

$$= \frac{\varepsilon^2}{\mu}E$$

$$S = \frac{2}{3c^3}\overline{\ddot{f}^2}\tau = \frac{(2\pi\nu_0)^4}{3c^3}\overline{f_0^2}\tau = \cancel{k_1 E}\underbrace{\frac{e^2}{3c^3\mu}(2\pi\nu)^2}_{k_1}E$$

$$\frac{1}{2}(2\pi\nu_0)^2\overline{\dot{f}_0^2}$$

$$\cos a \cos b = \frac{1}{2}\{\cos(a+b) + \cos(a-b)\}$$

$$\sin \overset{\alpha+\beta}{a} - \sin \overset{\alpha-\beta}{b} = 2\cos\frac{a+b}{2}\sin\frac{a-b}{2}$$

$$\varepsilon = \int_0^\tau \sum A_n \cos\left(2\pi n\frac{t}{T} + \vartheta_n\right) f_0 \cos(2\pi\nu_0 t)\,dt$$

$$= \frac{1}{2}\sum A_n f_0 \underbrace{\int_0^\tau \cos\left[2\pi\left(\frac{n}{T} - \nu_0\right)t + \vartheta_n\right]dt}$$

$$\left|\frac{\sin\left(2\pi\left(\frac{n}{T} - \nu_0\right)t + \vartheta_\nu\right)}{2\pi\left(\frac{n}{T} - \nu_0\right)}\right|_0^\tau = \frac{\cos\left(\pi\left(\frac{n}{T} - \nu_0\right)\tau + \vartheta_\nu\right)\sin\pi\left(\frac{n}{T} - \nu_0\right)\tau}{\pi\left(\frac{n}{T} - \nu_0\right)}$$

$$= \frac{1}{2}\sum A_n f_0 \frac{\sin\pi\left(\frac{n}{T} - \nu_0\right)\tau}{\pi\left(\frac{n}{T} - \nu_0\right)}\cos\gamma_n \qquad\qquad \frac{\pi\tau}{T} = \Delta x$$

$$\overline{\varepsilon^2} = \frac{1}{8}\sum \overline{A_n^2}\,\overline{f_0^2}\frac{\sin^2 x}{x^2}\tau^2 \cdot \Delta x\frac{T}{\pi\tau} = \frac{1}{8}\overline{A_n^2}\,T\overline{f_0^2}\tau = k_2\rho E$$

$$\overline{E} = \frac{\overline{\varepsilon^2}}{2S} = \frac{1}{8}\overline{A_n^2}T\cancel{\overline{f_0^2}\tau} \cdot \frac{1}{2} \cdot \frac{3c^3}{(2\pi\nu_0)^2} \cdot \frac{1}{\cancel{\overline{f_0^2}\tau}} = \frac{3c^3}{16}\frac{A_n^2 T}{(2\pi\nu_0)^2}$$

In Übereinstimmung mit der Formel 195 des Planck'schen Buches.

[p. 02 R]

Statistische Gleichung

$$FS + \frac{1}{2}\overline{\varepsilon^2}\frac{\partial F}{\partial E} = 0$$

$$F = e^{-\frac{E}{\kappa T}} \qquad S = k_1 E$$

$$\frac{\partial F}{\partial E} = -\frac{1}{\kappa T}e^{-\frac{E}{\kappa T}} \qquad \overline{\varepsilon^2} = k_2 \rho E$$

$$E = \xi^2$$

$$\varepsilon = (\xi + \Delta)^2 - \xi^2 = 2\Delta\xi + \Delta^2$$

$$\overline{\varepsilon^2} = 4\xi^2\Delta^2$$

Also $\cancel{\frac{1}{\kappa T} = \frac{1}{2}}$ $\quad k_1 = \frac{1}{2}\frac{k_2\rho}{\kappa T} \qquad \rho = \frac{2k_1}{k_2}\kappa T$ Formel von Jeans.

$$FdE = F^x dE^x$$

$$=$$

$$\frac{S}{S^x} = \frac{dE}{dE^x}\frac{\overline{\varepsilon^2}}{\overline{\varepsilon^{x^2}}} = \left(\frac{dE}{dE^x}\right)^2$$

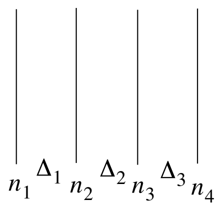
$$F^x S^x + \frac{1}{2}\overline{\varepsilon^{x^2}} \cdot \underbrace{\left(\frac{dE}{dE^x}\right)^2 \cdot \frac{d}{dE}\left(F^x \frac{dE^x}{dE}\right)} = 0$$

$$\underbrace{\frac{dE}{dE^x}\frac{d}{dE^x}\left(F^x \cdot \frac{1}{\frac{dE}{dE^x}}\right)}$$

$$\frac{\partial F^x}{\partial E^x} - \underbrace{F^x \frac{\frac{d^2E}{dE^{x^2}}}{\left(\frac{dE}{dE^x}\right)^2} \cdot \frac{dE}{dE^x}}$$

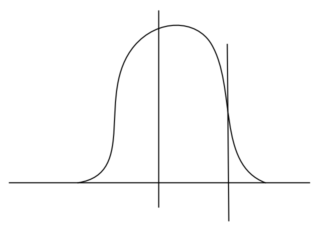
$$F^x \frac{\frac{d^2E}{dE^{x^2}}}{\frac{dE}{dE^x}}$$

Gleichung von derselben Form mit anderem S.

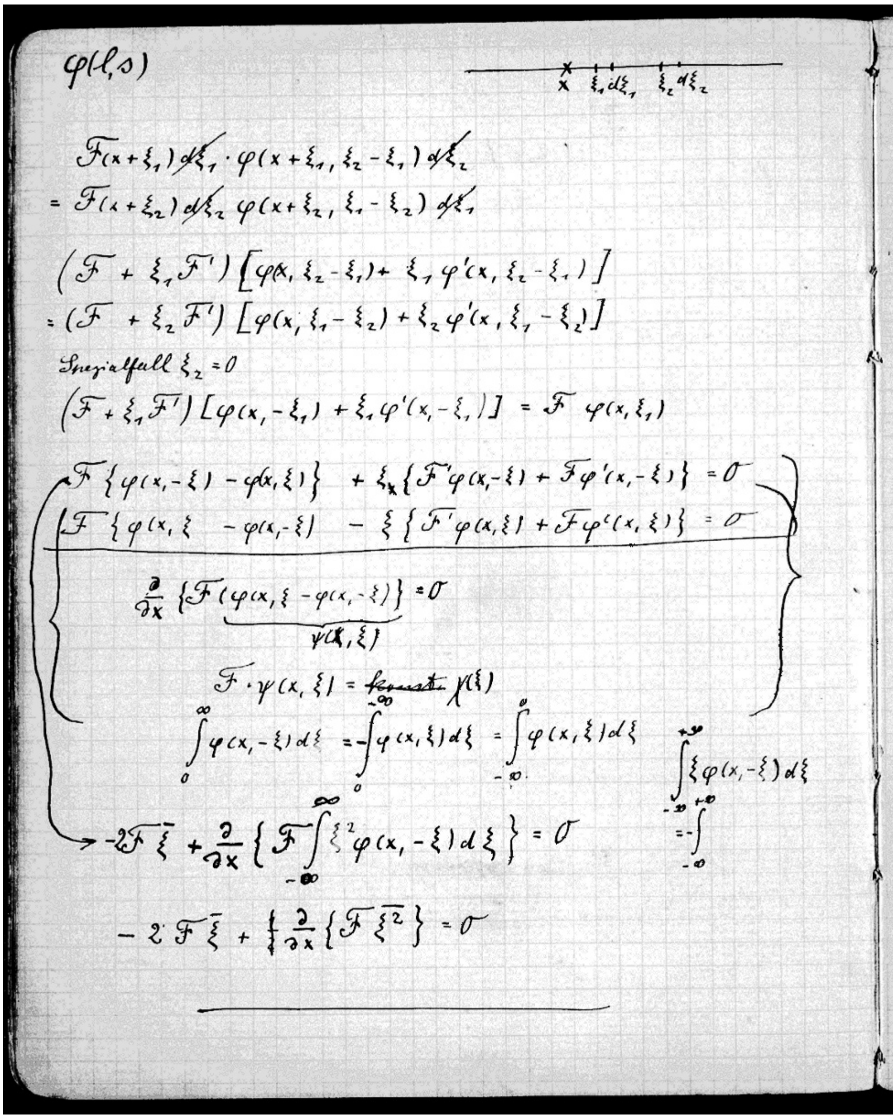


$$\varphi(\Delta)d\Delta$$

$$(\Delta + \delta_\Delta)\varphi(\Delta)d\Delta$$



[p. 03 L]



$\varphi(l, s)$

$x \quad \xi_1 \, d\xi_1 \quad \xi_2 \, d\xi_2$

$$\mathfrak{F}(x+\xi_1)\, \cancel{d\xi_1} \cdot \varphi(x+\xi_1, \xi_2-\xi_1)\, \cancel{d\xi_2}$$
$$= \mathfrak{F}(x+\xi_2)\, \cancel{d\xi_2}\, \varphi(x+\xi_2, \xi_1-\xi_2)\, \cancel{d\xi_1}$$

$$(\mathfrak{F} + \xi_1 \mathfrak{F}')\left[\varphi(x, \xi_2-\xi_1) + \xi_1 \varphi'(x, \xi_2-\xi_1)\right]$$
$$= (\mathfrak{F} + \xi_2 \mathfrak{F}')\left[\varphi(x, \xi_1-\xi_2) + \xi_2 \varphi'(x, \xi_1-\xi_2)\right]$$

Spezialfall $\xi_2 = 0$

$$(\mathfrak{F} + \xi_1 \mathfrak{F}')\left[\varphi(x, -\xi_1) + \xi_1 \varphi'(x, -\xi_1)\right] = \mathfrak{F}\, \varphi(x, \xi_1)$$

$$\mathfrak{F}\{\varphi(x, -\xi) - \varphi(x, \xi)\} + \xi_x\{\mathfrak{F}'\varphi(x, -\xi) + \mathfrak{F}\varphi'(x, -\xi)\} = 0$$
$$\mathfrak{F}\{\varphi(x, \xi - \varphi(x, -\xi)\} - \xi\{\mathfrak{F}'\varphi(x, \xi) + \mathfrak{F}\varphi'(x, \xi)\} = 0$$

$$\frac{\partial}{\partial x}\{\mathfrak{F}\underbrace{(\varphi(x, \xi - \varphi(x, -\xi))}_{\psi(x, \xi)}\} = 0$$

$$\mathfrak{F} \cdot \psi(x, \xi) = \cancel{\text{konst.}}\; \chi(\xi)$$

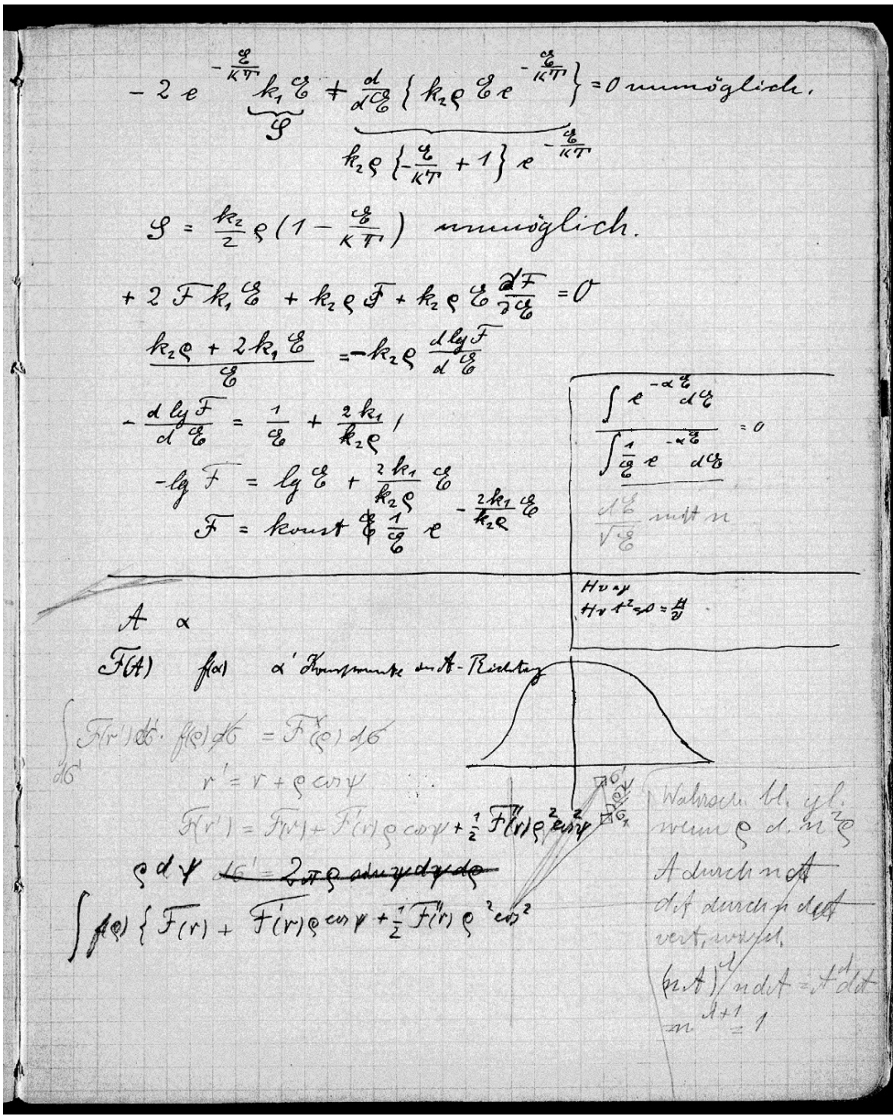
$$\int_0^\infty \varphi(x, -\xi)\, d\xi = -\int_0^{-\infty} \varphi(x, \xi)\, d\xi = \int_{-\infty}^0 \varphi(x, \xi)\, d\xi$$

$$\int_{-\infty}^{+\infty} \xi \varphi(x, -\xi)\, d\xi = -\int_{-\infty}^{+\infty}$$

$$-2\mathfrak{F}\bar{\xi} + \frac{\partial}{\partial x}\left\{\mathfrak{F}\int_{-\infty}^{\infty} \xi^2 \varphi(x, -\xi)\, d\xi\right\} = 0$$

$$-2\mathfrak{F}\bar{\xi} + \frac{1}{2}\frac{\partial}{\partial x}\{\mathfrak{F}\overline{\xi^2}\} = 0$$

[p. 03 R]



$$-2e^{-\frac{\mathfrak{E}}{\kappa T}}\underbrace{k_1\mathfrak{E}}_{\mathcal{S}}+\frac{d}{d\mathfrak{E}}\underbrace{\left\{k_2\varrho\,\mathfrak{E}\,e^{-\frac{\mathfrak{E}}{\kappa T}}\right\}}_{k_2\varrho\left\{-\frac{\mathfrak{E}}{\kappa T}+1\right\}e^{-\frac{\mathfrak{E}}{\kappa T}}}=0 \text{ unmöglich.}$$

$$\mathcal{S}=\frac{k_2}{2}\varrho\left(1-\frac{\mathfrak{E}}{\kappa T}\right) \text{ unmöglich.}$$

$$+2\mathcal{F}k_1\mathfrak{E}+k_2\varrho\mathcal{F}+k_2\varrho\,\mathfrak{E}\frac{\partial\mathcal{F}}{\partial\mathfrak{E}}=0$$

$$\frac{k_2\varrho+2k_1\mathfrak{E}}{\mathfrak{E}}=-k_2\varrho\frac{d\lg\mathcal{F}}{d\mathfrak{E}}$$

$$-\frac{d\lg\mathcal{F}}{d\mathfrak{E}}=\frac{1}{\mathfrak{E}}+\frac{2k_1}{k_2\varrho}$$

$$-\lg\mathcal{F}=\lg\mathfrak{E}+\frac{2k_1}{k_2\varrho}\mathfrak{E}$$

$$\mathcal{F}=\text{konst}\,\frac{1}{\mathfrak{E}}\,e^{-\frac{2k_1}{k_2\varrho}\mathfrak{E}}$$

$$\frac{\int e^{-\alpha\mathfrak{E}}\,d\mathfrak{E}}{\int\frac{1}{\mathfrak{E}}e^{-\alpha\mathfrak{E}}\,d\mathfrak{E}}=0$$

$\frac{d\mathfrak{E}}{\sqrt{\mathfrak{E}}}$ mit n

$H v \varphi$

$H v t^2 \varphi\varrho = \frac{H}{\vartheta}$

A α

$\mathcal{F}(A)$ $f(\alpha)$ α' Komponente in A-Richtung

$$\int\mathcal{F}(r')\,d\sigma'\cdot f(\varrho)\,d\sigma=\mathcal{F}^*(\varrho)\,d\sigma$$

$$r'=r+\varrho\cos\psi$$

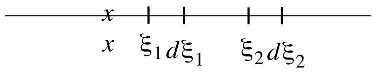
$$\mathcal{F}(r')=\mathcal{F}(r)+\mathcal{F}'(r)\varrho\cos\psi+\frac{1}{2}\mathcal{F}''(r)\varrho^2\cos^2\psi$$

$\varrho\,d\psi$ $\cancel{d\sigma'=2\pi\varrho\sin\psi\,d\psi\,d\varrho}$

$$\int f(\varrho)\left\{F(r)+F'(r)\varrho\cos\psi+\frac{1}{2}F''(r)\varrho^2\cos^2\right.$$

Wahrsch. bl. gl.
wenn ϱ d. $n^2\varrho$
A durch nA
dA durch $n\,dA$
verk. ...
$(nA)^\lambda\,n\,dA=A^\lambda\,dA$
$n^{\lambda+1}=1$

[p. 03 L]



$$\varphi(l, s)$$

$$F(x+\xi_1)\cancel{d\xi_1}\cdot\varphi(x+\xi_1,\,\xi_2-\xi_1)\cancel{d\xi_2}$$

$$= F(x+\xi_2)\cancel{d\xi_2}\,\varphi(x+\xi_2,\,\xi_1-\xi_2)\cancel{d\xi_1}$$

$$(F+\xi_1F')[\varphi(x,\,\xi_2-\xi_1)+\xi_1\varphi'(x,\,\xi_2-\xi_1)]$$

$$= (F+\xi_2F')[\varphi(x,\,\xi_1-\xi_2)+\xi_2\varphi'(x,\,\xi_1-\xi_2)]$$

Spezialfall $\xi_2 = 0$

$$(F+\xi_1F')[\varphi(x,-\xi_1)+\xi_1\varphi'(x,-\xi_1)] = F\ \ \varphi(x,\xi_1)$$

$$F\{\varphi(x,-\xi)-\varphi(x,\xi)\}\quad+\xi_{\cancel{1}}\{F'\varphi(x,-\xi)+F\varphi'(x,-\xi)\} = 0$$

$$F\{\varphi(x,\xi)-\varphi(x,-\xi)\}\quad-\xi\ \{F'\varphi(x,\ \ \xi)+F\varphi'(x,\ \ \xi)\} = 0$$

$$\frac{\partial}{\partial x}\{F(\underbrace{\varphi(x\,,\xi-\varphi(x,-\xi)\}}_{\psi(x,\,\xi)} = 0$$

$$F\cdot\psi(x,\xi) = \cancel{\text{konst.}}\ \chi^{(\xi)}$$

$$\int_0^{\infty}\varphi(x,-\xi)d\xi = -\int_0^{-\infty}\varphi(x,\xi)d\xi = \int_{-\infty}^{0}\varphi(x,\xi)d\xi$$

$$\int_{-\infty}^{+\infty}\xi\varphi(x,-\xi)d\xi$$

$$= -\int_{-\infty}^{+\infty}$$

$$-2F\bar{\xi}+\frac{\partial}{\partial x}\left\{F\int_{-\infty}^{\infty}\xi^2\varphi(x,-\xi)d\xi\right\} = 0$$

$$-2F\bar{\xi}+\frac{\cancel{1}}{\cancel{2}}\frac{\partial}{\partial x}\{F\overline{\xi^2}\} = 0$$

[p. 03 R]

$$-2e^{-\frac{E}{\kappa T}}\underbrace{k_1 E}_{S} + \underbrace{\frac{d}{dE}\left\{k_2\rho E e^{-\frac{E}{\kappa T}}\right\}}_{k_2\rho\left\{-\frac{E}{\kappa T}+1\right\}e^{-\frac{E}{\kappa T}}} = 0 \quad \text{unmöglich.}$$

$$S = \frac{k_2}{2}\rho\left(1-\frac{E}{\kappa T}\right) \quad \text{unmöglich.}$$

$$+2Fk_1E + k_2\rho F + k_2\rho E\frac{dF}{\partial E} = 0$$

$$\frac{k_2\rho + 2k_1E}{E} = -k_2\rho\frac{d\lg F}{dE}$$

$$-\frac{d\lg F}{dE} = \frac{1}{E}+\frac{2k_1}{k_2\rho}$$

$$-\lg F = \lg E + \frac{2k_1}{k_2\rho}E$$

$$F = \text{konst}\ \cancel{E}\frac{1}{E}e^{-\frac{2k_1}{k_2\rho}E}$$

$$\frac{\int e^{-\alpha E}dE}{\underline{\int \frac{1}{E}e^{-\alpha E}dE}} = 0$$

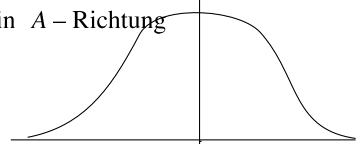
$\frac{dE}{\sqrt{E}}$ mit n.

$H\nu\alpha\gamma$

$H\nu t^2\alpha\beta = \frac{H}{\nu}$

$A \quad \alpha$

$F(A) \quad f(\alpha) \quad \alpha'$ Komponente in A – Richtung



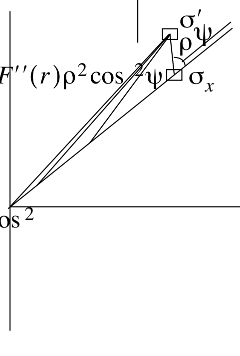
$$\int_{d\sigma'} F(r')d\sigma'\cdot f(\rho)\cancel{d\sigma} = F^x(\rho)\cancel{d\sigma}$$

$$r' = r+\rho\cos\psi$$

$$F(r') = F(r)+F'(r)\rho\cos\psi + \frac{1}{2}F''(r)\rho^2\cos^2\psi$$

$\rho d\psi$ ~~$d\sigma' = 2\pi\rho\sin\psi d\psi d\rho$~~

$$\int f(\rho)\left\{F(r)+F'(r)\rho\cos\psi+\frac{1}{2}F''(r)\rho^2\cos^2\right.$$



Wahrsch. bl. gl.

wenn ρ d. $n^2\rho$

A durch nA

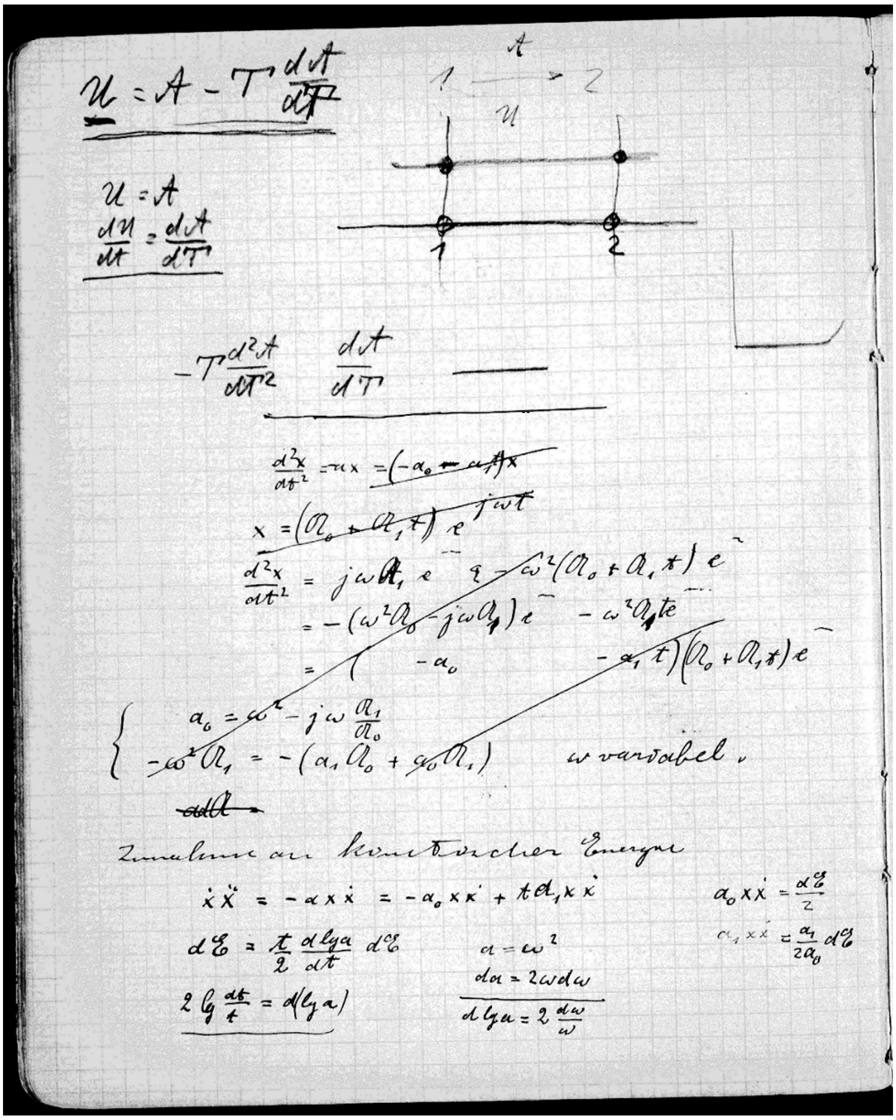
dA durch ndA

vert. wird

$(nA)^\lambda ndA = A^\lambda dt$

$= n^{\lambda+1} = 1$

[p. 04 L]



$$\underline{U = A - T\frac{dA}{dT}}$$

$$U = A$$
$$\frac{dU}{dt} = \frac{dA}{dT}$$

A

$1 \longrightarrow 2$

U

1 2

$$-T\frac{d^2A}{dT^2} \qquad \frac{dA}{dT}$$

$$\cancel{\frac{d^2x}{dt^2} = -ax = (-a_0 - a_1 t)x}$$

$$\cancel{x = (\mathfrak{A}_0 + \mathfrak{A}_1 t)\,e^{j\omega t}}$$

$$\frac{d^2x}{dt^2} = j\omega\mathfrak{A}_1 e^{\cdots} \cancel{- \omega^2(\mathfrak{A}_0 + \mathfrak{A}_1 t)\,e^{\cdots}}$$
$$\cancel{= -(\omega^2\mathfrak{A}_0 - j\omega\mathfrak{A}_1)\,e^{\cdots} - \omega^2\mathfrak{A}_1 t e^{\cdots}}$$
$$\cancel{= (\quad -a_0 \qquad - a_1 t)(\mathfrak{A}_0 + \mathfrak{A}_1 t)\,e^{\cdots}}$$

$$\left\{\begin{array}{l} \cancel{a_0 = \omega^2 - j\omega\frac{\mathfrak{A}_1}{\mathfrak{A}_0}} \\ \cancel{-\omega^2\mathfrak{A}_1 = -(a_1\mathfrak{A}_0 + a_0\mathfrak{A}_1)} \end{array}\right. \qquad \omega \text{ variabel.}$$

~~$d\mathfrak{A}$~~

Zunahme an kinetischer Energie

$$\dot{x}\ddot{x} = -ax\dot{x} = -a_0 x\dot{x} + t a_1 x\dot{x} \qquad a_0 x\dot{x} = \frac{d\mathfrak{E}}{2}$$

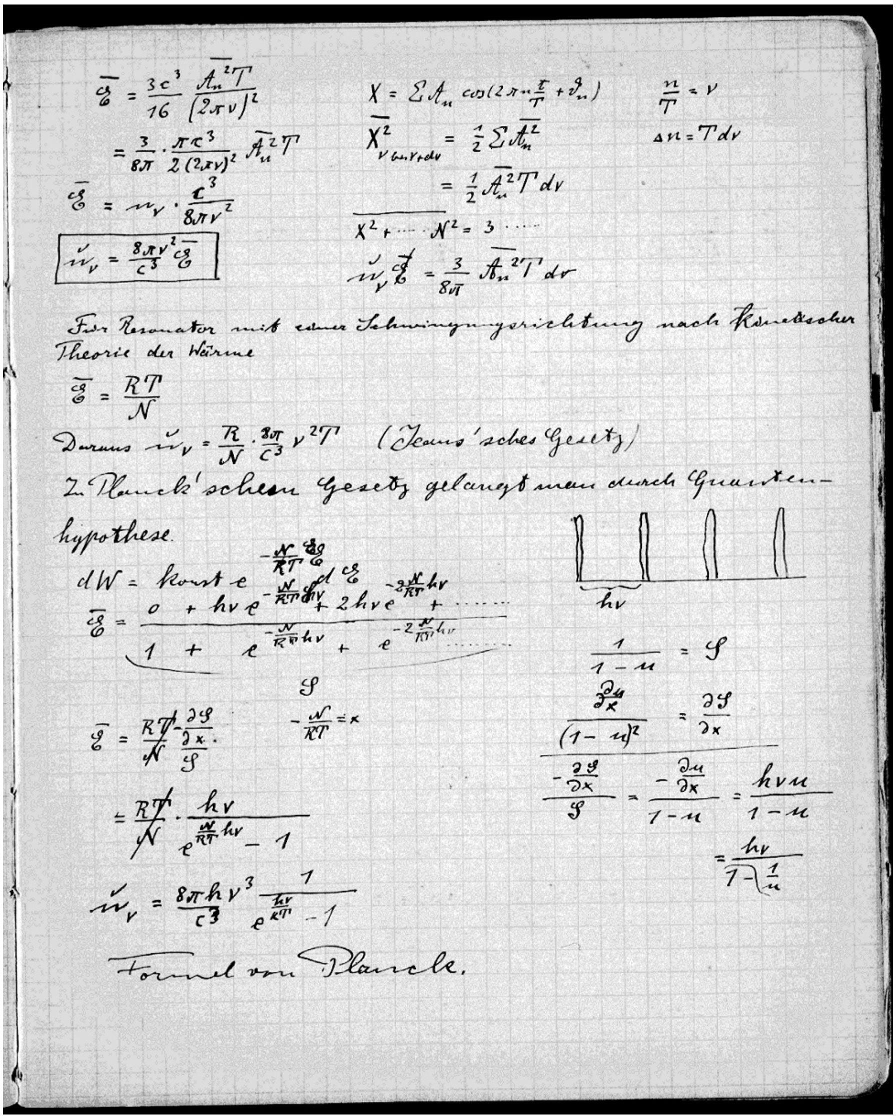
$$a_1 x\dot{x} = \frac{a_1}{2a_0}\,d\mathfrak{E}$$

$$d\mathfrak{E} = \frac{t}{2}\frac{d\lg a}{dt}\,d\mathfrak{E} \qquad a = \omega^2$$

$$da = 2\omega\,d\omega$$

$$\underline{2\lg\frac{dt}{t} = d(\lg a)} \qquad d\lg a = 2\frac{d\omega}{\omega}$$

[p. 04 R]



$$\overline{\mathfrak{E}} = \frac{3c^3}{16}\frac{\overline{A_n^2}T}{(2\pi\nu)^2}$$

$$= \frac{3}{8\pi}\cdot\frac{\pi c^3}{2(2\pi\nu)^2}\overline{A_n^2}T$$

$$\overline{\mathfrak{E}} = u_\nu\cdot\frac{c^3}{8\pi\nu^2}$$

$$\boxed{\check{u}_\nu = \frac{8\pi\nu^2}{c^3}\overline{\mathfrak{E}}}$$

$$X = \sum A_n \cos\left(2\pi n\frac{t}{T} + \vartheta_n\right) \qquad \frac{n}{T} = \nu$$

$$\overline{X^2}_{\nu \text{ bis } \nu+d\nu} = \frac{1}{2}\sum\overline{A_n^2} \qquad \Delta n = T\,d\nu$$

$$= \frac{1}{2}\overline{A_n^2}T\,d\nu$$

$$\overline{X^2 + \cdots Z^2} = 3\cdots$$

$$\check{u}_\nu\cancel{\overline{\mathfrak{E}}} = \frac{3}{8\pi}\overline{A_n^2}T\,d\nu$$

Für Resonator mit einer Schwingungsrichtung nach kinetischer Theorie der Wärme

$$\overline{\mathfrak{E}} = \frac{RT}{N}$$

Daraus $\check{u}_\nu = \frac{R}{N}\cdot\frac{8\pi}{c^3}\nu^2 T$ (Jeans'sches Gesetz)

Zu Planck'schem Gesetz gelangt man durch Quantenhypothese.

$h\nu$

$$dW = \text{konst}\; e^{-\frac{N}{RT}\mathfrak{E}}\,d\mathfrak{E}$$

$$\overline{\mathfrak{E}} = \frac{0 + h\nu e^{-\frac{N}{RT}h\nu} + 2h\nu e^{-2\frac{N}{RT}h\nu} + \cdots\cdots}{\underbrace{1 + e^{-\frac{N}{RT}h\nu} + e^{-2\frac{N}{RT}h\nu}\cdots\cdots}_{S}}$$

$$-\frac{N}{RT} = x$$

$$\frac{1}{1-u} = S$$

$$\frac{\frac{\partial u}{\partial x}}{(1-u)^2} = \frac{\partial S}{\partial x}$$

$$\overline{\mathfrak{E}} = \cancel{\frac{RT}{N}}\frac{-\frac{\partial S}{\partial x}}{S}$$

$$\frac{-\frac{\partial S}{\partial x}}{S} = \frac{-\frac{\partial u}{\partial x}}{1-u} = \frac{h\nu u}{1-u}$$

$$= \frac{h\nu}{1-\frac{1}{u}}$$

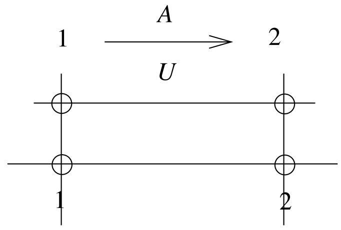
$$= \cancel{\frac{RT}{N}}\cdot\frac{h\nu}{e^{\frac{N}{RT}h\nu}-1}$$

$$\check{u}_\nu = \frac{8\pi h\nu^3}{c^3}\frac{1}{e^{\frac{h\nu}{kT}}-1}$$

Formel von Planck.

[p. 04 L]

$$\underline{U} = A - T\underline{\frac{dA}{dT}}$$



$$U = A$$

$$\frac{dU}{dt} = \frac{dA}{dT}$$

$$-T\frac{d^2A}{dT^2} \qquad \frac{dA}{dT}$$

$$\frac{d^2x}{dt^2} = -ax = \cancel{(-a_0 - a_1 t)x}$$

$$\cancel{x = (A_0 + A_1 t)e^{j\omega t}}$$

$$\frac{d^2x}{dt^2} = j\omega\langle A\rangle A_1 e^{\sim} \quad [-] - \omega^2(A_0 + A_1 t)e^{\sim}$$

$$= -(\omega^2 A_0 - j\omega A_1)e^{\sim} \quad - \omega^2 A_1 t e^{\sim}$$

$$= (\qquad - a_0 \qquad -a_1 t)(A_0 + A_1 t)e^{\sim}$$

$$\begin{cases} a_0 = \omega^2 - j\omega\dfrac{A_1}{A_0} \\ -\omega^2 A_1 = -(a_1 A_0 + a_0 A_1) \end{cases} \quad \omega \text{ variabel.}$$

$$\cancel{adA =}$$

Zunahme an kinetischer Energie

$$\dot{x}\ddot{x} = -ax\dot{x} = -a_0 x\dot{x} + ta_1 x\dot{x} \qquad a_0 x\dot{x} = \frac{dE}{2}$$

$$dE = \frac{t}{2}\frac{d\lg a}{dt}dE \qquad a = \omega^2 \qquad a_1 x\dot{x} = \frac{a_1}{2a_0}dE$$

$$da = 2\omega d\omega$$

$$\underline{2 \lg\frac{dt}{t} = d(\lg a)} \qquad d\lg a = 2\frac{d\omega}{\omega}$$

[p. 04 R]

$$\bar{E} = \frac{3c^3}{16}\frac{\overline{A_n^2}T}{(2\pi\nu)^2} \qquad X = \sum A_n \cos\left(2\pi n\frac{\tau}{T} + \vartheta_n\right) \qquad \frac{n}{T} = \nu$$

$$= \frac{3}{8\pi}\cdot\frac{\pi c^3}{2(2\pi\nu)^2}\overline{A_n^2}T \qquad \overline{X^2_{\nu\cos\nu + d\nu}} = \frac{1}{2}\sum\overline{A_n^2} \qquad \Delta n = Td\nu$$

$$= \frac{1}{2}\overline{A_n^2}Td\nu$$

$$\bar{E} = u_\nu\cdot\frac{c^3}{8\pi\nu^2}$$

$$\overline{X^2 + \cdots\cdots N^2} = 3\cdot\cdots\cdots$$

$$\boxed{u_\nu = \frac{8\pi\nu^2}{c^3}\bar{E}} \qquad u_\nu\cancel{\bar{E}} = \frac{3}{8\pi}\overline{A_n^2}T\cancel{d\nu}$$

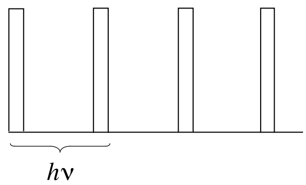
Für Resonator mit einer Schwingungsrichtung nach kinetischer Theorie der Wärme

$$\bar{E} = \frac{RT}{N}$$

Daraus $u_\nu = \frac{R}{N}\cdot\frac{3\pi}{c^3}\nu^2T$ (Jeans'sches Gesetz)

Zu Planck'schem Gesetz gelangt man durch Quantenhypothese.

$$dW = \text{konst}\ e^{-\frac{N}{RT}E}dE$$



$$\bar{E} = \frac{0 + h\nu e^{-\frac{N}{RT}h\nu} + 2h\nu e^{-2\frac{N}{RT}h\nu} + \cdots\cdots\cdots}{\underbrace{1 + e^{-\frac{N}{RT}h\nu} + e^{-2\frac{N}{RT}h\nu}\cdots\cdots\cdots}_{S}}$$

$$\frac{1}{1-u} = S$$

$$\bar{E} = \frac{\cancel{RT}}{\cancel{N}}\frac{-\frac{\partial S}{\partial x}}{S}\cdot \qquad -\frac{N}{RT} = x$$

$$\frac{\frac{\partial u}{\partial x}}{(1-u)^2} = \frac{\partial S}{\partial x}$$

$$= \frac{\cancel{RT}}{\cancel{N}}\cdot\frac{h\nu}{e^{\frac{N}{RT}h\nu} - 1}$$

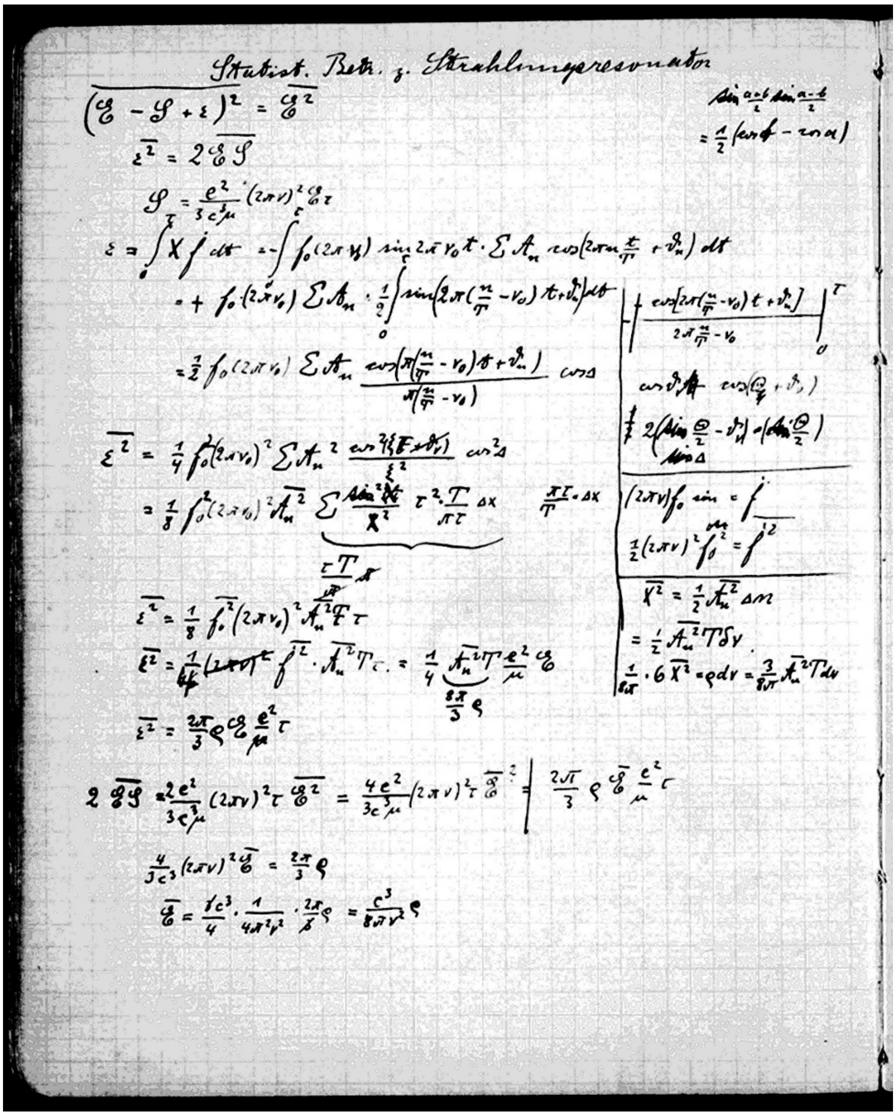
$$\frac{-\frac{\partial S}{\partial x}}{S} = \frac{-\frac{\partial u}{\partial x}}{1-u} = \frac{h\nu u}{1-u}$$

$$= \frac{h\nu}{1 - \frac{1}{u}}$$

$$u_\nu = \frac{8\pi h\nu^3}{c^3}\frac{1}{e^{\frac{h\nu}{RT}} - 1}$$

Formel von Planck.

[p. 05 L]



Statist. Betr. z. Strahlungsresonator

$$\overline{(\mathfrak{E} - \mathfrak{J} + \varepsilon)^2} = \overline{\mathfrak{E}^2}$$

$$\overline{\varepsilon^2} = 2\,\overline{\mathfrak{E}\mathfrak{J}}$$

$$\mathfrak{J} = \frac{e^2}{3c^3\mu}(2\pi\nu)^2\,\mathfrak{E}\tau$$

$$\varepsilon = \int_0^\tau X\dot{f}\,dt = -\int_0^\tau f_0(2\pi\nu_0)\sin 2\pi\nu_0 t \cdot \sum A_n \cos\left(2\pi n\frac{t}{T} + \vartheta_n\right) dt$$

$$= + f_0(2\pi\nu_0)\sum A_n \cdot \frac{1}{2}\int_0^\tau \sin\left(2\pi\left(\frac{n}{T} - \nu_0\right)t + \vartheta_n\right) dt$$

$$= \frac{1}{2} f_0(2\pi\nu_0)\sum A_n \frac{\cos\left(\pi\left(\frac{n}{T} - \nu_0\right)\tau + \vartheta_n\right)}{\pi\left(\frac{n}{T} - \nu_0\right)} \cos\Delta$$

$$\overline{\varepsilon^2} = \frac{1}{4} f_0^2(2\pi\nu_0)^2 \sum A_n^2 \frac{\cos^2(\xi + \vartheta_\nu)}{\xi^2}\cos^2\Delta$$

$$= \frac{1}{8} f_0^2(2\pi\nu_0)^2\overline{A_n^2}\underbrace{\sum \frac{\sin^2 x}{x^2}\,\tau^2 \cdot \frac{T}{\pi\tau}\,\Delta x}_{\tau T} \qquad \frac{\pi\tau}{T} \cdot \Delta x$$

$$\overline{\varepsilon^2} = \frac{1}{8}\overline{f_0^2}(2\pi\nu_0)^2\overline{A_n^2}T\tau$$

$$\overline{\varepsilon^2} = \frac{1}{4}\overline{\dot{f}^2}\cdot\overline{A_n^2}T\tau = \frac{1}{4}\underbrace{\overline{A_n^2}T}_{\frac{8\pi}{3}\varrho}\frac{e^2}{\mu}\mathfrak{E}$$

$$\overline{\varepsilon^2} = \frac{2\pi}{3}\varrho\,\mathfrak{E}\frac{e^2}{\mu}\tau$$

$$2\overline{\mathfrak{E}\mathfrak{J}} = \frac{2e^2}{3c^3\mu}(2\pi\nu)^2\tau\,\overline{\mathfrak{E}^2} = \frac{4e^2}{3c^3\mu}(2\pi\nu)^2\tau\,\overline{\mathfrak{E}}^2 = \frac{2\pi}{3}\varrho\,\overline{\mathfrak{E}}\frac{e^2}{\mu}\tau$$

$$\frac{4}{3c^3}(2\pi\nu)^2\overline{\mathfrak{E}} = \frac{2\pi}{3}\varrho$$

$$\overline{\mathfrak{E}} = \frac{c^3}{4}\cdot\frac{1}{4\pi^2\nu^2}\cdot\frac{2\pi}{}\varrho = \frac{c^3}{8\pi\nu^2}\varrho$$

$$\sin\frac{a+b}{2}\sin\frac{a-b}{2} = \frac{1}{2}(\cos b - \cos a)$$

$$-\left|\frac{\cos\left[2\pi\left(\frac{n}{T} - \nu_0\right)t + \vartheta_n\right]}{2\pi\frac{n}{T} - \nu_0}\right|_0^\tau$$

$$\cos\vartheta \quad \cos\left(\frac{\Theta}{2} + \vartheta_\nu\right)$$

$$2\left(\sin\frac{\Theta}{2} - \vartheta_\nu\right)\left(\sin\frac{\Theta}{2}\right) \quad \sin\Delta$$

$$(2\pi\nu)f_0 \sin = \dot{f}$$

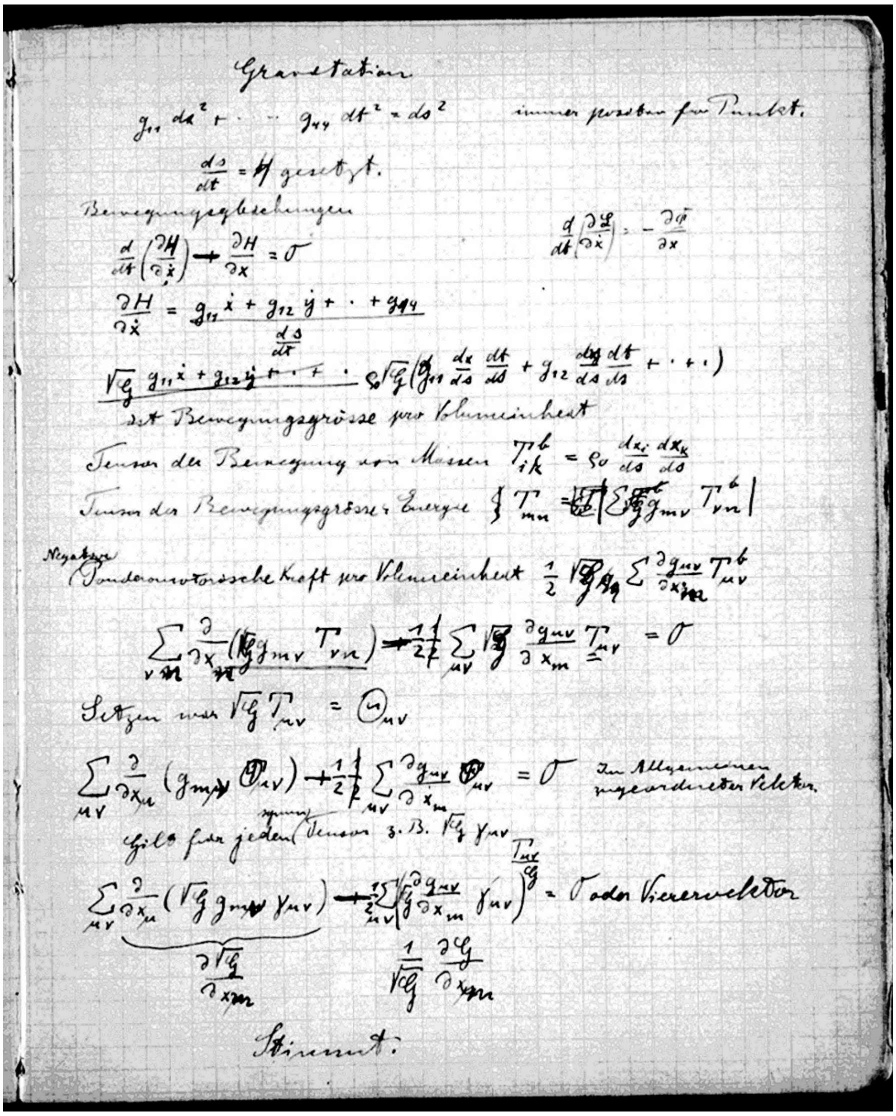
$$\frac{1}{2}(2\pi\nu)^2 f_0^2 = \overline{\dot{f}^2}$$

$$\overline{X^2} = \frac{1}{2}\overline{A_n^2}\,\Delta n$$

$$= \frac{1}{2}\overline{A_n^2}\,T\delta\nu$$

$$\frac{1}{8\pi}\cdot 6\,\overline{X^2} = \varrho\,d\nu = \frac{3}{8\pi}\overline{A_n^2}\,T d\nu$$

[p. 05 R]



Gravitation

$g_{11}\,dx^2 + \cdot\ \cdot\ g_{44}\,dt^2 = ds^2$ immer positiv für Punkt.

$\frac{ds}{dt} = H$ gesetzt.

Bewegungsgleichungen

$$\frac{d}{dt}\left(\frac{\partial H}{\partial \dot{x}}\right) - \frac{\partial H}{\partial x} = 0 \qquad \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}}\right) = -\frac{\partial \Phi}{\partial x}$$

$$\frac{\partial H}{\partial \dot{x}} = \frac{g_{11}\dot{x} + g_{12}\dot{y} + \cdot + g_{14}}{\frac{ds}{dt}}$$

$$\sqrt{G}\,\frac{g_{11}\dot{x} + g_{12}\dot{y} + \cdot + }{\frac{ds}{dt}} = \rho_0\sqrt{G}\left(g_{11}\frac{dx}{ds}\frac{dt}{ds} + g_{12}\frac{dy}{ds}\frac{dt}{ds} + \cdot + \cdot\right)$$

ist Bewegungsgrösse pro Volumeneinheit

Tensor der Bewegung von Massen $T_{ik} = \rho_0 \frac{dx_i}{ds}\frac{dx_k}{ds}$

Tensor der Bewegungsgrösse u. Energie $T_{mn} = \sum \sqrt{G}\, g_{m\nu} T_{\nu n}$

Negative Ponderomotorische Kraft pro Volumeneinheit $\frac{1}{2}\sqrt{G} \sum \frac{\partial g_{\mu\nu}}{\partial x_m} T_{\mu\nu}$

$$\sum_{\nu n} \frac{\partial}{\partial x_n}\left(\sqrt{G}\, g_{m\nu} T_{\nu n}\right) - \frac{1}{2} \sum_{\mu\nu} \sqrt{G}\,\frac{\partial g_{\mu\nu}}{\partial x_m} T_{\mu\nu} = 0$$

Setzen wir $\sqrt{G}\, T_{\mu\nu} = \Theta_{\mu\nu}$

$$\sum_{\mu\nu} \frac{\partial}{\partial x_\mu}\left(g_{m\nu}\Theta_{\mu\nu}\right) - \frac{1}{2} \sum_{\mu\nu} \frac{\partial g_{\mu\nu}}{\partial x_m}\Theta_{\mu\nu} = 0$$ Zu allgemeinen zugeordneten Vektor.

Gilt für jeden symm. Tensor z.B. $\sqrt{G}\,\gamma_{\mu\nu}$ $\frac{T_{\mu\nu}}{G}$

$$\sum_{\mu\nu} \underbrace{\frac{\partial}{\partial x_\mu}\left(\sqrt{G}\, g_{m\nu}\gamma_{\mu\nu}\right)}_{\frac{\partial\sqrt{G}}{\partial x_m}} - \frac{1}{2}\sum_{\mu\nu}\left(\sqrt{G}\,\frac{\partial g_{\mu\nu}}{\partial x_m}\gamma_{\mu\nu}\right) = 0 \text{ oder Vierervektor}$$

$$\frac{1}{\sqrt{G}}\frac{\partial G}{\partial x_m}$$

Stimmt.

[p. 05 L]

Statist. Betr. z. Strahlungsresonator

$$\overline{(E-S+\varepsilon)^2} = \overline{E^2}$$

$$\sin\frac{a+b}{2}\sin\frac{a-b}{2} = \frac{1}{2}(\cos\, b - \cos\, a)$$

$$\overline{\varepsilon^2} = 2\overline{ES}$$

$$S = \frac{e^2}{3c^3\mu}(2\pi\nu)^2 E\tau$$

$$\varepsilon = \int_0^\tau x\dot{f}\,dt = -\int_0^\tau f_0(2\pi\nu_0)\sin 2\pi\nu_0 t \cdot \sum A_n \cos\left(2\pi n\frac{t}{T} + \vartheta_n\right)dt$$

$$= +\, f_0(2\pi\nu_0)\sum A_n \cdot \frac{1}{2}\int_0^\tau \sin\left(2\pi\left(\frac{n}{T}-\nu_0\right)t + \vartheta_n\right)dt$$

$$-\left|\frac{\cos\left[2\pi\left(\frac{n}{T}-\nu_0\right)t+\vartheta_n\right]}{2\pi\frac{n}{T}-\nu_0}\right|_0^\tau$$

$$= \frac{1}{2}f_0(2\pi\nu_0)\sum A_n \frac{\cos\left(\pi\left(\frac{n}{T}-\nu_0\right)t+\vartheta_n\right)}{\pi\left(\frac{n}{T}-\nu_0\right)}\cos\Delta$$

$$\cos\vartheta_\nu \cancel{A_n} - \cos(\Theta + \vartheta_\nu)$$

$$\overline{\varepsilon^2} = \frac{1}{4}f_0^2(2\pi\nu_0)^2\sum A_n{}^2 \frac{\cos^2(\xi \cancel{t + \vartheta_\nu})}{\xi^2}\cos^2\Delta$$

$$\cancel{\tfrac{1}{2}}2\left(\sin\frac{\Theta}{2} - \vartheta_\nu\right)\cdot\left(\sin\frac{\Theta}{2}\right)$$

$$\sin\Delta$$

$$= \frac{1}{8}f_0^2(2\pi\nu_0)^2\overline{A_n^2}\underbrace{\sum\frac{\sin^2\langle\vartheta_2\rangle x}{x^2}\tau^2\cdot\frac{T}{\pi\tau}\Delta x}_{\frac{\tau T}{\cancel{\pi}}\cancel{\pi}} \qquad \frac{\pi\tau}{T} = \Delta x$$

$$(2\pi\nu)f_0\sin = \dot{f}$$

$$\frac{1}{2}(2\pi\nu)^2 f_0^2 = \overline{\dot{f}^2}$$

$$\overline{\varepsilon^2} = \frac{1}{8}\overline{f_0^2}(2\pi\nu_0)^2\overline{A_n^2}T\tau$$

$$\overline{x^2} = \frac{1}{2}\overline{A_n^2}\Delta n$$

$$= \frac{1}{2}\overline{A_n^2}T\delta\nu$$

$$\overline{\varepsilon^2} = \frac{1}{4}\cancel{(2\pi\nu)^2}\overline{\dot{f}^2}\cdot\overline{A_n^2}T\bar{c}. = \frac{1}{4}\underbrace{\overline{A_n^2}T}_{\frac{8\pi}{3}\rho}\frac{e^2}{\mu}E$$

$$\frac{1}{8\pi}\cdot 6\overline{x^2} = \rho\, d\nu = \frac{3}{8\pi}\overline{A_n^2}T d\nu$$

$$\overline{\varepsilon^2} = \frac{2\pi}{3}\rho E\frac{e^2}{\mu}\tau$$

$$2\overline{ES} = \frac{2e^2}{3c^3\mu}(2\pi\nu)^2\tau\overline{E^2} = \frac{4e^2}{3c^3\mu}(2\pi\nu)^2\tau\bar{E}^2 = \left|\ \frac{2\pi}{3}\rho\bar{E}\frac{e^2}{\mu}\tau\right.$$

$$\frac{4}{3c^3}(2\pi\nu)^2\bar{E} = \frac{2\pi}{3}\rho$$

$$\bar{E} = \frac{\cancel{3}c^3}{4}\cdot\frac{1}{4\pi^2\nu^2}\cdot\frac{2\pi}{\cancel{3}}\rho = \frac{c^3}{8\pi\nu^2}\rho$$

[p. 05 R]

Gravitation

$g_{11}dx^2 + \cdot \cdot \cdot \; g_{44}dt^2 = ds^2$ immer positiv für Punkt.

$\frac{ds}{dt} = H$ gesetzt.

Bewegungsgleichungen

$$\frac{d}{dt}\left(\frac{\partial H}{\partial \dot{x}}\right)\langle + \rangle - \frac{\partial H}{\partial x} = 0 \qquad \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}}\right) = -\frac{\partial \Phi}{\partial x}$$

$$\frac{\partial H}{\partial \dot{x}} = \frac{g_{11}\dot{x} + g_{12}\dot{y} + \cdot + g_{14}}{\frac{ds}{dt}}$$

$$\cancel{\frac{\sqrt{G}g_{11}\dot{x} + g_{12}\dot{y} + \cdot + \cdot}{}} \qquad \rho_0\sqrt{G}\left(g_{11}\frac{dx}{ds}\frac{dt}{ds} + g_{12}\frac{dy}{ds}\frac{dt}{ds} + \cdot + \cdot\right)$$

ist Bewegungsgrösse pro Volumeinheit

Tensor der Bewegung von Massen $T^b_{ik} = \rho_0\frac{dx_i}{ds}\frac{dx_\kappa}{ds}$

Tensor der Bewegungsgrösse u Energie $T_{mn} = \left|\sum \sqrt{G}g_{m\nu}T^b_{\nu n}\right|$

Negative

Ponderomotorische Kraft pro Volumeinheit $\frac{1}{2}\sqrt{G}\sum\frac{\partial g_{\mu\nu}}{\partial x_m}T^b_{\mu\nu}$

$$\sum_{\nu n}\frac{\partial}{\partial x_n}(\underline{\sqrt{G}g_{m\nu}T_{\nu n}})\langle + \rangle - \frac{1}{2}\left\langle\frac{1}{2}\right\rangle\sum_{\mu\nu}\sqrt{G}\frac{\partial g_{\mu\nu}}{\partial x_m}\underline{\underline{T}}_{\mu\nu} = 0$$

Setzen wir $\sqrt{G}T_{\mu\nu} = \Theta_{\mu\nu}$

$$\sum_{\mu\nu}\frac{\partial}{\partial x_\mu}(g_{m\nu}\Theta_{\mu\nu})\langle + \rangle - \frac{1}{2}\left\langle\frac{1}{2}\right\rangle\sum_{\mu\nu}\frac{\partial g_{\mu\nu}}{\partial x_m}\Theta_{\mu\nu} = 0$$

Im Allgemeinen zugeordneter Vektor.

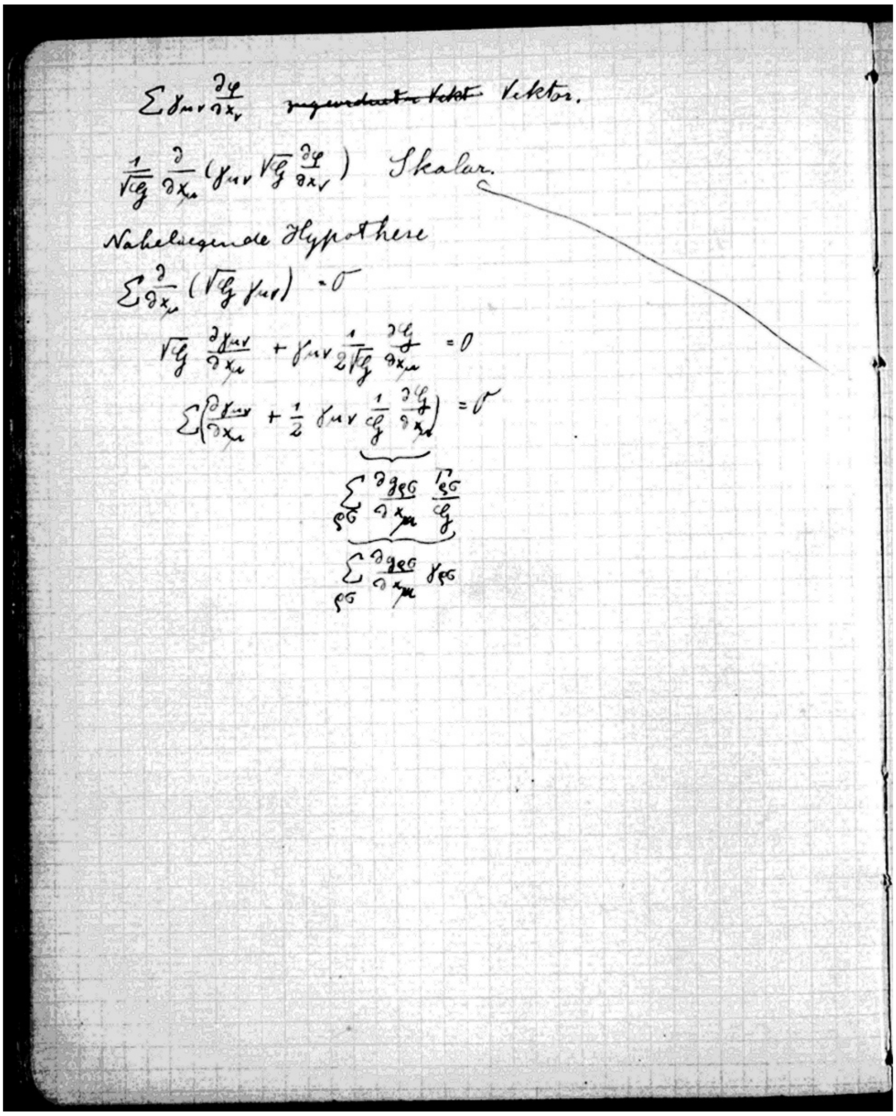
Gilt für jeden symm. Tensor z. B. $\sqrt{G}\gamma_{\mu\nu}$

$$\underbrace{\sum_{\mu\nu}\frac{\partial}{\partial x_\mu}(\sqrt{G}g_{m\nu}\gamma_{\mu\nu})}_{\frac{\partial\sqrt{G}}{\partial x_m}}\langle + \rangle - \frac{1}{2}\underbrace{\sum_{\mu\nu}\left(\sqrt{G}\frac{\partial g_{\mu\nu}}{\partial x_m}\overset{\frac{\Gamma_{\mu\nu}}{G}}{\gamma_{\mu\nu}}\right)}_{\frac{1}{\sqrt{G}}\frac{\partial G}{\partial x_m}} = 0$$

oder Vierervektor

Stimmt.

[p. 06 L]



$\sum \gamma_{\mu\nu} \frac{\partial \varphi}{\partial x_\nu}$ ~~zugeordneter Vektor~~ Vektor.

$\frac{1}{\sqrt{G}} \frac{\partial}{\partial x_\mu}\left(\gamma_{\mu\nu}\sqrt{G}\frac{\partial\varphi}{\partial x_\nu}\right)$ Skalar.

Naheliegende Hypothese

$$\sum \frac{\partial}{\partial x_\nu}\left(\sqrt{G}\,\gamma_{\mu\nu}\right) = 0$$

$$\sqrt{G}\,\frac{\partial\gamma_{\mu\nu}}{\partial x_\nu} + \gamma_{\mu\nu}\frac{1}{2\sqrt{G}}\frac{\partial G}{\partial x_\mu} = 0$$

$$\sum\left(\frac{\partial\gamma_{\mu\nu}}{\partial x_\nu} + \frac{1}{2}\gamma_{\mu\nu}\underbrace{\frac{1}{G}\frac{\partial G}{\partial x_\mu}}_{\underbrace{\sum_{\rho\sigma}\frac{\partial g_{\rho\sigma}}{\partial x_\mu}\frac{\Gamma_{\rho\sigma}}{G}}_{\sum_{\rho\sigma}\frac{\partial g_{\rho\sigma}}{\partial x_\mu}\gamma_{\rho\sigma}}}\right) = 0$$

[p. 06 R]

$\alpha_\nu = \frac{\partial\varphi}{\partial x_\nu}$ zugeordneter Vektor.

$\sum \frac{1}{\sqrt{G}}\frac{\partial}{\partial x_\nu}(\sqrt{G}\,\alpha_\nu)$

$\sum \frac{1}{\sqrt{\Gamma}}\frac{\partial}{\partial \xi_\nu}(\sqrt{\Gamma}\,\alpha_\nu \gamma_{\mu\nu})$

$\xi_\mu = \sum g_{\mu\nu} x_\nu$

$x_\nu = \sum \gamma_{\mu\nu}\xi_\nu$

$\frac{\partial}{\partial\xi_\nu} = \sum \gamma_{\nu\sigma}\frac{\partial}{\partial x_\sigma}$

$\sum \sqrt{G}\,\gamma_{\nu\sigma}\frac{\partial}{\partial x_\sigma}\left(\frac{1}{\sqrt{G}}\frac{\partial\varphi}{\partial x_\nu}\right)$ Skalar

$G' = P^2 G$

$\Gamma' = \frac{1}{P^2}\Gamma$

$\sum_{\nu\sigma} \gamma_{\nu\sigma}\frac{\partial^2\varphi}{\partial x_\nu\partial x_\sigma} + \gamma_{\nu\sigma}\frac{\partial\varphi}{\partial x_\nu}\cdot\frac{1}{2G}\frac{\partial G}{\partial x_\sigma}$ Skalar

$\sum \gamma_{\nu\sigma}\frac{\partial^2\varphi}{\partial x\partial z} + \gamma_{\nu\sigma}\frac{\partial\varphi}{\partial x_\nu}\frac{1}{2G}\frac{\partial G}{\partial x_\sigma} + \frac{\partial\varphi}{\partial x_\nu}\frac{\partial\gamma_{\sigma\nu}}{\partial x_\sigma}$

Soll es nur einen derartigen Skalar geben so muss $\sum_\sigma \frac{\partial\gamma_{\sigma\nu}}{\partial x_\sigma} = 0$

$\sum_{\mu\nu\rho\sigma}\gamma_{\mu\nu}\frac{\partial}{\partial x_\mu}\left(\gamma_{\rho\sigma}\frac{\partial g_{\rho\sigma}}{\partial x_\nu}\right) = 0$

oder $\sum_{\mu\nu\rho\sigma}\gamma_{\mu\nu}\frac{\partial}{\partial x_\mu}\left(g_{\rho\sigma}\frac{\partial\gamma_{\rho\sigma}}{\partial x_\nu}\right) = 0.$

$$\begin{matrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & c^2 \end{matrix}$$

Spezialfall $\mu=\nu$ $\rho=\sigma$ $g_{11}=g_{22}=g_{33}=-1$ $g_{44}=c^2$ $G=-c^2$

$\gamma_{11}=\gamma_{22}=\gamma_{33}=+1$ $\gamma_{44}=\frac{1}{c^2}$

$\rho=\sigma=4$ $\mu=\nu=1\,2\,3$

$\sum_\nu \frac{\partial}{\partial x_\nu}\left(+\frac{1}{c^2}\cdot\underbrace{\frac{\partial c^2}{\partial x_\nu}}_{c\frac{\partial c}{\partial x_\nu}}\right) = 0$ bezw. $\sum \frac{\partial}{\partial x_\nu}\left(c^2\frac{\partial \frac{1}{c^2}}{\partial x_\nu}\right) = 0$

$\frac{\partial^2 \lg c}{\partial x_\nu^2}$ So nicht unterscheidbar.

[p. 06 L]

$$\sum \gamma_{\mu\nu} \frac{\partial \varphi}{\partial x_\nu}$$ ⟨zugeordneter Vekt⟩ Vektor.

$$\frac{1}{\sqrt{G}} \frac{\partial}{\partial x_\mu}\left(\gamma_{\mu\nu} \sqrt{G} \frac{\partial \varphi}{\partial x_\nu}\right)$$ Skalar.

Naheliegende Hypothese

$$\sum \frac{\partial}{\partial x_\mu}(\sqrt{G}\gamma_{\mu\nu}) = 0$$

$$\sqrt{G}\frac{\partial \gamma_{\mu\nu}}{\partial x_\mu} + \gamma_{\mu\nu}\frac{1}{2\sqrt{G}}\frac{\partial G}{\partial x_\mu} = 0$$

[connects p. 06 R]

$$\sum\left(\frac{\partial \gamma_{\mu\nu}}{\partial x_\mu} + \frac{1}{2}\gamma_{\mu\nu}\underbrace{\frac{1}{G}\frac{\partial G}{\partial x_\mu}}_{\underbrace{\sum_{\rho\sigma}\frac{\partial g_{\rho\sigma}}{\partial x_\mu}\frac{\Gamma_{\rho\sigma}}{G}}_{\sum_{\rho\sigma}\frac{\partial g_{\rho\sigma}}{\partial x_\mu}\gamma_{\rho\sigma}}}\right) = 0$$

[p. 06 R]

$$\alpha_\nu = \frac{\partial\varphi}{\partial x_\nu} \qquad \text{zugeordneter Vektor}$$

$$\sum \frac{1}{\sqrt{G}}\frac{\partial}{\partial x_\nu}(\sqrt{G}a_\nu \cancel{\gamma_{\mu\nu}}))$$

$$\xi_\mu = \sum g_{\mu\nu}x_\nu$$

$$x_\nu = \sum \gamma_{\mu\nu}\xi_\nu$$

$$\sum \frac{1}{\sqrt{\Gamma}}\frac{\partial}{\partial \xi_\nu}(\sqrt{\Gamma}\alpha_\nu \cancel{g}_{\mu\nu}) \qquad \frac{\partial}{\partial \xi_\nu} = \sum \gamma_{\nu\sigma}\frac{\partial}{\partial x_\sigma}$$

$$\sum \sqrt{G}\gamma_{\nu\sigma}\frac{\partial}{\partial x_\sigma}\left(\frac{1}{\sqrt{G}}\frac{\partial\varphi}{\partial x_\nu}\right) \qquad \text{Skalar}$$

$$G' = P^2 G$$

$$\Gamma' = \frac{1}{P^2}\Gamma$$

$$\sum_{\nu\sigma} \gamma_{\nu\sigma}\frac{\partial^2\varphi}{\partial x_\nu \partial x_\sigma} + \gamma_{\nu\sigma}\frac{\partial\varphi}{\partial x_\nu} \cdot \frac{1}{2G}\frac{\partial G}{\partial x_\sigma} \qquad \text{Skalar}$$

[connects p. 06 L]

$$\sum \gamma_{\nu\sigma}\frac{\partial^2\varphi}{\partial x \partial z} + \gamma_{\nu\sigma}\frac{\partial\varphi}{\partial x_\nu}\frac{1}{2G}\frac{\partial G}{\partial x_\sigma} + \frac{\partial\varphi}{\partial x_\nu}\frac{\partial\gamma_{\sigma\nu}}{\partial x_\sigma}$$

Soll es nur <u>einen</u> derartigen Skalar geben so muss $\sum_\sigma \frac{\partial\gamma_{\sigma\nu}}{\partial x_\sigma} = 0$

$$\sum_{\mu\nu\rho\sigma} \gamma_{\mu\nu}\frac{\partial}{\partial x_\mu}\left(\gamma_{\rho\sigma}\frac{\partial g_{\rho\sigma}}{\partial x_\nu}\right) = 0$$

$$\begin{matrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & c^2 \end{matrix}$$

oder $\sum_{\mu\nu\rho\sigma} \gamma_{\mu\nu}\frac{\partial}{\partial x_\mu}\left(g_{\rho\sigma}\frac{\partial\gamma_{\rho\sigma}}{\partial x_\nu}\right) = 0.$

$$G = -c^2$$

Spezialf<ä>all<e> $\mu = \nu \quad \rho = \sigma \quad g_{11} = g_{22} = g_{33} = -1 \quad g_{44} = c^2$

$$\gamma_{11} = \gamma_{22} = \gamma_{33} = +1 \quad \gamma_{44} = -\frac{1}{c^2}$$

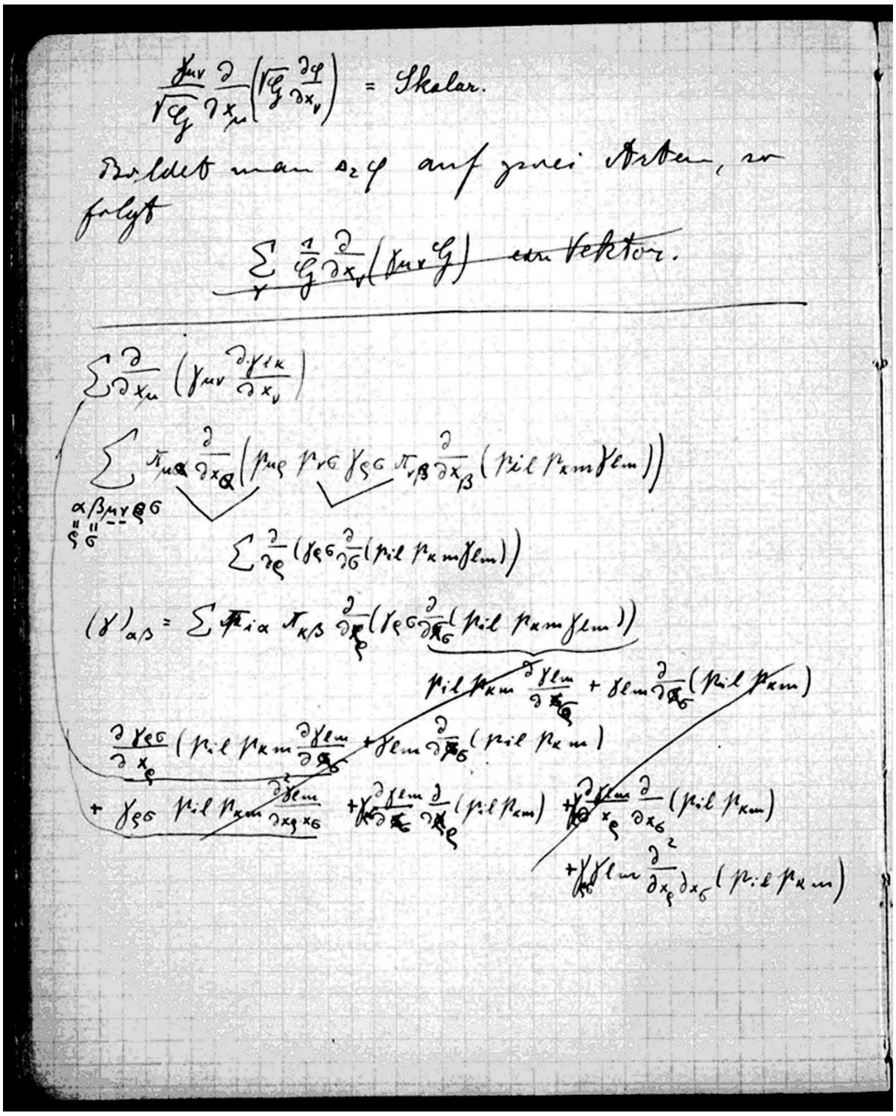
$$\rho = \sigma = 4 \qquad \mu = \nu = 1 \ 2 \ 3$$

$$\sum_\nu \frac{\partial}{\partial x_\nu}\left(+\frac{1}{c^2} \cdot \frac{\partial c^2}{\partial x_\nu}\right) = 0 \qquad \text{bezw} \qquad \sum \frac{\partial}{\partial x_\nu}\left(c^2 \frac{\partial \frac{1}{c^2}}{\partial x_\nu}\right) = 0$$

$$c\frac{\partial c}{\partial x_\nu}$$

$$\frac{\partial^2 \lg c}{\partial x_\nu^2}$$ So nicht unterscheidbar.

[p. 07 L]



$$\frac{\gamma_{\mu\nu}}{\sqrt{G}}\frac{\partial}{\partial x_\mu}\left(\sqrt{G}\frac{\partial\varphi}{\partial x_\nu}\right) = \text{Skalar.}$$

Bildet man $\Delta_2\varphi$ auf zwei Arten, so folgt

$$\cancel{\sum_\nu \frac{1}{G}\frac{\partial}{\partial x_\nu}(\gamma_{\mu\nu}G)} \quad \text{~~ein Vektor.~~}$$

$$\sum \frac{\partial}{\partial x_\mu}\left(\gamma_{\mu\nu}\frac{\partial \gamma_{ik}}{\partial x_\nu}\right)$$

$$\sum_{\substack{\alpha\beta\mu\nu\varrho\sigma \\ \varrho = \sigma}} \pi_{\mu\alpha}\frac{\partial}{\partial x_\alpha}\left(p_{\mu\varrho}\, p_{\nu\sigma}\, \gamma_{\varrho\sigma}\, \pi_{\nu\beta}\frac{\partial}{\partial x_\beta}(p_{il}\, p_{km}\, \gamma_{lm})\right)$$

$$\sum \frac{\partial}{\partial \varrho}\left(\gamma_{\varrho\sigma}\frac{\partial}{\partial \sigma}(p_{il}\, p_{km}\, \gamma_{lm})\right)$$

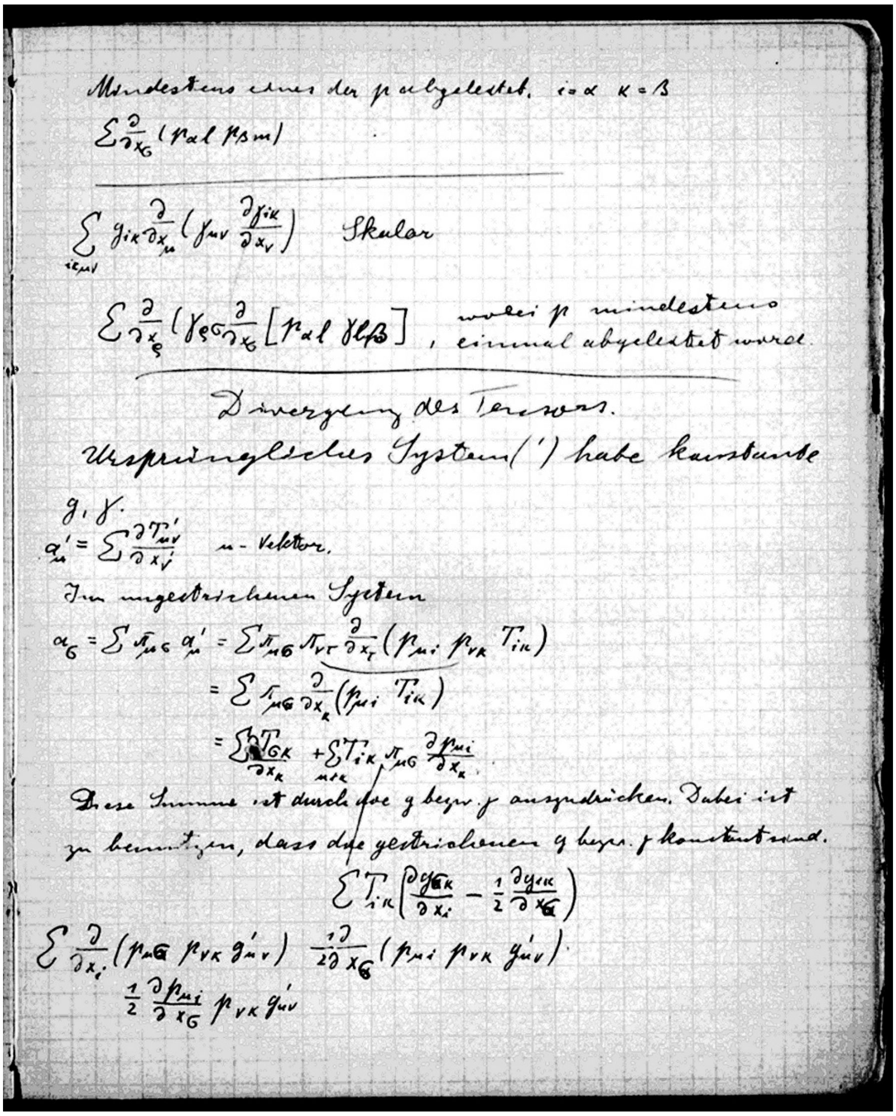
$$(\gamma)_{\alpha\beta} = \sum \pi_{i\alpha}\, \pi_{k\beta}\frac{\partial}{\partial x_\varrho}\left(\gamma_{\varrho\sigma}\underbrace{\frac{\partial}{\partial x_\sigma}(p_{il}\, p_{km}\, \gamma_{lm})}_{p_{il}\, p_{km}\frac{\partial \gamma_{lm}}{\partial x_\sigma} + \gamma_{lm}\frac{\partial}{\partial x_\sigma}(p_{il}\, p_{km})}\right)$$

$$\frac{\partial \gamma_{\varrho\sigma}}{\partial x_\varrho}\left(p_{il}\, p_{km}\frac{\partial \gamma_{lm}}{\partial x_\sigma} + \gamma_{lm}\frac{\partial}{\partial x_\sigma}(p_{il}\, p_{km})\right)$$

$$+ \gamma_{\varrho\sigma}\, p_{il}\, p_{km}\frac{\partial^2 \gamma_{lm}}{\partial x_\varrho \partial x_\sigma} + \gamma_{\varrho\sigma}\frac{\partial \gamma_{lm}}{\partial x_\sigma}\frac{\partial}{\partial x_\varrho}(p_{il}\, p_{km}) + \gamma_{\varrho\sigma}\frac{\partial \gamma_{lm}}{\partial x_\varrho}\frac{\partial}{\partial x_\sigma}(p_{il}\, p_{km})$$

$$+ \gamma_{\varrho\sigma}\gamma_{lm}\frac{\partial^2}{\partial x_\varrho \partial x_\sigma}(p_{il}\, p_{km})$$

[p. 07 R]



Mindestens eines der p abgeleitet. $i = \alpha \quad \kappa = \beta$

$\sum \frac{\partial}{\partial x_\sigma}(p_{\alpha l}\, p_{\beta m})$

$$\sum_{ik\mu\nu} g_{ik} \frac{\partial}{\partial x_\mu}\left(\gamma_{\mu\nu}\frac{\partial \gamma_{ik}}{\partial x_\nu}\right) \quad \text{Skalar}$$

$$\sum \frac{\partial}{\partial x_\rho}\left(\gamma_{\rho\sigma}\frac{\partial}{\partial x_\sigma}\left[p_{\alpha l}\,\gamma_{l\beta}\right]\right)$$, wobei p mindestens einmal abgeleitet werde.

Divergenz des Tensors.

Ursprüngliches System (') habe konstante g, γ.

$\alpha'_\mu = \sum \frac{\partial T'_{\mu\nu}}{\partial x'_\nu}$ μ-Vektor.

Im ungestrichenen System

$$\alpha_\sigma = \sum \pi_{\mu\sigma}\alpha'_\mu = \sum \pi_{\mu\sigma}\pi_{\nu\tau}\frac{\partial}{\partial x_\tau}(p_{\mu i}\, p_{\nu k}\, T_{ik})$$

$$= \sum \pi_{\mu\sigma}\frac{\partial}{\partial x_k}(p_{\mu i}\, T_{ik})$$

$$= \sum \frac{\partial T_{\sigma k}}{\partial x_k} + \sum_{\mu i k} T_{ik}\,\pi_{\mu\sigma}\frac{\partial p_{\mu i}}{\partial x_k}.$$

Diese Summe ist durch die g bezw. γ auszudrücken. Dabei ist zu benutzen, dass die gestrichenen g bezw. γ konstant sind.

$$\sum T_{ik}\left(\frac{\partial g_{\sigma k}}{\partial x_i} - \frac{1}{2}\frac{\partial g_{ik}}{\partial x_\sigma}\right)$$

$$\sum \frac{\partial}{\partial x_i}(p_{\mu\sigma}\, p_{\nu k}\, g'_{\mu\nu}) \quad \frac{1\,\partial}{2\,\partial x_\sigma}(p_{\mu i}\, p_{\nu k}\, g'_{\mu\nu})$$

$$\frac{1}{2}\frac{\partial p_{\mu i}}{\partial x_\sigma}\, p_{\nu k}\, g'_{\mu\nu}$$

[p. 07 L]

$$\frac{\gamma_{\mu\nu}}{\sqrt{G}}\frac{\partial}{\partial x_\mu}\left(\sqrt{G}\frac{\partial\varphi}{\partial x_\nu}\right) = \text{Skalar.}$$

Bildet man $\Delta_2\varphi$ auf zwei Arten, so folgt

$$\cancel{\sum_\nu \frac{1}{G}\frac{\partial}{\partial x_\nu}(\gamma_{\mu\nu}G) \quad \text{ein Vektor.}}$$

$$\sum\frac{\partial}{\partial x_\mu}\left(\gamma_{\mu\nu}\frac{\partial\gamma_{i\kappa}}{\partial x_\nu}\right)$$

$$\sum_{\substack{\alpha\beta\mu\nu\rho\sigma\\ \rho\sigma}}\pi_{\mu\alpha}\frac{\partial}{\partial x_\alpha}\left(p_{\mu\rho}p_{\nu\sigma}\gamma_{\rho\sigma}\pi_{\nu\beta}\frac{\partial}{\partial x_\beta}(p_{il}p_{km}\gamma_{lm})\right)$$

$$\sum\frac{\partial}{\partial\rho}\left(\gamma_{\rho\sigma}\frac{\partial}{\partial\sigma}(p_{il}p_{\kappa m}\gamma_{lm})\right)$$

$$(\gamma)_{\alpha\beta} = \sum\pi_{i\alpha}\pi_{\kappa\beta}\frac{\partial}{\partial x_\rho}\left(\gamma_{\rho\sigma}\underbrace{\frac{\partial}{\partial x_\sigma}(p_{il}p_{\kappa m}\gamma_{lm})}\right)$$

$$\cancel{p_{il}p_{\kappa m}\frac{\partial\gamma_{lm}}{\partial x_\sigma}+\gamma_{lm}\frac{\partial}{\partial x_\sigma}(p_{il}p_{\kappa m})}$$

$$\frac{\partial\gamma_{\rho\sigma}}{\partial x_\rho}(p_{il}p_{\kappa m}\frac{\partial\gamma_{lm}}{\partial x_\sigma}+\gamma_{lm}\frac{\partial}{\partial x_\sigma}(p_{il}p_{\kappa m})$$

$$+\gamma_{\rho\sigma}p_{il}p_{\kappa m}\frac{\partial^2\gamma_{lm}}{\partial x_\rho\cdot x_\sigma}+\gamma_{\rho\sigma}\frac{\partial\gamma_{lm}}{\partial x_\sigma}\frac{\partial}{\partial x_\rho}(p_{il}p_{\kappa m})+\gamma_{\rho\sigma}\frac{\partial\gamma_{lm}}{\partial x_\rho}\frac{\partial}{\partial x_\sigma}(p_{il}p_{\kappa m})$$

$$+\gamma_{\rho\sigma}\gamma_{lm}\frac{\partial^2}{\partial x_\rho x_\sigma}(p_{il}p_{\kappa m})$$

[p. 07 R]

Mindestens eines der p abgeleitet. $i = \alpha \quad \kappa = \beta$

$$\sum \frac{\partial}{\partial x_\sigma}(p_{\alpha l} p_{\beta m})$$

$$\sum_{i\kappa\mu\nu} g_{i\kappa} \frac{\partial}{\partial x_\mu}\left(\gamma_{\mu\nu} \frac{\partial \gamma_{i\kappa}}{\partial x_\nu}\right) \qquad \text{Skalar}$$

$$\sum \frac{\partial}{\partial x_\rho}(\gamma_{\rho\sigma} \frac{\partial}{\partial x_\sigma}[p_{\alpha l} \gamma_{l\beta}] \quad , \quad \text{wobei } p \text{ mindestens einmal abgeleitet wird.}$$

Divergenz des Tensors.

Ursprüngliches System ($'$) habe konstante

g, γ.

$$a'_\mu = \sum \frac{\partial T'_{\mu\nu}}{\partial x'_\nu} \qquad \mu \text{ - Vektor.}$$

Im ungestrichenen System

$$a_\sigma = \sum \pi_{\mu\sigma} a'_\mu = \sum \pi_{\mu\sigma} \underbrace{\pi_{\nu\tau} \frac{\partial}{\partial x_\tau}}(p_{\mu i} p_{\nu\kappa} T_{i\kappa})$$

$$= \sum \pi_{\mu\sigma} \frac{\partial}{\partial x_\kappa}(p_{\mu i} T_{i\kappa})$$

$$= \sum \frac{\partial T_{\sigma\kappa}}{\partial x_\kappa} + \sum_{\mu i\kappa} T_{i\kappa} \pi_{\mu\sigma} \frac{\partial p_{\mu i}}{\partial x_\kappa}$$

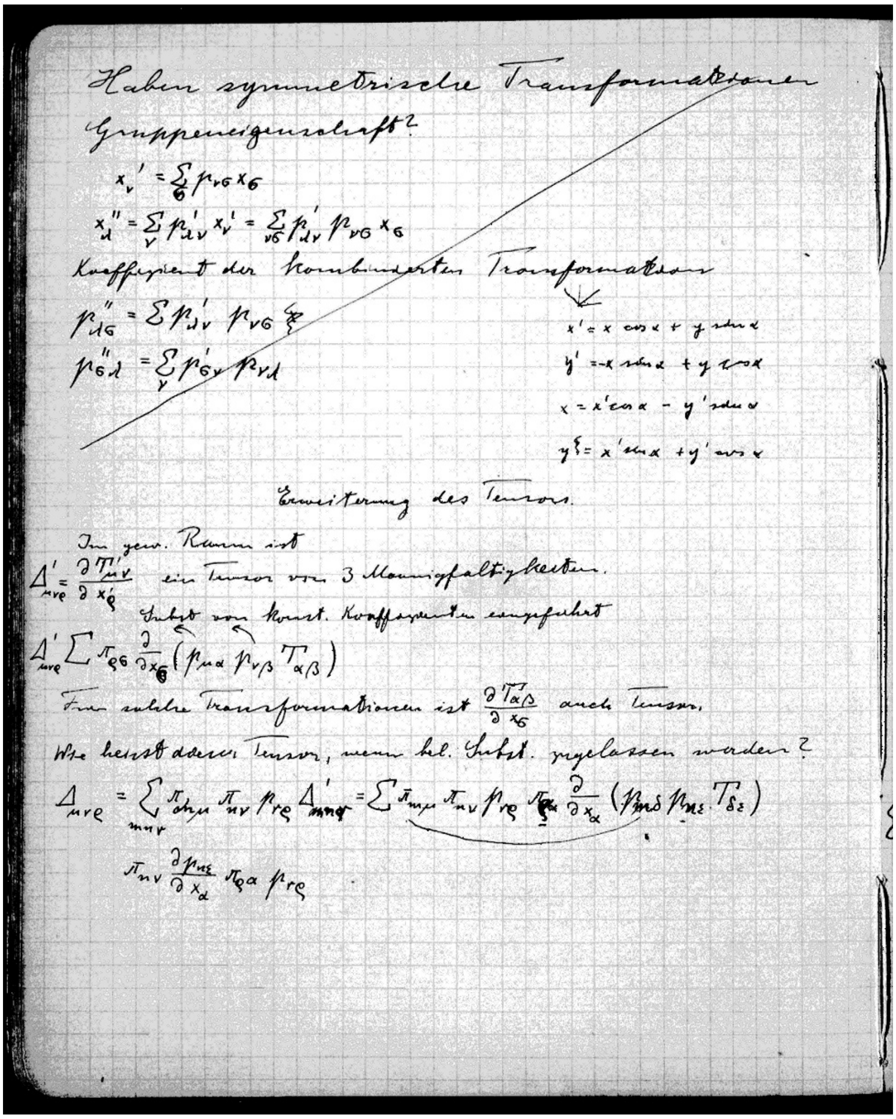
Diese Summe ist durch die g bezw. γ auszudrücken. Dabei ist zu benutzen, dass die gestrichenen g bezw. γ konstant sind.

$$\sum T_{i\kappa}\left(\frac{\partial g_{\sigma\kappa}}{\partial x_i} - \frac{1}{2} \frac{\partial g_{i\kappa}}{\partial x_\sigma}\right)$$

$$\sum \frac{\partial}{\partial x_i}(p_{\mu\sigma} p_{\nu\kappa} g'_{\mu\nu}) - \frac{1}{2} \frac{\partial}{\partial x_\sigma}(p_{\mu i} p_{\nu\kappa} g'_{\mu\nu})$$

$$\frac{1}{2} \frac{\partial p_{\mu i}}{\partial x_\sigma} p_{\nu\kappa} g'_{\mu\nu}$$

[p. 08 L]



Haben symmetrische Transformationen Gruppeneigenschaft?

$x_\nu' = \sum_\sigma p_{\nu\sigma} x_\sigma$

$x_\lambda'' = \sum_\nu p'_{\lambda\nu} x_\nu' = \sum_{\nu\sigma} p'_{\lambda\nu} p_{\nu\sigma} x_\sigma$

Koeffizient der kombinierten Transformation

$p''_{\lambda\sigma} = \sum p'_{\lambda\nu} p_{\nu\sigma}$ ξ

$p''_{\sigma\lambda} = \sum_\nu p'_{\sigma\nu} p_{\nu\lambda}$

$x' = x \cos\alpha + y \sin\alpha$

$y' = -x \sin\alpha + y \cos\alpha$

$x = x' \cos\alpha - y' \sin\alpha$

$y = x' \sin\alpha + y' \cos\alpha$

Erweiterung des Tensors.

Im gew. Raum ist

$\Delta'_{\mu\nu\varrho} = \frac{\partial T'_{\mu\nu}}{\partial x'_\varrho}$ ein Tensor von 3 Mannigfaltigkeiten.

Subst. von konst. Koeffizienten eingeführt

$\Delta'_{\mu\nu\varrho} \sum \pi_{\varrho\sigma} \frac{\partial}{\partial x_\sigma}(p_{\mu\alpha} p_{\nu\beta} T_{\alpha\beta})$

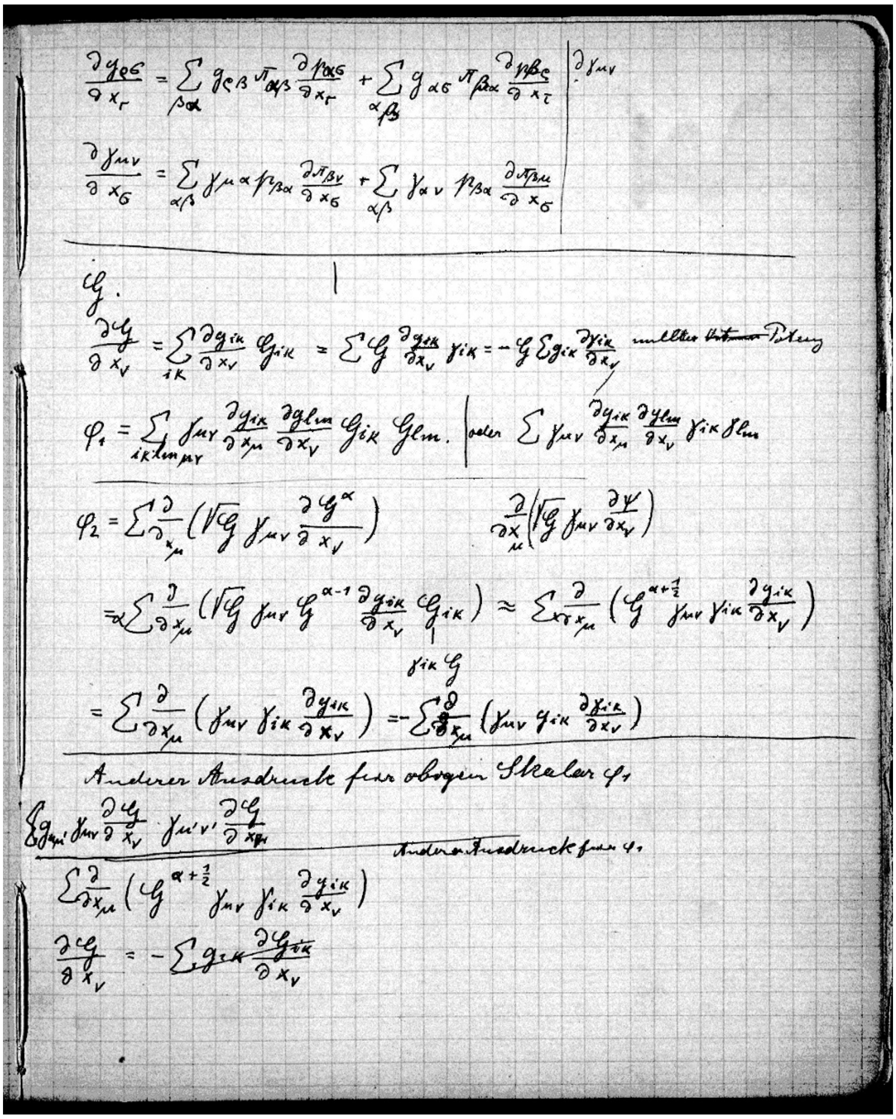
Für solche Transformationen ist $\frac{\partial T_{\alpha\beta}}{\partial x_\sigma}$ auch Tensor.

Wie heisst dieser Tensor, wenn bel. Subst. zugelassen werden?

$\Delta_{\mu\nu\varrho} = \sum_{mnr} \pi_{m\mu} \pi_{n\nu} p_{r\varrho} \Delta'_{mnr} = \sum \pi_{m\mu} \pi_{n\nu} p_{r\varrho} \pi_{\xi\alpha} \frac{\partial}{\partial x_\alpha}(p_{m\delta} p_{n\varepsilon} T_{\delta\varepsilon})$

$\pi_{n\nu} \frac{\partial p_{m\varepsilon}}{\partial x_\alpha} \pi_{\varrho\alpha} p_{r\varrho}$

[p. 08 R]



$$\frac{\partial g_{\rho\sigma}}{\partial x_\tau} = \sum_{\beta\alpha} g_{\rho\beta}\,\pi_{\alpha\beta}\frac{\partial p_{\alpha\sigma}}{\partial x_\tau} + \sum_{\alpha\beta} g_{\alpha\sigma}\,\pi_{\beta\alpha}\frac{\partial p_{\beta\rho}}{\partial x_\tau} \quad \Bigg|\ \partial\gamma_{\mu\nu}$$

$$\frac{\partial \gamma_{\mu\nu}}{\partial x_\sigma} = \sum_{\alpha\beta} \gamma_{\mu\alpha}\, p_{\beta\alpha}\frac{\partial \pi_{\beta\nu}}{\partial x_\sigma} + \sum_{\alpha\beta} \gamma_{\alpha\nu}\, p_{\beta\alpha}\frac{\partial \pi_{\beta\mu}}{\partial x_\sigma}$$

$\mathcal{G}$.

$$\frac{\partial \mathcal{G}}{\partial x_\nu} = \sum_{ik}\frac{\partial g_{ik}}{\partial x_\nu}\,\mathcal{G}_{ik} = \sum \mathcal{G}\,\frac{\partial g_{ik}}{\partial x_\nu}\gamma_{ik} = -\mathcal{G}\sum g_{ik}\frac{\partial \gamma_{ik}}{\partial x_\nu}$$ [illegible] ~~[illegible]~~ Potenz

$$\varphi_1 = \sum_{iklm\mu\nu} \gamma_{\mu\nu}\frac{\partial g_{ik}}{\partial x_\mu}\frac{\partial g_{lm}}{\partial x_\nu}\,\mathcal{G}_{ik}\,\mathcal{G}_{lm}. \quad \Big|\ \text{oder } \sum \gamma_{\mu\nu}\frac{\partial g_{ik}}{\partial x_\mu}\frac{\partial g_{lm}}{\partial x_\nu}\gamma_{ik}\gamma_{lm}$$

$$\varphi_2 = \sum \frac{\partial}{\partial x_\mu}\left(\sqrt{\mathcal{G}}\,\gamma_{\mu\nu}\frac{\partial \mathcal{G}^\alpha}{\partial x_\nu}\right) \qquad \frac{\partial}{\partial x_\mu}\left(\sqrt{\mathcal{G}}\,\gamma_{\mu\nu}\frac{\partial \psi}{\partial x_\nu}\right)$$

$$= \alpha\sum\frac{\partial}{\partial x_\mu}\left(\sqrt{\mathcal{G}}\,\gamma_{\mu\nu}\,\mathcal{G}^{\alpha-1}\frac{\partial g_{ik}}{\partial x_\nu}\,\underset{\gamma_{ik}\mathcal{G}}{\mathcal{G}_{ik}}\right) \approx \sum\frac{\partial}{\partial x_\mu}\left(\mathcal{G}^{\alpha+\frac{1}{2}}\gamma_{\mu\nu}\gamma_{ik}\frac{\partial g_{ik}}{\partial x_\nu}\right)$$

$$= \sum\frac{\partial}{\partial x_\mu}\left(\gamma_{\mu\nu}\gamma_{ik}\frac{\partial g_{ik}}{\partial x_\nu}\right) = -\sum\frac{\partial}{\partial x_\mu}\left(\gamma_{\mu\nu}\, g_{ik}\frac{\partial \gamma_{ik}}{\partial x_\nu}\right)$$

Anderer Ausdruck für obigen Skalar φ_1

$$\sum g_{\mu\mu'}\gamma_{\mu\nu}\frac{\partial \mathcal{G}}{\partial x_\nu}\,\gamma_{\mu'\nu'}\frac{\partial \mathcal{G}}{\partial x_{\nu'}}$$

Anderer Ausdruck für φ_1

$$\sum\frac{\partial}{\partial x_\mu}\left(\mathcal{G}^{\alpha+\frac{1}{2}}\gamma_{\mu\nu}\gamma_{ik}\frac{\partial g_{ik}}{\partial x_\nu}\right)$$

$$\frac{\partial \mathcal{G}}{\partial x_\nu} = -\sum g_{ik}\frac{\partial \mathcal{G}_{ik}}{\partial x_\nu}$$

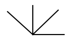
[p. 08 L]

Haben symmetrische Transformationen Gruppeneigenschaft?

$$x_\nu' = \sum_\sigma p_{\nu\sigma} x_\sigma$$

$$x_\lambda'' = \sum_\nu p'_{\lambda\nu}\, x_\nu' = \sum_{\nu\sigma} p'_{\lambda\nu}\, p_{\nu\sigma} x_\sigma$$

Koeffizient der kombinierten Transformation



$$p''_{\lambda\sigma} = \sum p'_{\lambda\nu}\, p_{\nu\sigma} \qquad x' = x\cos\alpha + y\sin\alpha$$

$$p''_{\sigma\lambda} = \sum_\nu p'_{\sigma\nu}\, p_{\nu\lambda} \qquad y' = -x\sin\alpha + y\cos\alpha$$

$$x = x'\cos\alpha - y'\sin\alpha$$

$$y = x'\sin\alpha + y'\cos\alpha$$

Erweiterung des Tensors.

Im gew. Raum ist

$$\Delta'_{\mu\nu\rho} = \frac{\partial T'_{\mu\nu}}{\partial x_\rho'} \qquad \text{Ein Tensor von 3 Mannigfaltigkeiten.}$$

Subst von konst. Koeffizienten eingeführt

$$\Delta'_{\mu\nu\rho} \sum \pi_{\rho\sigma} \frac{\partial}{\partial x_\sigma}(p_{\mu\alpha} p_{\nu\beta} T_{\alpha\beta})$$

Für solche Transformationen ist $\frac{\partial T_{\alpha\beta}}{\partial x_\sigma}$ auch Tensor.

Wie heisst dieser Tensor, wenn bel. Subst. zugelassen werden?

$$\Delta_{\mu\nu\rho} = \sum_{mnr} \pi_{m\mu}\pi_{n\nu}p_{r\rho}\Delta'_{mnr} = \sum \underbrace{\pi_{m\mu}\pi_{n\nu}p_{r\rho}\pi_{\underline{r}\alpha}\frac{\partial}{\partial x_\alpha}}(p_{m\delta}p_{n\varepsilon}T_{\delta\varepsilon})$$

$$\pi_{n\nu}\frac{\partial p_{n\varepsilon}}{\partial x_\alpha}\pi_{\rho\alpha}p_{r\rho}$$

[p. 08 R]

$$\frac{\partial g_{\rho\sigma}}{\partial x_\tau} = \sum_{\beta\alpha} g_{\rho\beta}\pi_{\alpha\beta}\frac{\partial p_{\alpha\sigma}}{\partial x_\tau} + \sum_{\alpha\beta} g_{\alpha\sigma}\pi_{\beta\alpha}\frac{\partial p_{\beta\rho}}{\partial x_\tau} \qquad \partial\gamma_{\mu\nu}$$

$$\frac{\partial \gamma_{\mu\nu}}{\partial x_\sigma} = \sum_{\alpha\beta} \gamma_{\mu\alpha}p_{\beta\alpha}\frac{\partial \pi_{\beta\nu}}{\partial x_\sigma} + \sum_{\alpha\beta} \gamma_{\alpha\nu}p_{\beta\alpha}\frac{\partial \pi_{\beta\mu}}{\partial x_\sigma}$$

$G.$

$$\frac{\partial G}{\partial x_\nu} = \sum_{i\kappa}\frac{\partial g_{i\kappa}}{\partial x_\nu}G_{i\kappa} = \sum G\frac{\partial g_{i\kappa}}{\partial x_\nu}\gamma_{i\kappa} = -G\sum g_{i\kappa}\frac{\partial \gamma_{i\kappa}}{\partial x_\nu}$$ nullter<Ortnun> Potenz

$$\varphi_1 = \sum_{i\kappa lm\mu\nu}\gamma_{\mu\nu}\frac{\partial g_{i\kappa}}{\partial x_\mu}\frac{\partial g_{lm}}{\partial x_\nu}G_{i\kappa}G_{lm.} \quad \text{oder} \quad \sum\gamma_{\mu\nu}\frac{\partial g_{i\kappa}}{\partial x_\mu}\frac{\partial g_{lm}}{\partial x_\nu}\gamma_{i\kappa}\gamma_{lm}$$

$$\varphi_2 = \sum\frac{\partial}{\partial x_\mu}\left(\sqrt{G}\gamma_{\mu\nu}\frac{\partial G^\alpha}{\partial x_\nu}\right) \qquad \frac{\partial}{\partial x_\mu}\left(\sqrt{G}\gamma_{\mu\nu}\frac{\partial \psi}{\partial x_\nu}\right)$$

$$= \alpha\sum\frac{\partial}{\partial x_\mu}\left(\sqrt{G}\gamma_{\mu\nu}G^{\alpha-1}\frac{\partial g_{i\kappa}}{\partial x_\nu}\underbrace{G_{i\kappa}}_{\gamma_{i\kappa}G}\right) \approx \sum\frac{\partial}{\partial x_\mu}\left(G^{\alpha+\frac{1}{2}}\gamma_{\mu\nu}\gamma_{i\kappa}\frac{\partial g_{i\kappa}}{\partial x_\nu}\right)$$

$$= \sum\frac{\partial}{\partial x_\mu}\left(\gamma_{\mu\nu}\gamma_{i\kappa}\frac{\partial g_{i\kappa}}{\partial x_\nu}\right) = -\sum\frac{\partial}{\partial x_\mu}\left(\gamma_{\mu\nu}g_{i\kappa}\frac{\partial \gamma_{i\kappa}}{\partial x_\nu}\right)$$

Anderer Ausdruck für obigen Skalar φ_1

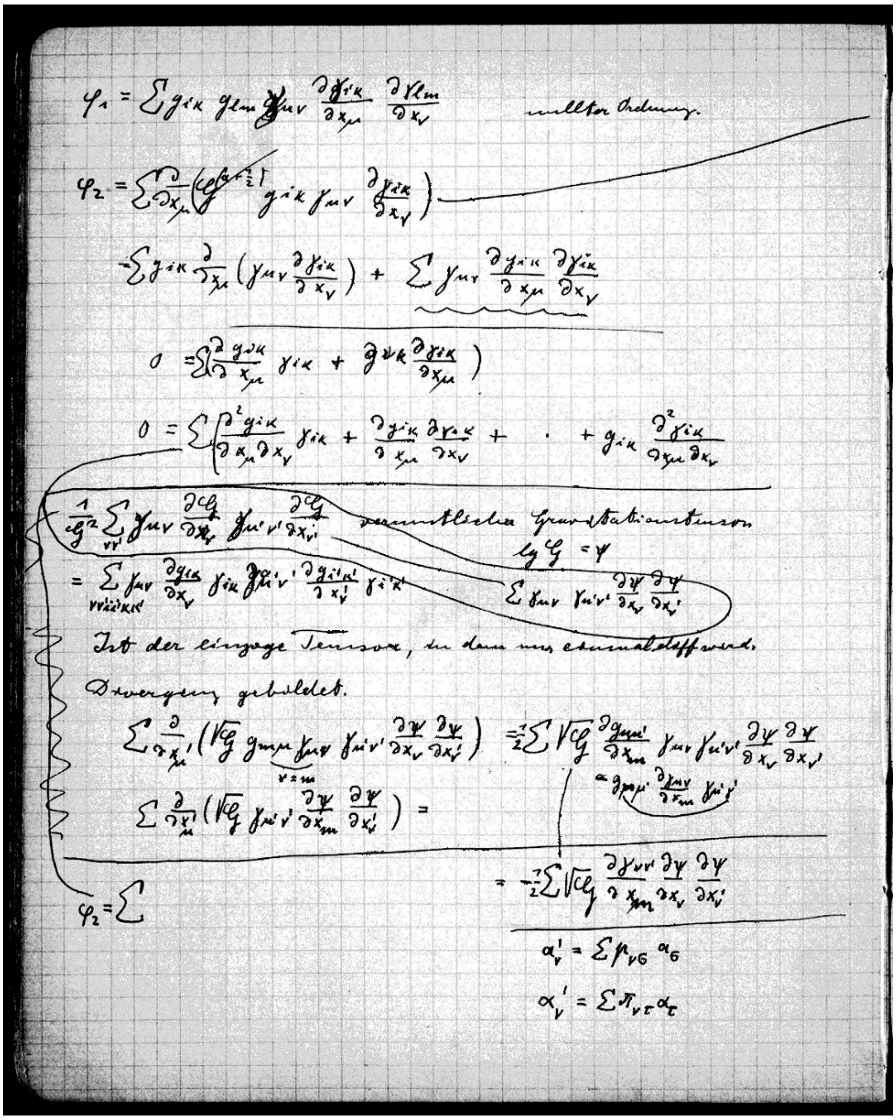
$$\sum g_{\mu\mu'}\gamma_{\mu\nu}\frac{\partial G}{\partial x_\nu}\ \gamma_{\mu'\nu'}\frac{\partial G}{\partial x_{\nu'}}$$

Anderer Ausdruck für φ_1

$$\sum\frac{\partial}{\partial x_\mu}\left(G^{\alpha+\frac{1}{2}}\gamma_{\mu\nu}\gamma_{i\kappa}\frac{\partial g_{i\kappa}}{\partial x_\nu}\right)$$

$$\frac{\partial G}{\partial x_\nu} = -\cancel{\sum g_{i\kappa}\frac{\partial G_{i\kappa}}{\partial x_\nu}}$$

[p. 09 L]



$$\varphi_1 = \sum g_{ik}\, g_{lm}\, \gamma_{\mu\nu} \frac{\partial \gamma_{ik}}{\partial x_\mu} \frac{\partial \gamma_{lm}}{\partial x_\nu}$$ nullter Ordnung.

$$\varphi_2 = \sum \frac{\partial}{\partial x_\mu}\left(\cancel{G^{\alpha+\frac{1}{2}}}\, g_{ik}\, \gamma_{\mu\nu} \frac{\partial \gamma_{ik}}{\partial x_\nu}\right)$$

$$= \sum g_{ik} \frac{\partial}{\partial x_\mu}\left(\gamma_{\mu\nu} \frac{\partial \gamma_{ik}}{\partial x_\nu}\right) + \sum \gamma_{\mu\nu} \frac{\partial g_{ik}}{\partial x_\mu} \frac{\partial \gamma_{ik}}{\partial x_\nu}$$

$$0 = \sum\left(\frac{\partial g_{\nu k}}{\partial x_\mu} \gamma_{ik} + g_{\nu k} \frac{\partial \gamma_{ik}}{\partial x_\mu}\right)$$

$$0 = \sum\left(\frac{\partial^2 g_{ik}}{\partial x_\mu \partial x_\nu} \gamma_{ik} + \frac{\partial g_{ik}}{\partial x_\mu} \frac{\partial \gamma_{ik}}{\partial x_\nu} + \cdot + g_{ik} \frac{\partial^2 \gamma_{ik}}{\partial x_\mu \partial x_\nu}\right.$$

$$\frac{1}{G^2} \sum_{\nu\nu'} \gamma_{\mu\nu} \frac{\partial G}{\partial x_\nu} \gamma_{\mu'\nu'} \frac{\partial G}{\partial x_{\nu'}}$$ sämtlicher Gravitationstensoren

$$\lg G = \psi$$

$$= \sum_{\nu\nu' i i' k k'} \gamma_{\mu\nu} \frac{\partial g_{ik}}{\partial x_\nu} \gamma_{ik}\, \gamma_{\mu'\nu'} \frac{\partial g_{i'k'}}{\partial x'_\nu} \gamma_{i'k'} \qquad \sum \gamma_{\mu\nu} \gamma_{\mu'\nu'} \frac{\partial \psi}{\partial x_\nu} \frac{\partial \psi}{\partial x'_\nu}$$

Ist der einzige Tensor, in dem nur einmal differenz. Divergenz gebildet.

$$\sum \frac{\partial}{\partial x'_\mu}\left(\sqrt{G}\, g_{m\mu}\, \underbrace{\gamma_{\mu\nu}}_{\nu=m}\, \gamma_{\mu'\nu'} \frac{\partial \psi}{\partial x_\nu} \frac{\partial \psi}{\partial x'_\nu}\right) = \frac{1}{2} \sum \sqrt{G} \frac{\partial g_{\mu\mu'}}{\partial x_m} \gamma_{\mu\nu} \gamma_{\mu'\nu'} \frac{\partial \psi}{\partial x_\nu} \frac{\partial \psi}{\partial x_{\nu'}}$$

$$- g_{\mu\mu'} \frac{\partial \gamma_{\mu\nu}}{\partial x_m} \gamma_{\mu'\nu'}$$

$$\sum \frac{\partial}{\partial x'_\mu}\left(\sqrt{G}\, \gamma_{\mu'\nu'} \frac{\partial \psi}{\partial x_m} \frac{\partial \psi}{\partial x'_\nu}\right) =$$

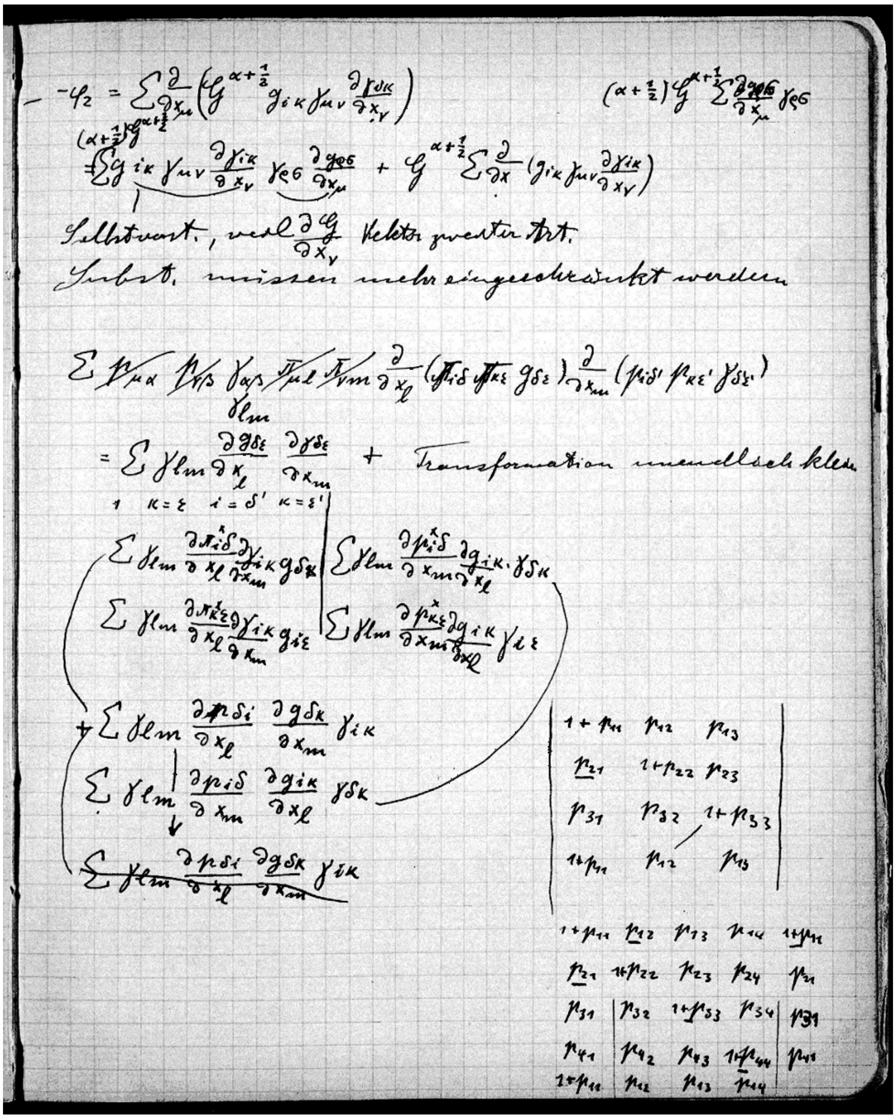
$$= -\frac{1}{2} \sum \sqrt{G} \frac{\partial \gamma_{\nu\nu'}}{\partial x_m} \frac{\partial \psi}{\partial x_\nu} \frac{\partial \psi}{\partial x'_\nu}$$

$$\varphi_2 = \sum$$

$$\alpha'_\nu = \sum p_{\nu\sigma} \alpha_\sigma$$

$$\alpha'_\nu = \sum \pi_{\nu\tau} \alpha_\tau$$

[p. 09 R]


$$-\varphi_2 = \sum \frac{\partial}{\partial x_\mu}\left(G^{\alpha+\frac{1}{2}} g_{ik}\gamma_{\mu\nu}\frac{\partial \gamma_{ik}}{\partial x_\nu}\right) \qquad (\alpha+\tfrac{1}{2})G^{\alpha+\frac{1}{2}}\sum\frac{\partial g_{\varrho\sigma}}{\partial x_\mu}\gamma_{\varrho\sigma}$$

$$= (\alpha+\tfrac{1}{2})G^{\alpha+\frac{1}{2}}\sum g_{ik}\gamma_{\mu\nu}\frac{\partial\gamma_{ik}}{\partial x_\nu}\gamma_{\varrho\sigma}\frac{\partial g_{\varrho\sigma}}{\partial x_\mu} + G^{\alpha+\frac{1}{2}}\sum\frac{\partial}{\partial x}\left(g_{ik}\gamma_{\mu\nu}\frac{\partial\gamma_{ik}}{\partial x_\nu}\right)$$

Selbstverst., weil $\frac{\partial G}{\partial x_\nu}$ Vektor zweiter Art.

Subst. müssen mehr eingeschränkt werden

$$\sum p_{\mu\alpha}\,p_{\nu\beta}\,\gamma_{\alpha\beta}\,\pi_{\mu l}\,\pi_{\nu m}\frac{\partial}{\partial x_l}(\pi_{i\delta}\pi_{\kappa\varepsilon}g_{\delta\varepsilon})\frac{\partial}{\partial x_m}(p_{i\delta'}p_{\kappa\varepsilon'}\gamma_{\delta'\varepsilon'})$$

γ_{lm}

$$= \sum \gamma_{lm}\frac{\partial g_{\delta\varepsilon}}{\partial x_l}\frac{\partial\gamma_{\delta\varepsilon}}{\partial x_m} + \text{Transformation unendlich klein}$$

$1 \quad \kappa=\varepsilon \quad i=\delta' \quad \kappa=\varepsilon'$

$$\sum\gamma_{lm}\frac{\partial\pi_{i\delta}}{\partial x_l}\frac{\partial\gamma_{ik}}{\partial x_m}g_{\delta\kappa} \qquad \sum\gamma_{lm}\frac{\partial p_{i\delta}}{\partial x_m}\frac{\partial g_{ik}}{\partial x_l}\gamma_{\delta\kappa}$$

$$\sum\gamma_{lm}\frac{\partial\pi_{\kappa\varepsilon}}{\partial x_l}\frac{\partial\gamma_{ik}}{\partial x_m}g_{i\varepsilon} \qquad \sum\gamma_{lm}\frac{\partial p_{\kappa\varepsilon}}{\partial x_m}\frac{\partial g_{ik}}{\partial x_l}\gamma_{i\varepsilon}$$

$$+\sum\gamma_{lm}\frac{\partial p_{\delta i}}{\partial x_l}\frac{\partial g_{\delta\kappa}}{\partial x_m}\gamma_{ik}$$

$$\sum\gamma_{lm}\frac{\partial p_{i\delta}}{\partial x_m}\frac{\partial g_{ik}}{\partial x_l}\gamma_{\delta\kappa}$$

$$\cancel{\sum\gamma_{lm}\frac{\partial p_{\delta i}}{\partial x_l}\frac{\partial g_{\delta\kappa}}{\partial x_m}\gamma_{ik}}$$

$$\begin{vmatrix} 1+p_{11} & p_{12} & p_{13} \\ p_{21} & 1+p_{22} & p_{23} \\ p_{31} & p_{32} & 1+p_{33} \\ 1+p_{11} & p_{12} & p_{13} \end{vmatrix}$$

$$\begin{matrix} 1+p_{11} & p_{12} & p_{13} & p_{14} & 1+p_{11} \\ p_{21} & 1+p_{22} & p_{23} & p_{24} & p_{21} \\ p_{31} & p_{32} & 1+p_{33} & p_{34} & p_{31} \\ p_{41} & p_{42} & p_{43} & 1+p_{44} & p_{41} \\ 1+p_{11} & p_{12} & p_{13} & p_{14} & \end{matrix}$$

[p. 09 L]

$$\varphi_1 = \sum g_{i\kappa} g_{lm} \gamma_{\mu\nu} \frac{\partial \gamma_{i\kappa}}{\partial x_\mu} \frac{\partial \gamma_{lm}}{\partial x_\nu}$$ nullter Ordnung.

$$\varphi_2 = \sum \frac{\partial}{\partial x_\mu}\left(G^{\cancel{\left(\alpha+\frac{1}{2}\right)}} g_{i\kappa} \gamma_{\mu\nu} \frac{\partial \gamma_{i\kappa}}{\partial x_\nu}\right)$$

[connects p. 09 R]

$$= \sum g_{i\kappa} \frac{\partial}{\partial x_\mu}\left(\gamma_{\mu\nu} \frac{\partial \gamma_{i\kappa}}{\partial x_\nu}\right) + \sum \gamma_{\mu\nu} \frac{\partial g_{i\kappa}}{\partial x_\mu} \frac{\partial \gamma_{i\kappa}}{\partial x_\nu}$$

$$0 = \sum\left(\frac{\partial g_{i\kappa}}{\partial x_\mu}\gamma_{i\kappa} + g_{i\kappa}\frac{\partial \gamma_{i\kappa}}{\partial x_\mu}\right)$$

$$0 = \sum\left(\frac{\partial^2 g_{i\kappa}}{\partial x_\mu \partial x_\nu}\gamma_{i\kappa} + \frac{\partial g_{i\kappa}}{\partial x_\mu}\frac{\partial \gamma_{i\kappa}}{\partial x_\nu} + \quad \cdot \quad + g_{i\kappa}\frac{\partial^2 \gamma_{i\kappa}}{\partial x_\mu \partial x_\nu}\right)$$

$$\frac{1}{G^2}\sum_{\nu\nu'} \gamma_{\mu\nu} \frac{\partial G}{\partial x_\nu} \gamma_{\mu'\nu'} \frac{\partial G}{\partial x_{\nu'}}$$ vermutlicher Gravitationstensor

$$\lg G = \psi$$

$$\sum \gamma_{\mu\nu}\gamma_{\mu'\nu'}\frac{\partial \psi}{\partial x_\nu}\frac{\partial \psi}{\partial x_{\nu'}}$$

$$= \sum_{\nu\nu' ii' \kappa\kappa'} \gamma_{\mu\nu} \frac{\partial g_{i\kappa}}{\partial x_\nu} \gamma_{i\kappa} \gamma_{\mu'\nu'} \frac{\partial g_{i'\kappa'}}{\partial x_{\nu'}} \gamma_{i'\kappa'}$$

Ist der einzige Tensor, in dem nur einmal diff. wird.

Divergenz gebildet.

$$\sum \frac{\partial}{\partial x_{\mu'}}\left(\sqrt{G} g_{m\mu}\gamma_{\mu\nu}\gamma_{\mu'\nu'}\frac{\partial \psi}{\partial x_\nu}\frac{\partial \psi}{\partial x_{\nu'}}\right) = \frac{1}{2}\sqrt{G}\frac{\partial g_{\mu\mu'}}{\partial x_m}\gamma_{\mu\nu}\gamma_{\mu'\nu'}\frac{\partial \psi}{\partial x_\nu}\frac{\partial \psi}{\partial x_{\nu'}}$$

$$\nu = m$$

$$\sum \frac{\partial}{\partial x_{\mu'}}\left(\sqrt{G}\gamma_{\mu'\nu'}\frac{\partial \psi}{\partial x_m}\frac{\partial \psi}{\partial x_{\nu'}}\right) =$$

$$-g_{\mu\mu'}\frac{\partial \gamma_{\mu\nu}}{\partial x_m}\gamma_{\mu'\nu'}$$

$$\varphi_2 = \sum$$

$$= -\frac{1}{2}\sum \sqrt{G}\frac{\partial \gamma_{\nu\nu'}}{\partial x_m}\frac{\partial \psi}{\partial x_\nu}\frac{\partial \psi}{\partial x_{\nu'}}$$

$$a'_\nu = \sum p_{\nu\sigma} a_\sigma$$

$$\alpha'_\nu = \sum \pi_{\nu\tau}\alpha_\tau$$

[p. 09 R]

$$-\varphi_2 = \sum \frac{\partial}{\partial x_\mu}\left(G^{\alpha+\frac{1}{2}} g_{i\kappa}\gamma_{\mu\nu}\frac{\partial\gamma_{i\kappa}}{\partial x_\nu}\right) \qquad \left(\alpha+\frac{1}{2}\right)G^{\alpha+\frac{1}{2}}\sum\frac{\partial g_{\rho\sigma}}{\partial x_\mu}\gamma_{\rho\sigma}$$

[connects p. 09 L]

$$\left(\alpha+\frac{1}{2}\right)G^{\alpha+\frac{1}{2}}$$

$$= \sum g_{i\kappa}\gamma_{\mu\nu}\frac{\partial\gamma_{i\kappa}}{\partial x_\nu}\gamma_{\rho\sigma}\frac{\partial g_{\rho\sigma}}{\partial x_\mu} + G^{\alpha+\frac{1}{2}}\sum\frac{\partial}{\partial x}\left(g_{i\kappa}\gamma_{\mu\nu}\frac{\partial\gamma_{i\kappa}}{\partial x_\nu}\right)$$

Selbstverst., weil $\frac{\partial G}{\partial x_\nu}$ Vektor zweiter Art.
Subst. müssen mehr eingeschränkt werden

$$\sum \cancel{p_{\mu\alpha}}\cancel{p_{\nu\beta}}\underset{\gamma_{lm}}{\gamma_{\alpha\beta}}\cancel{\pi_{\mu l}}\cancel{\pi_{\nu m}}\frac{\partial}{\partial x_l}(\pi_{i\delta}\pi_{\kappa\varepsilon}g_{\delta\varepsilon})\frac{\partial}{\partial x_m}(p_{i\delta'}p_{\kappa\varepsilon'}\gamma_{\delta'\varepsilon'})$$

$$= \sum\gamma_{lm}\frac{\partial g_{\delta\varepsilon}}{\partial x_l}\frac{\partial\gamma_{\delta\varepsilon}}{\partial x_m} \quad + \quad \text{Transformation unendlich klein}$$

$$1 \quad \kappa=\varepsilon \quad i=\delta' \quad \kappa=\varepsilon'$$

$$\begin{array}{l|l}
\sum\gamma_{lm}\frac{\partial\pi^{x}_{i\delta}}{\partial x_l}\frac{\partial\gamma_{i\kappa}}{\partial x_m}g_{\delta\kappa} & \sum\gamma_{lm}\frac{\partial p^{x}_{i\delta}}{\partial x_m}\frac{\partial g_{i\kappa}}{\partial x_l}\cdot\gamma_{\delta\kappa} \\
\sum\gamma_{lm}\frac{\partial\pi^{x}_{\kappa\varepsilon}}{\partial x_l}\frac{\partial\gamma_{i\kappa}}{\partial x_m}g_{i\varepsilon} & \sum\gamma_{lm}\frac{\partial p^{x}_{\kappa\varepsilon}}{\partial x_m}\frac{\partial g_{i\kappa}}{\partial x_l}\gamma_{i\varepsilon}
\end{array}$$

$$+\sum\gamma_{lm}\frac{\partial p_{\delta i}}{\partial x_l}\frac{\partial g_{\delta\kappa}}{\partial x_m}\gamma_{i\kappa}$$

$$\sum\gamma_{lm}\frac{\partial p_{i\delta}}{\partial x_m}\frac{\partial g_{i\kappa}}{\partial x_l}\gamma_{\delta\kappa}$$

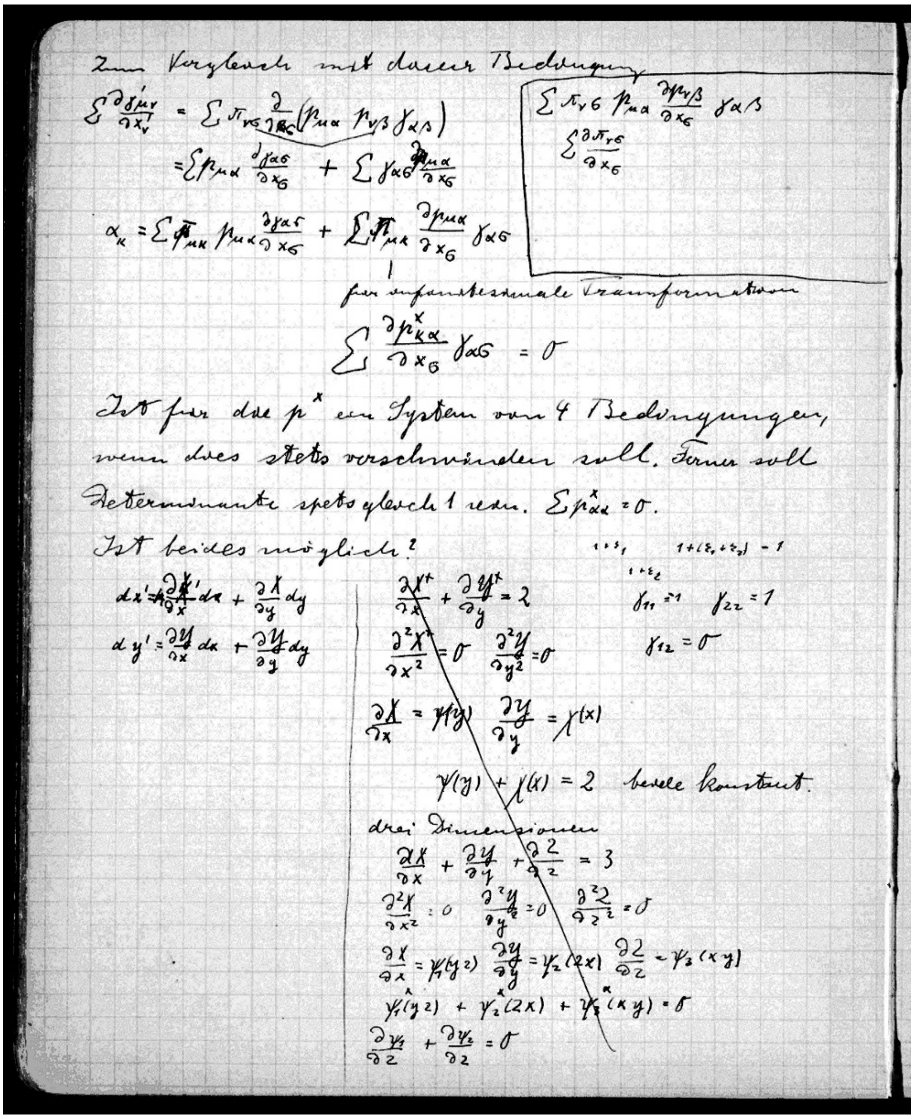
$$\downarrow$$

$$\cancel{\sum\gamma_{lm}\frac{\partial p_{\delta i}}{\partial x_l}\frac{\partial g_{\delta\kappa}}{\partial x_m}}\gamma_{i\kappa}$$

$$\left|\begin{array}{ccc}
1+p_{11} & p_{12} & p_{13} \\
\underline{p}_{21} & 1+p_{22} & p_{23} \\
p_{31} & p_{32} & 1+p_{33} \\
1+p_{11} & p_{12} & p_{13}
\end{array}\right|$$

$$\begin{array}{c|cccc|c}
1+p_{11} & \underline{p}_{12} & p_{13} & p_{14} & & 1+\underline{p}_{11} \\
\underline{p}_{21} & 1+p_{22} & p_{23} & p_{24} & & p_{21} \\
p_{31} & p_{32} & 1+\underline{p}_{33} & p_{34} & & p_{31} \\
p_{41} & p_{42} & p_{43} & 1+\underline{p}_{44} & & p_{41} \\
1+p_{11} & p_{12} & p_{13} & p_{14} & &
\end{array}$$

[p. 10 L]



Zum Vergleich mit dieser Bedingung

$$\sum \frac{\partial \gamma'_{\mu\nu}}{\partial x'_\nu} = \sum \pi_{\nu\sigma} \frac{\partial}{\partial x_\sigma}(p_{\mu\alpha} p_{\nu\beta} \gamma_{\alpha\beta})$$

$$= \sum p_{\mu\alpha} \frac{\partial \gamma_{\alpha\sigma}}{\partial x_\sigma} + \sum \gamma_{\alpha\sigma} \frac{\partial p_{\mu\alpha}}{\partial x_\sigma}$$

$$\boxed{\sum \pi_{\nu\sigma} p_{\mu\alpha} \frac{\partial p_{\nu\beta}}{\partial x_\sigma} \gamma_{\alpha\beta} \qquad \sum \frac{\partial \pi_{\nu\sigma}}{\partial x_\sigma}}$$

$$\alpha_\kappa = \sum \pi_{\mu\kappa} p_{\mu\alpha} \frac{\partial \gamma_{\alpha\sigma}}{\partial x_\sigma} + \sum \pi_{\mu\kappa} \frac{\partial p_{\mu\alpha}}{\partial x_\sigma} \gamma_{\alpha\sigma}$$

für infinitesimale Transformation

$$\sum \frac{\partial p^x_{\kappa\alpha}}{\partial x_\sigma} \gamma_{\alpha\sigma} = 0$$

Ist für die p^x ein System von 4 Bedingungen, wenn dies stets verschwinden soll. Ferner soll Determinante stets gleich 1 sein. $\sum p^x_{\alpha\alpha} = 0$.

Ist beides möglich? $1+\varepsilon_1$ $1+(\varepsilon_1+\varepsilon_2)-1$ $1+\varepsilon_2$

$$dx' = \frac{\partial X'}{\partial x} dx + \frac{\partial X}{\partial y} dy \qquad \frac{\partial X^x}{\partial x} + \frac{\partial Y^x}{\partial y} = 2 \qquad \gamma_{11} = 1 \quad \gamma_{22} = 1$$

$$dy' = \frac{\partial Y}{\partial x} dx + \frac{\partial Y}{\partial y} dy \qquad \frac{\partial^2 X^x}{\partial x^2} = 0 \quad \frac{\partial^2 Y}{\partial y^2} = 0 \qquad \gamma_{12} = 0$$

$$\frac{\partial X}{\partial x} = \psi(y) \quad \frac{\partial Y}{\partial y} = \chi(x)$$

$\psi(y) + \chi(x) = 2$ beide konstant.

drei Dimensionen

$$\frac{\partial X}{\partial x} + \frac{\partial Y}{\partial y} + \frac{\partial Z}{\partial z} = 3$$

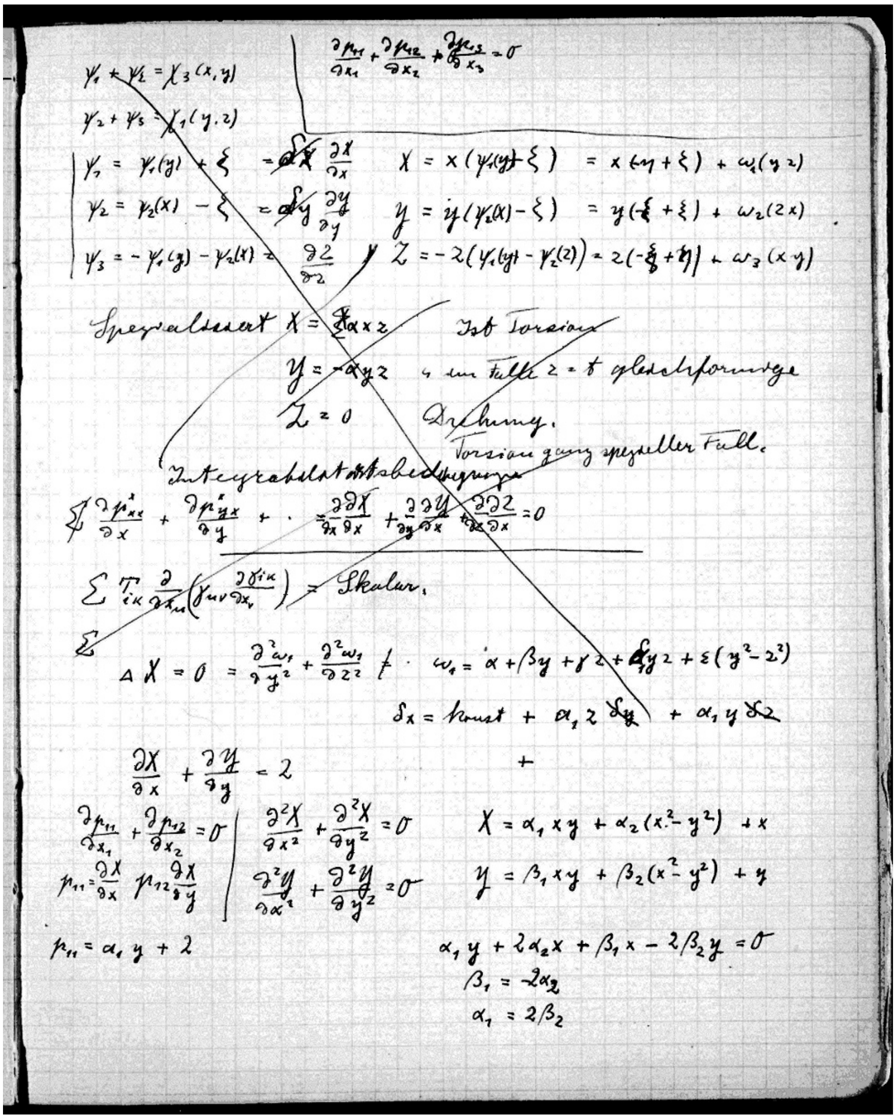
$$\frac{\partial^2 X}{\partial x^2} = 0 \quad \frac{\partial^2 Y}{\partial y^2} = 0 \quad \frac{\partial^2 Z}{\partial z^2} = 0$$

$$\frac{\partial X}{\partial x} = \psi_1(yz) \quad \frac{\partial Y}{\partial y} = \psi_2(zx) \quad \frac{\partial Z}{\partial z} = \psi_3(xy)$$

$$\psi_1^x(yz) + \psi_2^x(zx) + \psi_3^x(xy) = 0$$

$$\frac{\partial \psi_1}{\partial z} + \frac{\partial \psi_2}{\partial z} = 0$$

[p. 10 R]



$$\frac{\partial p_{11}}{\partial x_1}+\frac{\partial p_{12}}{\partial x_2}+\frac{\partial p_{13}}{\partial x_3}=0$$

$$\psi_1+\psi_2=\chi_3(x,y)$$

$$\psi_2+\psi_3=\chi_1(y,z)$$

$$\psi_1=\psi_1(y)+\xi \qquad \frac{\partial X}{\partial x} \qquad X=x(\psi_1(y)+\xi)=x(-\eta+\xi)+\omega_1(yz)$$

$$\psi_2=\psi_2(x)-\xi \qquad \frac{\partial Y}{\partial y} \qquad Y=y(\psi_2(x)-\xi)=y(-\xi+\xi)+\omega_2(zx)$$

$$\psi_3=-\psi_1(y)-\psi_2(x)=\frac{\partial Z}{\partial z} \qquad Z=-z(\psi_1(y)-\psi_2(z))=z(-\xi+\eta)+\omega_3(xy)$$

Spezialfall $X=\alpha xz$ Ist Torsion

$Y=-\alpha yz$ u. im Falle $z=t$ gleichförmige

$Z=0$ Drehung.

Torsion ganz spezieller Fall.

Integrabilitätsbedingungen

$$\sum \frac{\partial p_{xx}}{\partial x}+\frac{\partial p_{yx}}{\partial y}+\cdot\ \frac{\partial}{\partial x}\frac{\partial X}{\partial x}+\frac{\partial}{\partial y}\frac{\partial Y}{\partial x}+\frac{\partial}{\partial z}\frac{\partial Z}{\partial x}=0$$

$$\sum T_{ik}\frac{\partial}{\partial x_\mu}\left(\gamma_{\mu\nu}\frac{\partial \gamma_{ik}}{\partial x_\nu}\right)=\text{Skalar.}$$

$$\sum$$

$$\Delta X=0=\frac{\partial^2\omega_1}{\partial y^2}+\frac{\partial^2\omega_1}{\partial z^2} \qquad \omega_1=\alpha+\beta y+\gamma z+\delta yz+\varepsilon(y^2-z^2)$$

$$\delta_x=\text{konst}+\alpha_1 z\,\delta y+\alpha_1 y\,\delta z$$

$$+$$

$$\frac{\partial X}{\partial x}+\frac{\partial Y}{\partial y}=2$$

$$\frac{\partial p_{11}}{\partial x_1}+\frac{\partial p_{12}}{\partial x_2}=0 \qquad \frac{\partial^2 X}{\partial x^2}+\frac{\partial^2 X}{\partial y^2}=0 \qquad X=\alpha_1 xy+\alpha_2(x^2-y^2)+x$$

$$p_{11}=\frac{\partial X}{\partial x} \quad p_{12}\frac{\partial X}{\partial y} \qquad \frac{\partial^2 Y}{\partial x^2}+\frac{\partial^2 Y}{\partial y^2}=0 \qquad Y=\beta_1 xy+\beta_2(x^2-y^2)+y$$

$$p_{11}=\alpha_1 y+2$$

$$\alpha_1 y+2\alpha_2 x+\beta_1 x-2\beta_2 y=0$$

$$\beta_1=-2\alpha_2$$

$$\alpha_1=2\beta_2$$

[p. 10 L]

Zum Vergleich mit dieser Bedingung

$$\sum \frac{\partial \gamma'_{\mu\nu}}{\partial x'_\nu} = \sum \pi_{\nu\sigma} \frac{\partial}{\partial x_\sigma}(p_{\mu\alpha} p_{\nu\beta} \gamma_{\alpha\beta})$$

$$= \sum p_{\mu\alpha} \frac{\partial \gamma_{\alpha\sigma}}{\partial x_\sigma} + \sum \gamma_{\alpha\sigma} \frac{\partial p_{\mu\alpha}}{\partial x_\sigma}$$

$$\alpha_\kappa = \sum \pi_{\mu\kappa} p_{\mu\alpha} \frac{\partial \gamma_{\alpha\sigma}}{\partial x_\sigma} + \sum \pi_{\mu\kappa} \frac{\partial p_{\mu\alpha}}{\partial x_\sigma} \gamma_{\alpha\sigma}$$

$$\boxed{\begin{array}{l} \sum \pi_{\nu\sigma} p_{\mu\alpha} \frac{\partial p_{\nu\beta}}{\partial x_\sigma} \gamma_{\alpha\beta} \\ \sum \frac{\partial \pi_{\nu\sigma}}{\partial x_\sigma} \end{array}}$$

für infinitesimale Transformation

$$\sum \frac{\partial p^x_{\kappa\alpha}}{\partial x_\sigma} \gamma_{\alpha\sigma} = 0$$

Ist für die p^x ein System von 4 Bedingungen, wenn dies stets verschwinden soll. Ferner soll Determinante stets gleich 1 sein. $\sum p^x_{\alpha\alpha} = 0$.

Ist beides möglich?

$$1 + \varepsilon_1 \qquad 1 + (\varepsilon_1 + \varepsilon_2) = 1$$

$$1 + \varepsilon_2$$

$$\gamma_{11} = -1 \quad \gamma_{22} = 1$$

$$\gamma_{12} = 0$$

$$dx' = \frac{\partial X'}{\partial x} dx + \frac{\partial X}{\partial y} dy$$

$$dy' = \frac{\partial Y}{\partial x} dx + \frac{\partial Y}{\partial y} dy$$

$$\frac{\partial X^x}{\partial x} + \frac{\partial Y^x}{\partial y} = 2$$

$$\frac{\partial^2 X^x}{\partial x^2} = 0 \quad \frac{\partial^2 Y}{\partial y^2} = 0$$

$$\frac{\partial X}{\partial x} = \psi(y) \quad \frac{\partial Y}{\partial y} = \chi(x)$$

$$\psi(y) + \chi(x) = 2 \quad \text{beide konstant.}$$

drei Dimensionen

$$\frac{\partial X}{\partial x} + \frac{\partial Y}{\partial y} + \frac{\partial Z}{\partial z} = 3$$

$$\frac{\partial^2 X}{\partial x^2} = 0 \quad \frac{\partial^2 Y}{\partial y^2} = 0 \quad \frac{\partial^2 Z}{\partial z^2} = 0$$

$$\frac{\partial X}{\partial x} = \psi_1(yz) \quad \frac{\partial Y}{\partial y} = \psi_2(zx) \quad \frac{\partial Z}{\partial z} = \psi_3(xy)$$

$$\psi_1^x(yz) + \psi_2^x(zx) + \psi_3^x(xy) = 0$$

$$\frac{\partial \psi_1}{\partial z} + \frac{\partial \psi_2}{\partial z} = 0$$

[p. 10 R]

$$\psi_1 + \psi_2 = \chi_3(x, y)$$

$$\psi_2 + \psi_3 = \chi_1(y, z)$$

$$\psi_1 = \psi_1(y) + \zeta \quad = \langle \delta x \rangle \frac{\partial X}{\partial x}$$

$$\psi_2 = \psi_2(x) - \zeta \quad = \langle \delta y \rangle \frac{\partial Y}{\partial y}$$

$$\psi_3 = -\psi_1(y) - \psi_2(x) = \frac{\partial Z}{\partial z}$$

$$\frac{\partial p_{11}}{\partial x_1} + \frac{\partial p_{12}}{\partial x_2} + \frac{\partial p_{13}}{\partial x_3} = 0$$

$$X = x(\psi_1(y) + \zeta) = x(-\eta + \zeta) + \omega_1(yz)$$

$$Y = y(\psi_2(x) - \zeta) = y(-\zeta + \xi) + \omega_2(zx)$$

$$Z = -z(\psi_1(y) - \psi_2(z)) = z(-\xi + \eta) + \omega_3(xy)$$

Spezialisiert $X = \alpha xz$ Ist Torsion

$Y = -\alpha yz$ u im Falle $z = t$ gleichformige

$Z = 0$ Drehung.

Torsion ganz spezieller Fall.

Integrabilitätsbedingungen

$$\sum \frac{\partial p_{xx}^x}{\partial x} + \frac{\partial p_{yx}^x}{\partial y} + \cdot = \frac{\partial}{\partial x}\frac{\partial X}{\partial x} + \frac{\partial}{\partial y}\frac{\partial Y}{\partial x} + \frac{\partial}{\partial z}\frac{\partial Z}{\partial x} = 0$$

$$\sum T_{i\kappa} \frac{\partial}{\partial x_\mu}\left(\gamma_{\mu\nu}\frac{\partial \gamma_{i\kappa}}{\partial x_\nu}\right) = \text{Skalar.}$$

$$\sum$$

$$\Delta X = 0 = \frac{\partial^2 \omega_1}{\partial y^2} + \frac{\partial^2 \omega_1}{\partial z^2} \neq \cdot \quad \omega_1 = \alpha + \beta y + \gamma z + \langle \delta \rangle \alpha_1 yz + \varepsilon(y^2 - z^2)$$

$$\delta x = \text{konst.} + \alpha_1 z \langle \delta y \rangle + \alpha_2 y \langle \delta z \rangle$$

$$+$$

$$\frac{\partial X}{\partial x} + \frac{\partial Y}{\partial y} = 2$$

$$\frac{\partial p_{11}}{\partial x_1} + \frac{\partial p_{12}}{\partial x_2} = 0$$

$$p_{11} = \frac{\partial X}{\partial x} \quad p_{12}\frac{\partial X}{\partial y}$$

$$\frac{\partial^2 X}{\partial x^2} + \frac{\partial^2 X}{\partial y^2} = 0$$

$$\frac{\partial^2 Y}{\partial x^2} + \frac{\partial^2 Y}{\partial y^2} = 0$$

$$p_{11} = \alpha_1 y + 2$$

$$X = \alpha_1 xy + \alpha_2(x^2 - y^2) + x$$

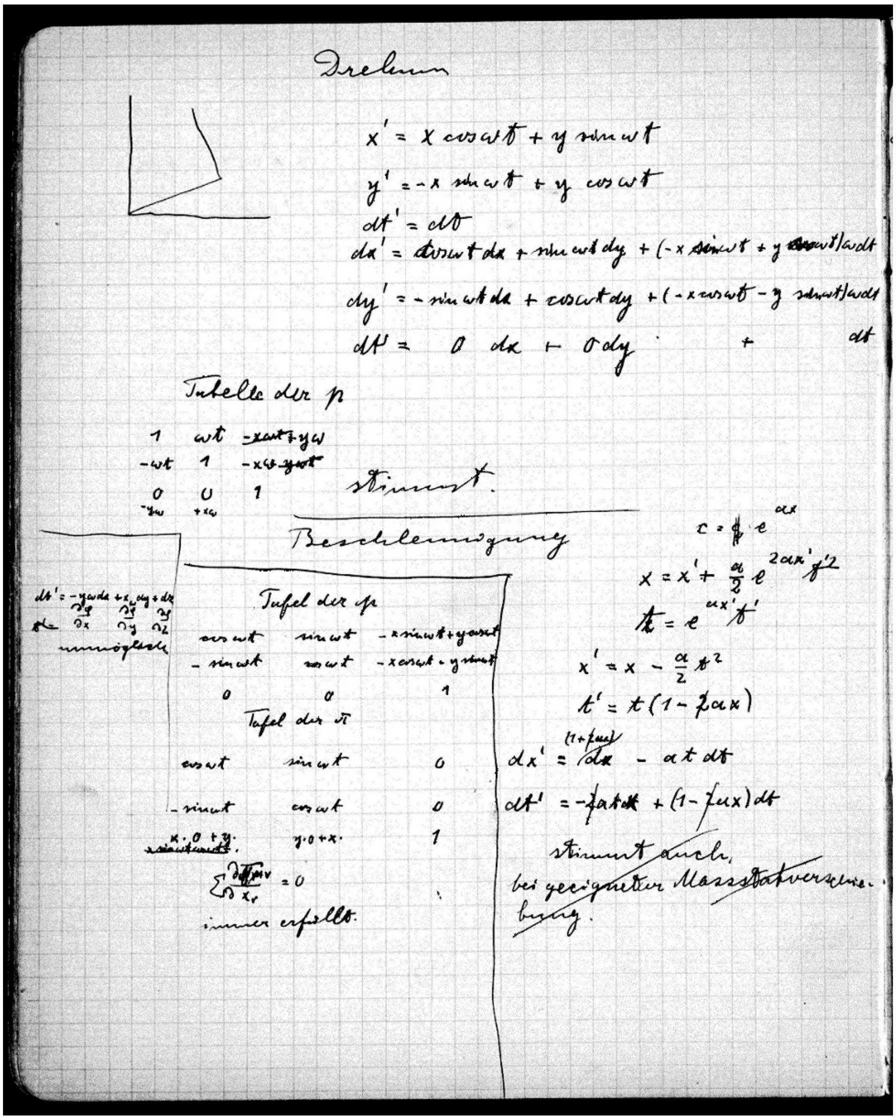
$$Y = \beta_1 xy + \beta_2(x^2 - y^2) + y$$

$$\alpha_1 y + 2\alpha_2 x + \beta_1 x - 2\beta_2 y = 0$$

$$\beta_1 = -2\alpha_2$$

$$\alpha_1 = 2\beta_2$$

[p. 11 L]


Drehung

$x' = x\cos\omega t + y\sin\omega t$

$y' = -x\sin\omega t + y\cos\omega t$

$t' = t$

$dx' = \cos\omega t\,dx + \sin\omega t\,dy + (-x\sin\omega t + y\cos\omega t)\omega dt$

$dy' = -\sin\omega t\,dx + \cos\omega t\,dy + (-x\cos\omega t - y\sin\omega t)\omega dt$

$dt' = 0\,dx + 0\,dy + dt$

Tabelle der p

1	ωt	$-x\omega t + y\omega$
$-\omega t$	1	$-x\omega - y\omega t$
0	0	1

stimmt.

Beschleunigung

$c = f e^{\alpha x}$

$x = x' + \frac{\alpha}{2} e^{2\alpha x'} t'^2$

$t = e^{\alpha x'} t'$

$x' = x - \frac{\alpha}{2} t^2$

$t' = t(1 - f\alpha x)$

$dx' = (1 + f\alpha x)\,dx - \alpha t\,dt$

$dt' = -f\alpha t\,dx + (1 - f\alpha x)\,dt$

~~stimmt auch.~~ bei geeigneter Massstabverschiebung.

$dt' = -y\omega dx + x\omega dy + dt$

$\frac{\partial g}{\partial x}\ \frac{\partial g}{\partial y}\ \frac{\partial g}{\partial z}$ unmöglich

Tafel der p

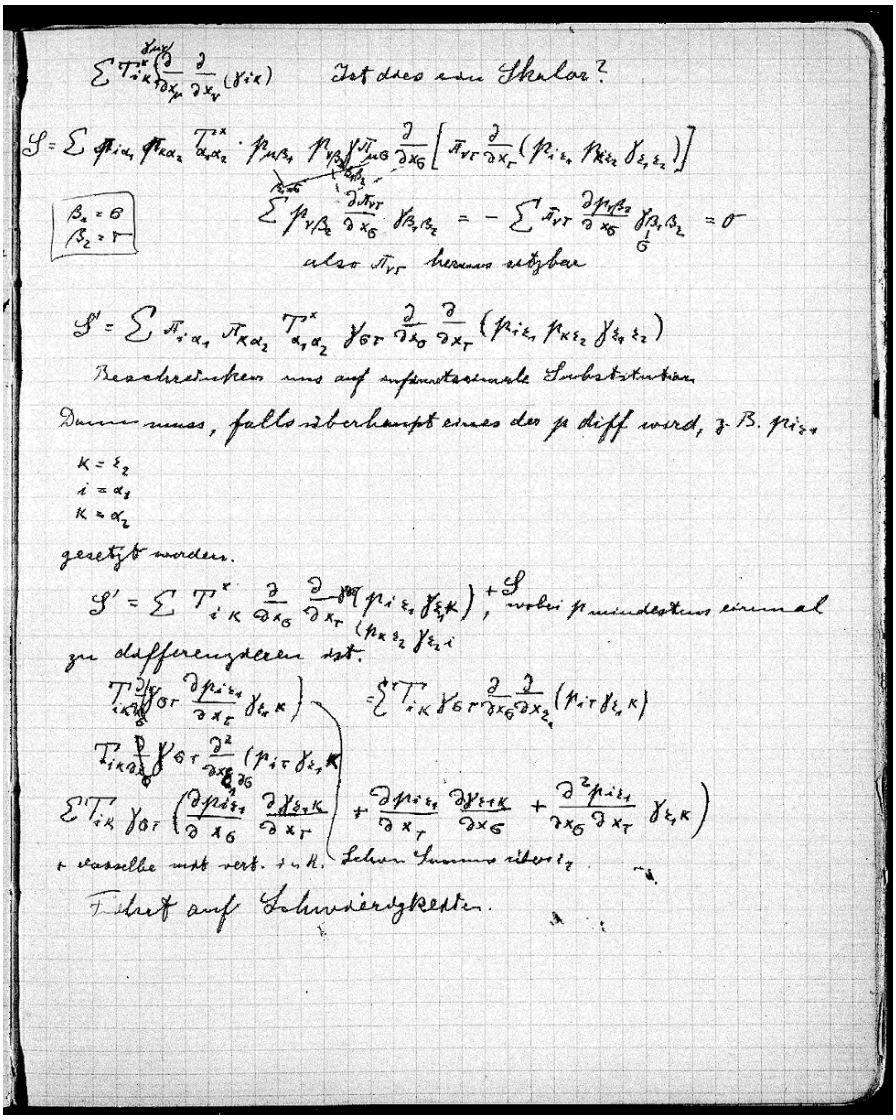
$\cos\omega t$	$\sin\omega t$	$-x\sin\omega t + y\cos\omega t$
$-\sin\omega t$	$\cos\omega t$	$-x\cos\omega t - y\sin\omega t$
0	0	1

Tafel der π

$\cos\omega t$	$\sin\omega t$	0
$-\sin\omega t$	$\cos\omega t$	0
$x\cdot 0 + y\cdot$	$y\cdot 0 + x\cdot$	1

$\sum \frac{\partial T_{\mu\nu}}{\partial x_\nu} = 0$ immer erfüllt.

[p. 11 R]



$$\sum T_{ik}^{x} \gamma_{\mu\nu} \frac{\partial}{\partial x_\mu} \frac{\partial}{\partial x_\nu}(\gamma_{ik})$$ Ist dies ein Skalar?

$$\mathfrak{S} = \sum p_{i\alpha_1} p_{\kappa\alpha_2} T_{\alpha_1\alpha_2}^{x} \cdot p_{\mu\beta_1} p_{\nu\beta_2} \gamma_{\beta_1\beta_2} \pi_{\mu\sigma} \frac{\partial}{\partial x_\sigma}\left[\pi_{\nu\tau} \frac{\partial}{\partial x_\tau}(p_{i\varepsilon_1} p_{\kappa\varepsilon_2} \gamma_{\varepsilon_1\varepsilon_2})\right]$$

$$\boxed{\begin{array}{l}\beta_1 = \sigma \\ \beta_2 = \tau\end{array}}$$

$$\sum p_{\nu\beta_2} \frac{\partial \pi_{\nu\tau}}{\partial x_\sigma} \gamma_{\beta_1\beta_2} = -\sum \pi_{\nu\tau} \frac{\partial p_{\nu\beta_2}}{\partial x_\sigma} \gamma_{\beta_1\beta_2} = 0$$

also $\pi_{\nu\tau}$ heraus setzbar

$$\mathfrak{S}' = \sum \pi_{i\alpha_1} \pi_{\kappa\alpha_2} T_{\alpha_1\alpha_2}^{x} \gamma_{\sigma\tau} \frac{\partial}{\partial x_\sigma} \frac{\partial}{\partial x_\tau}(p_{i\varepsilon_1} p_{\kappa\varepsilon_2} \gamma_{\varepsilon_1\varepsilon_2})$$

Beschränken uns auf infinitesimale Substitutionen

Dann muss, falls überhaupt eines der p diff wird, z. B. $p_{i\varepsilon_1}$

$\kappa = \varepsilon_2$
$i = \alpha_1$
$\kappa = \alpha_2$

gesetzt werden.

$$\mathfrak{S}' = \sum T_{i\kappa}^{x} \frac{\partial}{\partial x_\sigma} \frac{\partial}{\partial x_\tau}\left(p_{i\varepsilon_1} \gamma_{\varepsilon_1\kappa}\right) + \mathfrak{S}$$ wobei p mindestens einmal zu differenzieren ist.

$(p_{\kappa\varepsilon_2} \gamma_{\varepsilon_2 i}$

$$T_{i\kappa} \gamma_{\sigma\tau} \frac{\partial p_{i\varepsilon_1}}{\partial x_\sigma} \gamma_{\varepsilon_1\kappa}\Big) \qquad = \sum T_{i\kappa} \gamma_{\sigma\tau} \frac{\partial}{\partial x_\sigma} \frac{\partial}{\partial x_{\tau}} (p_{i\tau} \gamma_{\varepsilon_1\kappa})$$

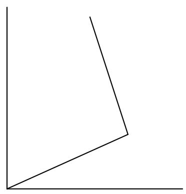
$$T_{i\kappa} \gamma_{\sigma\tau} \frac{\partial^2}{\partial x_\sigma \partial x_\tau}(p_{i\tau} \gamma_{\varepsilon_1\kappa}$$

$$\sum T_{i\kappa} \gamma_{\sigma\tau}\left(\frac{\partial p_{i\varepsilon_1}}{\partial x_\sigma} \frac{\partial \gamma_{\varepsilon_1\kappa}}{\partial x_\tau} + \frac{\partial p_{i\varepsilon_1}}{\partial x_\tau} \frac{\partial \gamma_{\varepsilon_1\kappa}}{\partial x_\sigma} + \frac{\partial^2 p_{i\varepsilon_1}}{\partial x_\sigma \partial x_\tau} \gamma_{\varepsilon_1\kappa}\right)$$

+ dasselbe mit vert. i u. κ. Schon Summe über ε_1

Führt auf Schwierigkeiten.

[p. 11 L]



Drehung

$$x' = x\cos\omega t + y\sin\omega t$$
$$y' = -x\sin\omega t + y\cos\omega t$$

$$dt' = dt$$

$$dx' = \cos\omega t\ dx + \sin\omega t\ dy + (-x\sin\omega t + y\cos\omega t)\omega dt$$

$$dy' = -\sin\omega t\ dx + \cos\omega t\ dy + (-x\cos\omega t - y\sin\omega t)\,\omega\, dt$$

$$dt' = \quad 0 \quad dx + \quad 0 \quad dy \qquad + \qquad dt$$

Tabelle der p

$$\begin{array}{ccc} 1 & \omega t & -\cancel{x\omega t} + y\omega \\ -\omega t & 1 & -x\omega \ \ \cancel{y\omega t} \\ 0 & 0 & 1 \\ -y\omega & +x\omega & \end{array}$$

$$dt' = -y\omega dx + x\omega dy + dz$$

$$\langle t' = \rangle \quad \frac{\partial\varphi}{\partial x} \quad \frac{\partial\varphi}{\partial y} \quad \frac{\partial\varphi}{\partial z}$$

unmöglich.

stimmt.

Beschleunigung

$$c = \langle c_0 \rangle e^{ax}$$

Tafel der p

$\cos\omega t$	$\sin\omega t$	$-x\sin\omega t + y\cos\omega t$
$-\sin\omega t$	$\cos\omega t$	$-x\cos\omega t + y\sin\omega t$
0	0	1

Tafel der π

$\cos\omega t$	$\sin\omega t$	0
$-\sin\omega t$	$\cos\omega t$	0
$x \cdot 0 + y \cdot$	$y \cdot 0 + x \cdot$	1

$$\langle x\sin\omega t\cos\omega t x + \cdot \rangle$$

$$\sum \frac{\partial \pi_{\mu\nu}}{\partial x_\nu} = 0$$

immer erfüllt.

$$x = x' + \frac{a}{2}e^{2ax'}t'^2$$
$$t = e^{ax'}t'$$

$$x' = x - \frac{a}{2}t^2$$
$$t' = t(1 - \langle 2 \rangle ax)$$
$$(1 + \langle 2 \rangle ax)$$
$$dx' = dx - atdt$$

$$dt' = -\langle 2 \rangle atdx + (1 - \langle 2 \rangle ax)dt$$

~~stimmt auch,~~
~~bei geeigneter Massstabverschiebung.~~

[p. 11 R]

$$\sum T^x_{i\kappa}\overbrace{\frac{\partial}{\partial x_\mu}\frac{\partial}{\partial x_\nu}}^{\gamma_{\mu\nu}}(\gamma_{i\kappa})$$ Ist dies ein Skalar?

$$S = \sum \pi_{i\alpha_1}\pi_{\kappa\alpha_2}T^x_{\alpha_1\alpha_2}\cdot p_{\mu\beta_1}p_{\nu\beta_2}\gamma_{\beta_1\beta_2}\pi_{\mu\sigma}\frac{\partial}{\partial x_\sigma}\left[\pi_{\nu\tau}\frac{\partial}{\partial x_\tau}(p_{i\varepsilon_1}p_{\kappa\varepsilon_2}\gamma_{\varepsilon_1\varepsilon_2})\right]$$

$\boxed{\begin{matrix}\beta_1 = \sigma\\ \beta_2 = \tau\end{matrix}}$ $\beta_1 = \sigma$

$$\sum p_{\nu\beta_2}\frac{\partial \pi_{\nu\tau}}{\partial x_\sigma}\gamma_{\beta_1\beta_2} = -\sum \pi_{\nu\tau}\frac{\partial p_{\nu\beta_2}}{\partial x_\sigma}\gamma_{\underset{\sigma}{\beta_1}\beta_2} = 0$$

also $\pi_{\nu\tau}$ heraus setzbar

$$S' = \sum \pi_{i\alpha_1}\pi_{\kappa\alpha_2}T^x_{\alpha_1\alpha_2}\gamma_{\sigma\tau}\frac{\partial}{\partial x_\sigma}\frac{\partial}{\partial x_\tau}(p_{i\varepsilon_1}p_{\kappa\varepsilon_2}\gamma_{\varepsilon_1\varepsilon_2})$$

Beschränken uns auf infinitesimale Substitution

Dann muss, falls überhaupt eines der p diff wird, z. B. $p_{i\varepsilon_1}$

$\kappa = \varepsilon_2$
$i = \alpha_1$
$\kappa = \alpha_2$

gesetzt werden.

$$S' = \sum T^x_{i\kappa}\frac{\partial}{\partial x_\sigma}\frac{\partial}{\partial x_\tau}\gamma_{\sigma\tau}\underset{(p_{\kappa\varepsilon_2}\gamma_{\varepsilon_2 i}}{(p_{i\varepsilon_1}\gamma_{\varepsilon_1\kappa})}^{+S},$$ wobei p mindestens einmal

zu differenzieren ist.

$$T_{i\kappa}\frac{\partial}{\partial x_\sigma}\left(\gamma_{\sigma\tau}\frac{\partial p_{i\varepsilon_1}}{\partial x_\tau}\gamma_{\varepsilon_1\kappa}\right) = \sum T_{i\kappa}\gamma_{\sigma\tau}\frac{\partial}{\partial x_\sigma}\frac{\partial}{\partial x_{\varepsilon_1}}(p_{i\tau}\gamma_{\varepsilon_1\kappa})$$

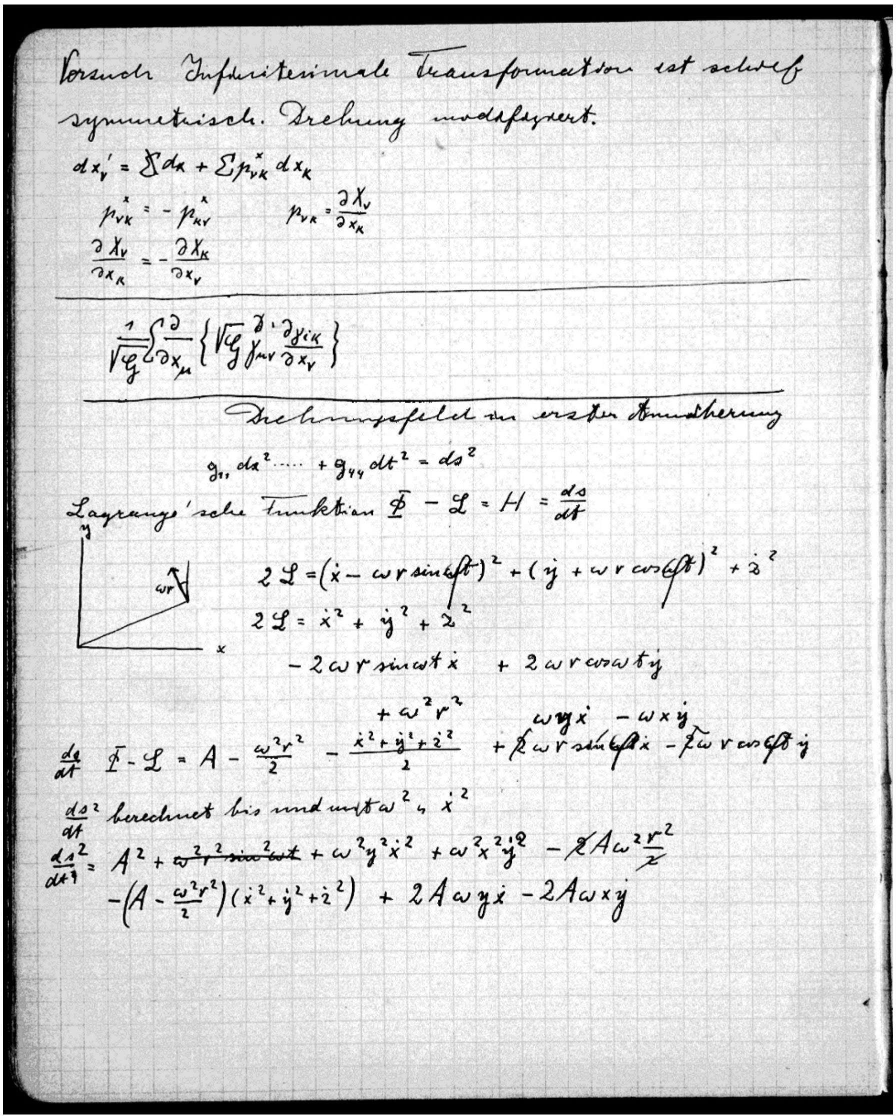
$$T_{i\kappa}\cancel{\frac{\partial}{\partial x_\sigma}}\gamma_{\sigma\tau}\frac{\partial^2}{\partial x_{\varepsilon_1}\partial\sigma}(p_{i\tau}\gamma_{\varepsilon_1\kappa})$$

$$\sum T_{i\kappa}\gamma_{\sigma\tau}\left(\frac{\partial p_{i\varepsilon_1}}{\partial x_\sigma}\frac{\partial\gamma_{\varepsilon_1\kappa}}{\partial x_\tau} + \frac{\partial p_{i\varepsilon_1}}{\partial x_\tau}\frac{\partial\gamma_{\varepsilon_1\kappa}}{\partial x_\sigma} + \frac{\partial^2 p_{i\varepsilon_1}}{\partial x_\sigma\partial x_\tau}\gamma_{\varepsilon_1\kappa}\right)$$

+ dasselbe mit vert. i u κ. Schon Summe über ε_1

Führt auf Schwierigkeiten.

[p. 12 L]



Versuch Infinitesimale Transformation ist schief symmetrisch. Drehung undefiniert.

$$dx_\nu' = \delta d_\kappa + \sum p_{\nu\kappa}^{x} dx_\kappa$$

$$p_{\nu\kappa}^{x} = -p_{\kappa\nu}^{x} \qquad p_{\nu\kappa} = \frac{\partial X_\nu}{\partial x_\kappa}$$

$$\frac{\partial X_\nu}{\partial x_\kappa} = -\frac{\partial X_\kappa}{\partial x_\nu}$$

$$\frac{1}{\sqrt{G}} \sum \frac{\partial}{\partial x_\mu} \left\{ \sqrt{G}\, \gamma_{\mu\nu} \frac{\partial \gamma_{i\kappa}}{\partial x_\nu} \right\}$$

Drehungsfeld in erster Annäherung

$$g_{11}\, dx^2 \ldots\ldots + g_{44}\, dt^2 = ds^2$$

Lagrange'sche Funktion $\Phi - \mathfrak{L} = H = \frac{ds}{dt}$

$$2\mathfrak{L} = (\dot{x} - \omega r \sin \omega t)^2 + (\dot{y} + \omega r \cos \omega t)^2 + \dot{z}^2$$

$$2\mathfrak{L} = \dot{x}^2 + \dot{y}^2 + \dot{z}^2 - 2\omega r \sin \omega t\, \dot{x} + 2\omega r \cos \omega t\, \dot{y} + \omega^2 r^2$$

$$\frac{ds}{dt} \quad \Phi - \mathfrak{L} = A - \frac{\omega^2 r^2}{2} - \frac{\dot{x}^2 + \dot{y}^2 + \dot{z}^2}{2} + \cancel{2}\omega r \sin \omega t\, \dot{x} - \cancel{2}\omega r \cos \omega t\, \dot{y}$$

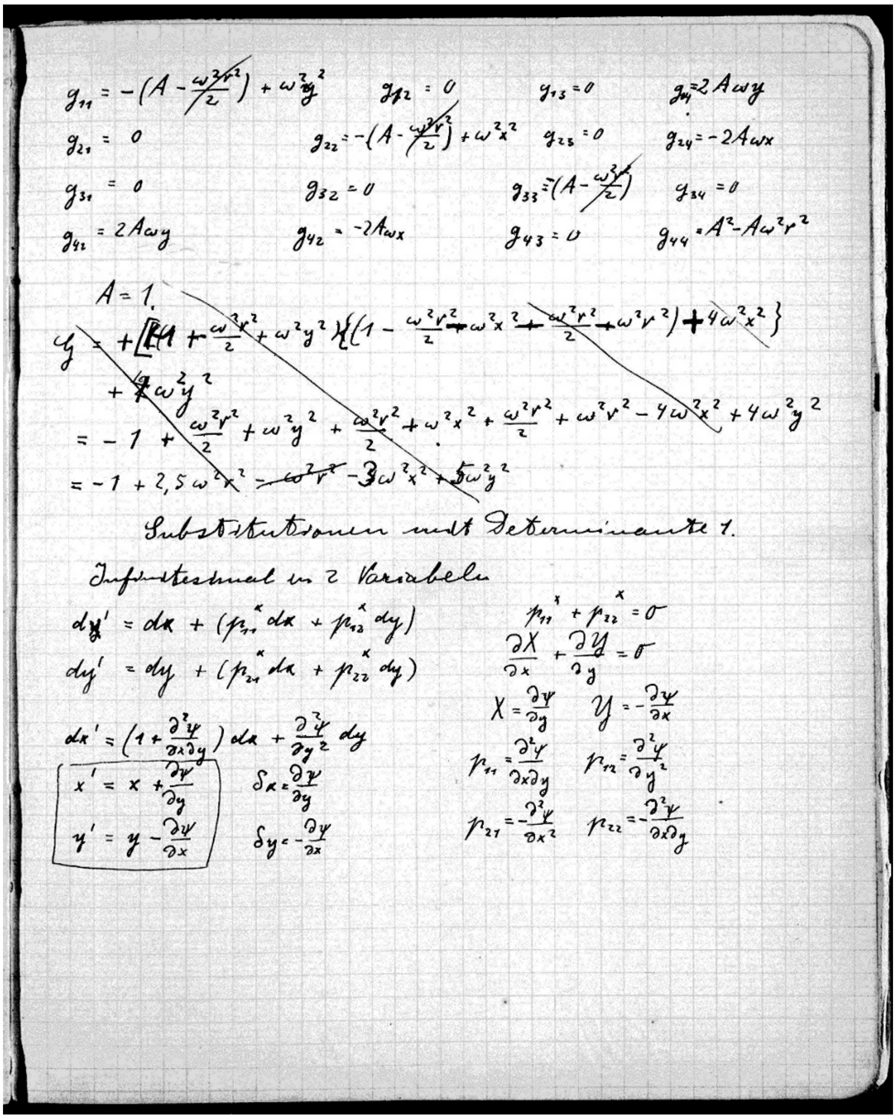
$$\omega y \dot{x} \quad - \omega x \dot{y}$$

$\frac{ds^2}{dt}$ berechnet bis und mit ω^2, $\dot{x}^2$

$$\frac{ds^2}{dt^2} = A^2 + \cancel{\omega^2 r^2 \sin^2 \omega t} + \omega^2 y^2 \dot{x}^2 + \omega^2 x^2 \dot{y}^2 - \cancel{2} A \omega^2 \frac{r^2}{\cancel{2}}$$

$$- \left(A - \frac{\omega^2 r^2}{2}\right)(\dot{x}^2 + \dot{y}^2 + \dot{z}^2) + 2A\omega y \dot{x} - 2A\omega x \dot{y}$$

[p. 12 R]



$$g_{11} = -\left(A - \cancel{\frac{\omega^2 r^2}{2}}\right) + \omega^2 y^2 \qquad g_{12} = 0 \qquad g_{13} = 0 \qquad g_{14} = 2A\omega y$$

$$g_{21} = 0 \qquad g_{22} = -\left(A - \cancel{\frac{\omega^2 r^2}{2}}\right) + \omega^2 x^2 \qquad g_{23} = 0 \qquad g_{24} = -2A\omega x$$

$$g_{31} = 0 \qquad g_{32} = 0 \qquad g_{33} = -\left(A - \cancel{\frac{\omega^2 r^2}{2}}\right) \qquad g_{34} = 0$$

$$g_{41} = 2A\omega y \qquad g_{42} = -2A\omega x \qquad g_{43} = 0 \qquad g_{44} = A^2 - A\omega^2 r^2$$

$A = 1.$

$$\cancel{G = +\left[\left(1 + \frac{\omega^2 r^2}{2} + \omega^2 y^2\right)\left\{\left(1 - \frac{\omega^2 r^2}{2} + \omega^2 x^2 + \frac{\omega^2 r^2}{2} + \omega^2 r^2\right) + 4\omega^2 x^2\right\}\right.}$$

$$\cancel{+ 4\omega^2 y^2}$$

$$\cancel{= -1 + \frac{\omega^2 r^2}{2} + \omega^2 y^2 + \frac{\omega^2 r^2}{2} + \omega^2 x^2 + \frac{\omega^2 r^2}{2} + \omega^2 r^2 - 4\omega^2 x^2 + 4\omega^2 y^2}$$

$$\cancel{= -1 + 2{,}5\,\omega^2 x^2 - \omega^2 r^2 - 3\omega^2 x^2 + 5\omega^2 y^2}$$

Substitutionen mit Determinante 1.

Infinitesimal in 2 Variabeln

$$dx' = dx + (p^{\times}_{11}\,dx + p^{\times}_{12}\,dy) \qquad p^{\times}_{11} + p^{\times}_{22} = 0$$

$$dy' = dy + (p^{\times}_{21}\,dx + p^{\times}_{22}\,dy) \qquad \frac{\partial X}{\partial x} + \frac{\partial Y}{\partial y} = 0$$

$$X = \frac{\partial \psi}{\partial y} \qquad Y = -\frac{\partial \psi}{\partial x}$$

$$dx' = \left(1 + \frac{\partial^2 \psi}{\partial x \partial y}\right) dx + \frac{\partial^2 \psi}{\partial y^2}\,dy \qquad p_{11} = \frac{\partial^2 \psi}{\partial x \partial y} \qquad p_{12} = \frac{\partial^2 \psi}{\partial y^2}$$

$$\boxed{\begin{aligned} x' &= x + \frac{\partial \psi}{\partial y} \\ y' &= y - \frac{\partial \psi}{\partial x} \end{aligned}} \qquad \begin{aligned} \delta x &= \frac{\partial \psi}{\partial y} \\ \delta y &= -\frac{\partial \psi}{\partial x} \end{aligned} \qquad p_{21} = -\frac{\partial^2 \psi}{\partial x^2} \qquad p_{22} = -\frac{\partial^2 \psi}{\partial x \partial y}$$

[p. 12 L]

versuch infinitesimale transformation ist schief symmetrisch. drehung modifiziert.

$$dx'_\nu = \cancel{\sum} dx + \sum p^x_{\nu\kappa}\, dx_\kappa$$

$$p^x_{\nu\kappa} = -p^x_{\kappa\nu} \qquad\qquad p_{\nu\kappa} = \frac{\partial X_\nu}{\partial x_\kappa}$$

$$\frac{\partial X_\nu}{\partial x_\kappa} = -\frac{\partial X_\kappa}{\partial x_\nu}$$

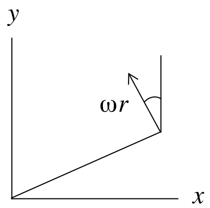
$$\frac{1}{\sqrt{G}}\sum \frac{\partial}{\partial x_\mu}\left\{\sqrt{G}\,\cancel{\partial}\,\gamma_{\mu\nu}\frac{\partial \gamma_{i\kappa}}{\partial x_\nu}\right\}$$

---

drehungsfeld in erster annäherung

$$g_{11}dx^2 \cdot\cdot\cdot\cdot\cdot + g_{44}dt^2 = ds^2$$

lagrange'sche funktion $\quad \Phi - L = H = \dfrac{ds}{dt}$



$$2L = (\dot{x} - \omega r \sin\langle\omega t\rangle\varphi)^2 + (\dot{y} + \omega r\cos\langle\omega t\rangle\varphi)^2 + \dot{z}^2$$

$$2L = \dot{x}^2 + \dot{y}^2 + \dot{z}^2$$

$$-\ 2\omega r\sin\omega t\dot{x} \qquad + \ 2\omega r\cos\omega t\dot{y}$$

$$+\ \omega^2 r^2$$

$$\omega y\dot{x} \qquad -\ \omega x\dot{y}$$

$$\frac{ds}{dt} \qquad \Phi - L = A - \frac{\omega^2 r^2}{2} - \frac{\dot{x}^2 + \dot{y}^2 + \dot{z}^2}{2} + \langle 2\rangle\omega r\sin\langle\omega t\rangle\varphi\dot{x} - \langle 2\rangle\omega r\cos\langle\omega t\rangle\varphi\dot{y}$$

$\dfrac{ds^2}{dt}$ berechnet bis und mit $\quad \omega^2$ u $\dot{x}^2$

$$\frac{ds^2}{dt} = A^2 + \langle\omega^2 r^2\sin^2\omega t\rangle + \omega^2 y^2\dot{x}^2 + \omega^2 x^2\dot{y}^2 - \cancel{2}A\omega^2\frac{r^2}{\cancel{2}}$$

$$-\left(A - \frac{\omega^2 r^2}{2}\right)(\dot{x}^2 + \dot{y}^2 + \dot{z}^2) + 2A\omega y\dot{x} - 2A\omega x\dot{y}$$

[p. 12 R]

$$
\begin{array}{llll}
g_{11} = -\left(A - \cancel{\frac{\omega^2 r^2}{2}}\right) + \omega^2 y^2 & g_{12} = 0 & g_{13} = 0 & g_{14} = 2A\omega y \\
g_{21} = 0 & g_{22} = -\left(A - \cancel{\frac{\omega^2 r^2}{2}}\right) + \omega^2 x^2 & g_{23} = 0 & g_{24} = -2A\omega x \\
g_{31} = 0 & g_{32} = 0 & g_{33} = -\left(A - \cancel{\frac{\omega^2 r^2}{2}}\right) & g_{34} = 0 \\
g_{41} = 2A\omega y & g_{42} = -2A\omega x & g_{43} = 0 & g_{44} = A^2 - A\omega^2 r^2
\end{array}
$$

$A = 1.$

$$G = + \left[\left(-1 + \frac{\omega^2 r^2}{2} + \omega^2 y^2\right)\left\{\left(1 - \frac{\omega^2 r^2}{2} - \omega^2 x^2 - \frac{\omega^2 r^2}{2} - \omega^2 r^2\right) + 4\omega^2 x^2\right\}\right.$$

$$+ \langle 2 \rangle 4\omega^2 y^2$$

$$= -1 + \frac{\omega^2 r^2}{2} + \omega^2 y^2 + \frac{\omega^2 r^2}{2} + \omega^2 x^2 + \frac{\omega^2 r^2}{2} + \omega^2 r^2 - 4\omega^2 x^2 + 4\omega^2 y^2$$

$$= -1 + 2{,}5\omega^2 r^2 - \langle \omega^2 r^2 \rangle - 3\omega^2 x^2 + 5\omega^2 y^2$$

Substitutionen mit Determinante 1.

Infinitesimal in 2 Variabeln

$$dx' = dx + (p^x_{11} dx + p^x_{12} dy) \qquad p^x_{11} + p^x_{22} = 0$$

$$dy' = dy + (p^x_{21} dx + p^x_{22} dy) \qquad \frac{\partial X}{\partial x} + \frac{\partial Y}{\partial y} = 0$$

$$X = \frac{\partial \psi}{\partial y} \qquad Y = -\frac{\partial \psi}{\partial x}$$

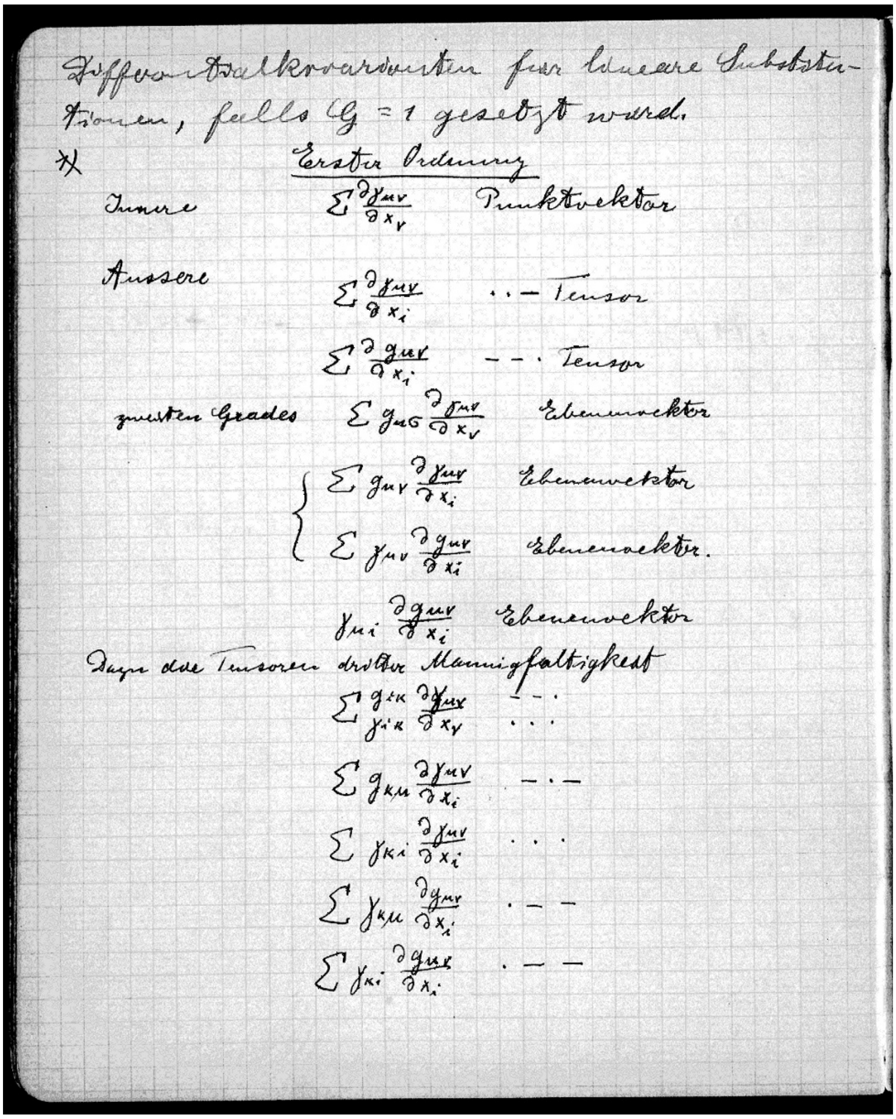
$$dx' = \left(1 + \frac{\partial^2 \psi}{\partial x \partial y}\right) dx + \frac{\partial^2 \psi}{\partial y^2} dy$$

$$p_{11} = \frac{\partial^2 \psi}{\partial x \partial y} \qquad p_{12} = \frac{\partial^2 \psi}{\partial y^2}$$

$$\boxed{\begin{aligned} x' &= x + \frac{\partial \psi}{\partial y} \\ y' &= x - \frac{\partial \psi}{\partial x} \end{aligned}} \qquad \begin{aligned} \delta x &= \frac{\partial \psi}{\partial y} \\ \delta y &= -\frac{\partial \psi}{\partial x} \end{aligned}$$

$$p_{21} = -\frac{\partial^2 \psi}{\partial x^2} \qquad p_{22} = -\frac{\partial^2 \psi}{\partial x \partial y}$$

[p. 13 L]



Differentialkovarianten für lineare Substitutionen, falls $G = 1$ gesetzt wird.

x) Erster Ordnung

Innere $\sum \frac{\partial \gamma_{\mu\nu}}{\partial x_\nu}$ Punktvektor

Äussere $\sum \frac{\partial \gamma_{\mu\nu}}{\partial x_i}$ ..— Tensor

$\sum \frac{\partial g_{\mu\nu}}{\partial x_i}$ —-. Tensor

zweiten Grades $\sum g_{\mu\sigma} \frac{\partial \gamma_{\mu\nu}}{\partial x_\nu}$ Ebenenvektor

$$\left\{ \begin{array}{ll} \sum g_{\mu\nu} \frac{\partial \gamma_{\mu\nu}}{\partial x_i} & \text{Ebenenvektor} \\ \sum \gamma_{\mu\nu} \frac{\partial g_{\mu\nu}}{\partial x_i} & \text{Ebenenvektor.} \end{array} \right.$$

$\gamma_{\mu i} \frac{\partial g_{\mu\nu}}{\partial x_i}$ Ebenenvektor

Dazu die Tensoren dritter Mannigfaltigkeit

$\sum \begin{matrix} g_{ik} \\ \gamma_{ik} \end{matrix} \frac{\partial \gamma_{\mu\nu}}{\partial x_\nu}$ $\begin{matrix} —-. \\ ... \end{matrix}$

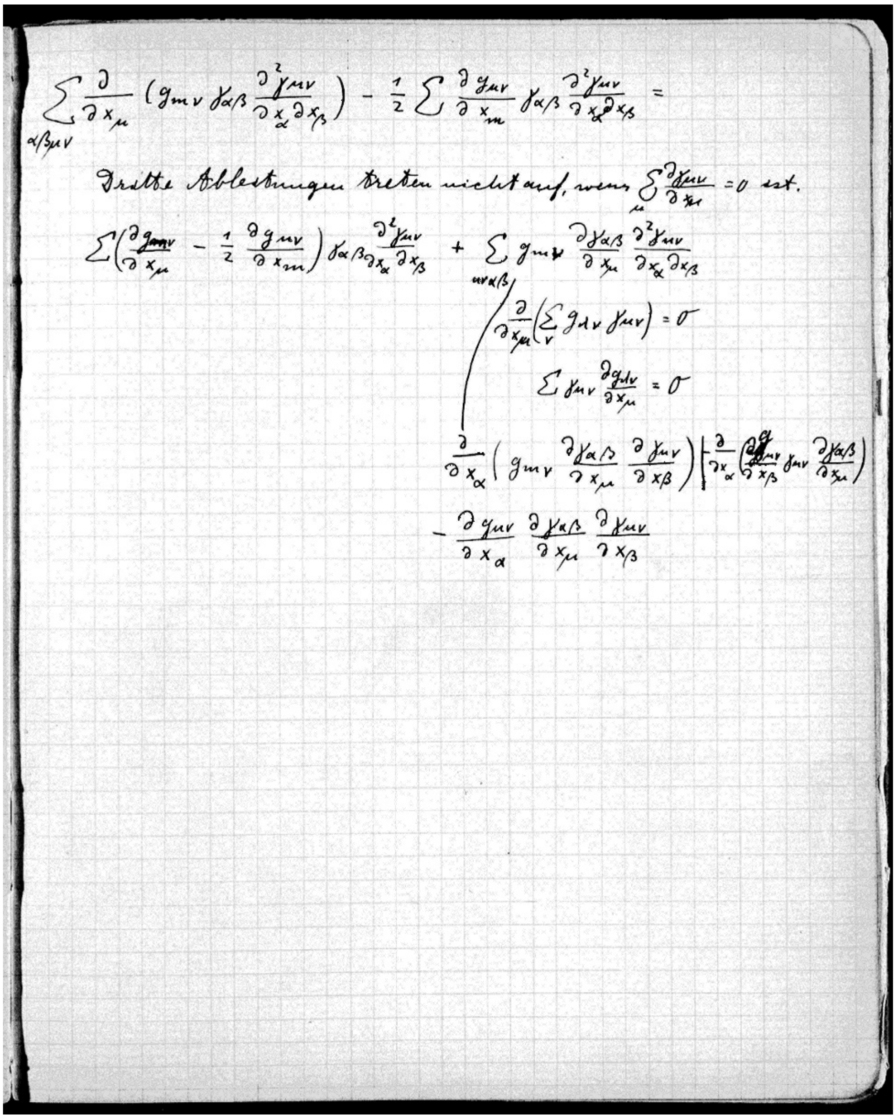
$\sum g_{\kappa\mu} \frac{\partial \gamma_{\mu\nu}}{\partial x_i}$ —.—

$\sum \gamma_{\kappa i} \frac{\partial \gamma_{\mu\nu}}{\partial x_i}$ ...

$\sum \gamma_{\kappa\mu} \frac{\partial g_{\mu\nu}}{\partial x_i}$.——

$\sum \gamma_{\kappa i} \frac{\partial g_{\mu\nu}}{\partial x_i}$.——

[p. 13 R]



$$\sum_{\alpha\beta\mu\nu} \frac{\partial}{\partial x_\mu}\left(g_{m\nu}\gamma_{\alpha\beta}\frac{\partial^2\gamma_{\mu\nu}}{\partial x_\alpha\partial x_\beta}\right) - \frac{1}{2}\sum \frac{\partial g_{\mu\nu}}{\partial x_m}\gamma_{\alpha\beta}\frac{\partial^2\gamma_{\mu\nu}}{\partial x_\alpha\partial x_\beta} =$$

Dritte Ableitungen treten nicht auf, wenn $\sum_\mu \frac{\partial\gamma_{\mu\nu}}{\partial x_\mu} = 0$ ist.

$$\sum\left(\frac{\partial g_{m\nu}}{\partial x_\mu} - \frac{1}{2}\frac{\partial g_{\mu\nu}}{\partial x_m}\right)\gamma_{\alpha\beta}\frac{\partial^2\gamma_{\mu\nu}}{\partial x_\alpha\partial x_\beta} + \sum_{\mu\nu\alpha\beta} g_{m\nu}\frac{\partial\gamma_{\alpha\beta}}{\partial x_\mu}\frac{\partial^2\gamma_{\mu\nu}}{\partial x_\alpha\partial x_\beta}$$

$$\frac{\partial}{\partial x_\mu}\left(\sum_\nu g_{\lambda\nu}\gamma_{\mu\nu}\right) = 0$$

$$\sum \gamma_{\mu\nu}\frac{\partial g_{\lambda\nu}}{\partial x_\mu} = 0$$

$$\frac{\partial}{\partial x_\alpha}\left(g_{m\nu}\frac{\partial\gamma_{\alpha\beta}}{\partial x_\mu}\frac{\partial\gamma_{\mu\nu}}{\partial x_\beta}\right) - \frac{\partial}{\partial x_\alpha}\left(\frac{\partial g_{m\nu}}{\partial x_\beta}\gamma_{\mu\nu}\frac{\partial\gamma_{\alpha\beta}}{\partial x_\mu}\right)$$

$$-\frac{\partial g_{\mu\nu}}{\partial x_\alpha}\frac{\partial\gamma_{\alpha\beta}}{\partial x_\mu}\frac{\partial\gamma_{\mu\nu}}{\partial x_\beta}$$

[p. 13 L]

Differentialkovarianten für lineare Substitutionen, falls $G = 1$ gesetzt wird.

~~1)~~ *Erster Ordnung*

Innere $\sum \frac{\partial \gamma_{\mu\nu}}{\partial x_\nu}$ Punktvektor

Aussere $\sum \frac{\partial \gamma_{\mu\nu}}{\partial x_i}$ · · – Tensor

$\sum \frac{\partial g_{\mu\nu}}{\partial x_i}$ – – · Tensor

zweiten Grades $\sum g_{\mu\sigma} \frac{\partial \gamma_{\mu\nu}}{\partial x_\nu}$ Ebenenvektor

$\sum g_{\mu\nu} \frac{\partial \gamma_{\mu\nu}}{\partial x_i}$ Ebenenvektor

$\sum \gamma_{\mu\nu} \frac{\partial g_{\mu\nu}}{\partial x_i}$ Ebenenvektor

$\gamma_{\mu i} \frac{\partial g_{\mu\nu}}{\partial x_i}$ Ebenenvektor

Dazu die Tensoren dritter Mannigfaltigkeit

$\sum \substack{g_{i\kappa} \\ \gamma_{i\kappa}} \frac{\partial \gamma_{\mu\nu}}{\partial x_\nu}$ – – ·
· · ·

$\sum g_{\kappa\mu} \frac{\partial \gamma_{\mu\nu}}{\partial x_i}$ – · –

$\sum \gamma_{\kappa i} \frac{\partial \gamma_{\mu\nu}}{\partial x_i}$ · · ·

$\sum \gamma_{\kappa\mu} \frac{\partial g_{\mu\nu}}{\partial x_i}$ · – –

$\sum \gamma_{\kappa i} \frac{\partial g_{\mu\nu}}{\partial x_i}$ · – –

[p. 13 R]

$$\sum_{\alpha\beta\mu\nu} \frac{\partial}{\partial x_\mu}\left(g_{m\nu}\gamma_{\alpha\beta}\frac{\partial^2\gamma_{\mu\nu}}{\partial x_\alpha \partial x_\beta}\right) - \frac{1}{2}\sum \frac{\partial g_{\mu\nu}}{\partial x_m}\gamma_{\alpha\beta}\frac{\partial^2\gamma_{\mu\nu}}{\partial x_\alpha \partial x_\beta} =$$

Dritte Ableitungen treten nicht auf, wenn $\sum_\mu \frac{\partial \gamma_{\mu\nu}}{\partial x_\mu} = 0$ ist.

$$\sum\left(\frac{\partial g_{m\nu}}{\partial x_\mu} - \frac{1}{2}\frac{\partial g_{\mu\nu}}{\partial x_m}\right)\gamma_{\alpha\beta}\frac{\partial^2\gamma_{\mu\nu}}{\partial x_\alpha \partial x_\beta} + \sum_{\mu\nu\alpha\beta} g_{m\nu}\frac{\partial \gamma_{\alpha\beta}}{\partial x_\mu}\frac{\partial^2\gamma_{\mu\nu}}{\partial x_\alpha \partial x_\beta}$$

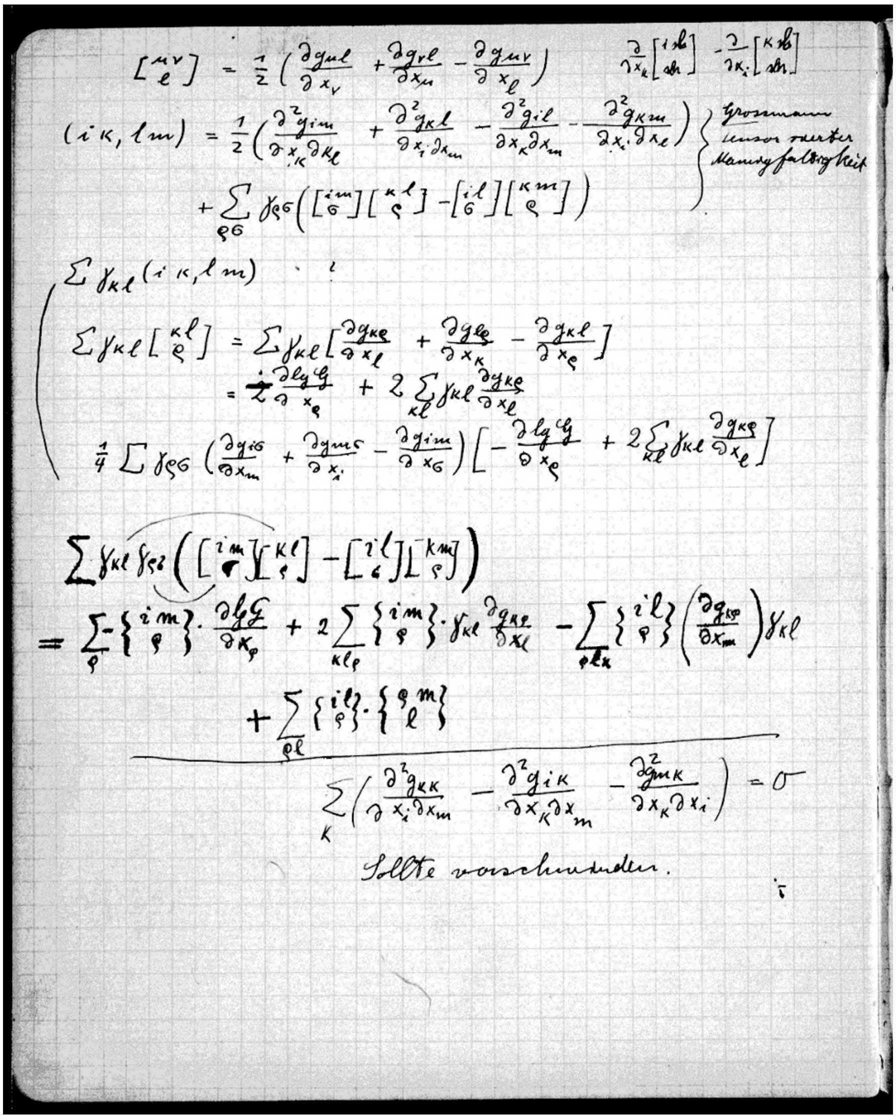
$$\frac{\partial}{\partial x_\mu}\left(\sum_\nu g_{\lambda\nu}\gamma_{\mu\nu}\right) = 0$$

$$\sum \gamma_{\mu\nu}\frac{\partial g_{\lambda\nu}}{\partial x_\mu} = 0$$

$$\frac{\partial}{\partial x_\alpha}\left(g_{m\nu}\frac{\partial \gamma_{\alpha\beta}}{\partial x_\mu}\frac{\partial \gamma_{\mu\nu}}{\partial x_\beta}\right) \quad \Bigg| \quad -\frac{\partial}{\partial x_\alpha}\left(\frac{\partial g_{\mu\nu}}{\partial x_\beta}\gamma_{\mu\nu}\frac{\partial \gamma_{\alpha\beta}}{\partial x_\mu}\right)$$

$$-\frac{\partial g_{\mu\nu}}{\partial x_\alpha}\frac{\partial \gamma_{\alpha\beta}}{\partial x_\mu}\frac{\partial \gamma_{\mu\nu}}{\partial x_\beta}$$

[p. 14 L]



$$\begin{bmatrix}\mu\nu\\ l\end{bmatrix} = \frac{1}{2}\left(\frac{\partial g_{\mu l}}{\partial x_\nu} + \frac{\partial g_{\nu l}}{\partial x_\mu} - \frac{\partial g_{\mu\nu}}{\partial x_l}\right) \qquad \frac{\partial}{\partial x_k}\begin{bmatrix}i\,l\\ m\end{bmatrix} - \frac{\partial}{\partial x_i}\begin{bmatrix}k\,l\\ m\end{bmatrix}$$

$$(ik, lm) = \frac{1}{2}\left(\frac{\partial^2 g_{im}}{\partial x_k \partial x_l} + \frac{\partial^2 g_{kl}}{\partial x_i \partial x_m} - \frac{\partial^2 g_{il}}{\partial x_k \partial x_m} - \frac{\partial^2 g_{km}}{\partial x_i \partial x_l}\right)$$

$$+ \sum_{\rho\sigma} \gamma_{\rho\sigma}\left(\begin{bmatrix}im\\ \sigma\end{bmatrix}\begin{bmatrix}kl\\ \rho\end{bmatrix} - \begin{bmatrix}il\\ \sigma\end{bmatrix}\begin{bmatrix}km\\ \rho\end{bmatrix}\right)$$

Grossmann Tensor vierter Mannigfaltigkeit

$$\sum \gamma_{kl}(ik, lm)$$

$$\sum \gamma_{kl}\begin{bmatrix}kl\\ \rho\end{bmatrix} = \sum \gamma_{kl}\left[\frac{\partial g_{k\rho}}{\partial x_l} + \frac{\partial g_{l\rho}}{\partial x_k} - \frac{\partial g_{kl}}{\partial x_\rho}\right]$$

$$= -\frac{\partial \lg G}{\partial x_\rho} + 2\sum_{kl} \gamma_{kl}\frac{\partial g_{k\rho}}{\partial x_l}$$

$$\frac{1}{4}\sum \gamma_{\rho\sigma}\left(\frac{\partial g_{i\sigma}}{\partial x_m} + \frac{\partial g_{m\sigma}}{\partial x_i} - \frac{\partial g_{im}}{\partial x_\sigma}\right)\left[-\frac{\partial \lg G}{\partial x_\rho} + 2\sum_{kl}\gamma_{kl}\frac{\partial g_{k\rho}}{\partial x_l}\right]$$

$$\sum \gamma_{kl}\gamma_{\rho\sigma}\left(\begin{bmatrix}im\\ \sigma\end{bmatrix}\begin{bmatrix}kl\\ \rho\end{bmatrix} - \begin{bmatrix}il\\ \sigma\end{bmatrix}\begin{bmatrix}km\\ \rho\end{bmatrix}\right)$$

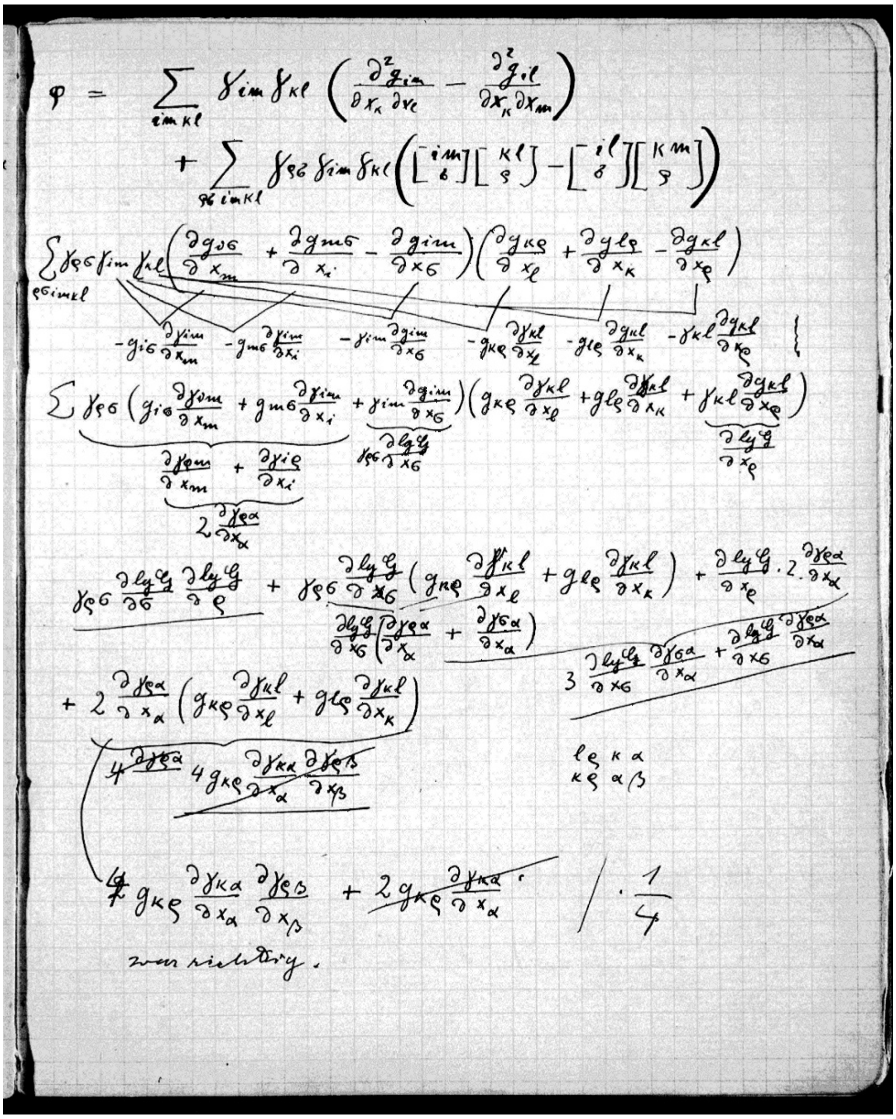
$$= \sum_\rho -\left\{\begin{matrix}im\\ \rho\end{matrix}\right\}\cdot\frac{\partial \lg G}{\partial x_\rho} + 2\sum_{kl\rho}\left\{\begin{matrix}im\\ \rho\end{matrix}\right\}\cdot\gamma_{kl}\frac{\partial g_{k\rho}}{\partial x_l} - \sum_{\rho lk}\left\{\begin{matrix}il\\ \rho\end{matrix}\right\}\left(\frac{\partial g_{k\rho}}{\partial x_m}\right)\gamma_{kl}$$

$$+ \sum_{\rho l}\left\{\begin{matrix}il\\ \rho\end{matrix}\right\}\cdot\left\{\begin{matrix}\rho m\\ l\end{matrix}\right\}$$

$$\sum_k\left(\frac{\partial^2 g_{kk}}{\partial x_i \partial x_m} - \frac{\partial^2 g_{ik}}{\partial x_k \partial x_m} - \frac{\partial^2 g_{mk}}{\partial x_k \partial x_i}\right) = 0$$

Sollte verschwinden.

[p. 14 R]



$$\varphi = \sum_{imkl} \gamma_{im}\gamma_{kl}\left(\frac{\partial^2 g_{im}}{\partial x_k \partial x_l} - \frac{\partial^2 g_{il}}{\partial x_k \partial x_m}\right) + \sum_{\rho\sigma imkl} \gamma_{\rho\sigma}\gamma_{im}\gamma_{kl}\left(\begin{bmatrix} im \\ \sigma \end{bmatrix}\begin{bmatrix} kl \\ \rho \end{bmatrix} - \begin{bmatrix} il \\ \sigma \end{bmatrix}\begin{bmatrix} km \\ \rho \end{bmatrix}\right)$$

$$\sum_{\rho\sigma imkl} \gamma_{\rho\sigma}\gamma_{im}\gamma_{kl}\left(\frac{\partial g_{i\sigma}}{\partial x_m} + \frac{\partial g_{m\sigma}}{\partial x_i} - \frac{\partial g_{im}}{\partial x_\sigma}\right)\left(\frac{\partial g_{k\rho}}{\partial x_l} + \frac{\partial g_{l\rho}}{\partial x_k} - \frac{\partial g_{kl}}{\partial x_\rho}\right)$$

$$-g_{i\sigma}\frac{\partial \gamma_{im}}{\partial x_m} \quad -g_{m\sigma}\frac{\partial \gamma_{im}}{\partial x_i} \quad -\gamma_{im}\frac{\partial g_{im}}{\partial x_\sigma} \quad -g_{k\rho}\frac{\partial \gamma_{kl}}{\partial x_l} \quad -g_{l\rho}\frac{\partial \gamma_{kl}}{\partial x_k} \quad -\gamma_{kl}\frac{\partial g_{kl}}{\partial x_\rho}$$

$$\sum \gamma_{\rho\sigma}\Bigg(\underbrace{g_{i\sigma}\frac{\partial \gamma_{im}}{\partial x_m} + g_{m\sigma}\frac{\partial \gamma_{im}}{\partial x_i}}_{\underbrace{\frac{\partial \gamma_{\sigma m}}{\partial x_m} + \frac{\partial \gamma_{i\sigma}}{\partial x_i}}_{2\frac{\partial \gamma_{\sigma\alpha}}{\partial x_\alpha}}} + \underbrace{\gamma_{im}\frac{\partial g_{im}}{\partial x_\sigma}}_{\gamma_{\rho\sigma}\frac{\partial \lg G}{\partial x_\sigma}}\Bigg)\left(g_{k\rho}\frac{\partial \gamma_{kl}}{\partial x_l} + g_{l\rho}\frac{\partial \gamma_{kl}}{\partial x_k} + \underbrace{\gamma_{kl}\frac{\partial g_{kl}}{\partial x_\rho}}_{\frac{\partial \lg G}{\partial x_\rho}}\right)$$

$$\gamma_{\rho\sigma}\frac{\partial \lg G}{\partial \sigma}\frac{\partial \lg G}{\partial \rho} + \gamma_{\rho\sigma}\frac{\partial \lg G}{\partial x_\sigma}\left(g_{k\rho}\frac{\partial \gamma_{kl}}{\partial x_l} + g_{l\rho}\frac{\partial \gamma_{kl}}{\partial x_k}\right) + \frac{\partial \lg G}{\partial x_\rho}\cdot 2\cdot\frac{\partial \gamma_{\rho\alpha}}{\partial x_\alpha}$$

$$\frac{\partial \lg G}{\partial x_\sigma}\left(\frac{\partial \gamma_{\rho\alpha}}{\partial x_\alpha} + \frac{\partial \gamma_{\sigma\alpha}}{\partial x_\alpha}\right)$$

$$3\frac{\partial \lg G}{\partial x_\sigma}\frac{\partial \gamma_{\sigma\alpha}}{\partial x_\alpha} + \frac{\partial \lg G}{\partial x_\sigma}\frac{\partial \gamma_{\sigma\alpha}}{\partial x_\alpha}$$

$$+\,2\frac{\partial \gamma_{\rho\alpha}}{\partial x_\alpha}\left(g_{k\rho}\frac{\partial \gamma_{kl}}{\partial x_l} + g_{l\rho}\frac{\partial \gamma_{kl}}{\partial x_k}\right)$$

$$\cancel{4\frac{\partial \gamma_{\rho\alpha}}{} \quad 4 g_{k\rho}\frac{\partial \gamma_{k\alpha}}{\partial x_\alpha}\frac{\partial \gamma_{\rho\beta}}{\partial x_\beta}}$$

$l\rho \quad k\alpha$
$k\rho \quad \alpha\beta$

$$4\, g_{k\rho}\frac{\partial \gamma_{k\alpha}}{\partial x_\alpha}\frac{\partial \gamma_{\rho\beta}}{\partial x_\beta} + \cancel{2 g_{k\rho}\frac{\partial \gamma_{k\alpha}}{\partial x_\alpha}} \qquad \Big/ \cdot \frac{1}{4}$$

war richtig.

[p. 14 L]

$$\begin{bmatrix}\mu\nu\\ l\end{bmatrix} = \frac{1}{2}\left(\frac{\partial g_{\mu l}}{\partial x_\nu} + \frac{\partial g_{\nu l}}{\partial x_\mu} - \frac{\partial g_{\mu\nu}}{\partial x_l}\right) \qquad \frac{\partial}{\partial x_\kappa}\begin{bmatrix}i & l\\ & m\end{bmatrix} - \frac{\partial}{\partial x_i}\begin{bmatrix}\kappa & l\\ & m\end{bmatrix}$$

$$\left.\begin{aligned}(i\kappa, lm) &= \frac{1}{2}\left(\frac{\partial^2 g_{im}}{\partial x_\kappa \partial \kappa_l} + \frac{\partial^2 g_{\kappa l}}{\partial x_i \partial x_m} - \frac{\partial^2 g_{il}}{\partial x_\kappa \partial x_m} - \frac{\partial^2 g_{\kappa m}}{\partial x_i \partial x_l}\right)\\ &\quad + \sum_{\rho\sigma}\gamma_{\rho\sigma}\left(\begin{bmatrix}i & m\\ \sigma\end{bmatrix}\begin{bmatrix}\kappa & l\\ \rho\end{bmatrix} - \begin{bmatrix}i & l\\ \sigma\end{bmatrix}\begin{bmatrix}\kappa & m\\ \rho\end{bmatrix}\right)\end{aligned}\right\}$$

Grossmann
Tensor vierter
Mannigfaltigkeit

$$\sum \gamma_{\kappa l}(i\kappa, lm) \quad ?$$

$$\sum \gamma_{\kappa l}\begin{bmatrix}\kappa & l\\ \rho\end{bmatrix} = \sum \gamma_{\kappa l}\left[\frac{\partial g_{\kappa\rho}}{\partial x_l} + \frac{\partial g_{l\rho}}{\partial x_\kappa} - \frac{\partial g_{\kappa l}}{\partial x_\rho}\right]$$

$$= \langle 2 \rangle - \frac{\partial \lg G}{\partial x_\rho} + 2\sum_{\kappa l}\gamma_{\kappa l}\frac{\partial g_{\kappa\rho}}{\partial x_l}$$

$$\frac{1}{4}\sum \gamma_{\rho\sigma}\left(\frac{\partial g_{i\sigma}}{\partial x_m} + \frac{\partial g_{m\sigma}}{\partial x_i} - \frac{\partial g_{im}}{\partial x_\sigma}\right)\left[-\frac{\partial \lg G}{\partial x_\rho} + 2\sum_{\kappa l}\gamma_{\kappa l}\frac{\partial g_{\kappa\rho}}{\partial x_l}\right]$$

$$\sum \gamma_{\kappa l}\gamma_{\rho\sigma}\left(\begin{bmatrix}i & m\\ \sigma\end{bmatrix}\begin{bmatrix}\kappa & l\\ \rho\end{bmatrix} - \begin{bmatrix}i & l\\ \sigma\end{bmatrix}\begin{bmatrix}\kappa & m\\ \rho\end{bmatrix}\right)$$

$$= \sum_\rho - \begin{Bmatrix}i & m\\ \rho\end{Bmatrix} \cdot \frac{\partial \lg G}{\partial x_\rho} + 2\sum_{\kappa l \rho}\begin{Bmatrix}i & m\\ \rho\end{Bmatrix} \cdot \gamma_{\kappa l}\frac{\partial g_{\kappa\rho}}{\partial x_l} - \sum_{\rho l \kappa}\begin{Bmatrix}i & l\\ \rho\end{Bmatrix}\left(\frac{\partial g_{\kappa\rho}}{\partial x_m}\right)\gamma_{\kappa l}$$

$$+ \sum_{\rho l}\begin{Bmatrix}i & l\\ \rho\end{Bmatrix} \cdot \begin{Bmatrix}\rho & m\\ l\end{Bmatrix}$$

$$\sum_\kappa\left(\frac{\partial^2 g_{\kappa\kappa}}{\partial x_i \partial x_m} - \frac{\partial^2 g_{i\kappa}}{\partial x_\kappa \partial x_m} - \frac{\partial^2 g_{m\kappa}}{\partial x_\kappa \partial x_i}\right) = 0$$

Sollte verschwinden.

[p. 14 R]

$$\varphi = \sum_{im\kappa l} \gamma_{im}\gamma_{\kappa l}\left(\frac{\partial^2 g_{im}}{\partial x_\kappa \partial x_l} - \frac{\partial^2 g_{il}}{\partial x_\kappa \partial x_m}\right) + \sum_{\rho\sigma im\kappa l} \gamma_{\rho\sigma}\gamma_{im}\gamma_{\kappa l}\left(\begin{bmatrix} i\ m \\ \sigma \end{bmatrix}\begin{bmatrix} \kappa\ l \\ \rho \end{bmatrix} - \begin{bmatrix} i\ l \\ \sigma \end{bmatrix}\begin{bmatrix} \kappa\ m \\ \rho \end{bmatrix}\right)$$

$$\sum_{\rho\sigma im\kappa l} \gamma_{\rho\sigma}\gamma_{im}\gamma_{\kappa l}\left(\frac{\partial g_{i\sigma}}{\partial x_m} + \frac{\partial g_{m\sigma}}{\partial x_i} - \frac{\partial g_{im}}{\partial x_\sigma}\right)\left(\frac{\partial g_{\kappa\rho}}{\partial x_l} + \frac{\partial g_{l\rho}}{\partial x_\kappa} - \frac{\partial g_{\kappa l}}{\partial x_\rho}\right)$$

$$-g_{i\sigma}\frac{\partial \gamma_{im}}{\partial x_m} \quad -g_{m\sigma}\frac{\partial \gamma_{im}}{\partial x_i} \quad -\gamma_{im}\frac{\partial g_{im}}{\partial x_\sigma} \quad -g_{\kappa\rho}\frac{\partial \gamma_{\kappa l}}{\partial x_l} \quad -g_{l\rho}\frac{\partial g_{\kappa l}}{\partial x_\kappa} \quad -\gamma_{\kappa l}\frac{\partial g_{\kappa l}}{\partial x_\rho}$$

$$\sum \gamma_{\rho\sigma}\underbrace{\left(g_{i\sigma}\frac{\partial \gamma_{im}}{\partial x_m} + g_{m\sigma}\frac{\partial \gamma_{im}}{\partial x_i}\right.}_{\underbrace{\frac{\partial \gamma_{\rho m}}{\partial x_m} + \frac{\partial \gamma_{i\rho}}{\partial x_i}}_{2\frac{\partial \gamma_{\rho\alpha}}{\partial x_\alpha}}} + \underbrace{\left.\gamma_{im}\frac{\partial g_{im}}{\partial x_\sigma}\right)}_{\gamma_{\rho\sigma}\frac{\partial \lg G}{\partial x_\sigma}}\left(g_{\kappa\rho}\frac{\partial \gamma_{\kappa l}}{\partial x_l} + g_{l\rho}\frac{\partial \gamma_{\kappa l}}{\partial x_\kappa} + \underbrace{\gamma_{\kappa l}\frac{\partial g_{\kappa l}}{\partial x_\rho}\right)}_{\frac{\partial \lg G}{\partial x_\rho}}$$

$$\underline{\gamma_{\rho\sigma}\frac{\partial \lg G}{\partial \sigma}\frac{\partial \lg G}{\partial \rho}} + \gamma_{\rho\sigma}\frac{\partial \lg G}{\partial x_\sigma}\left(g_{\kappa\rho}\frac{\partial \gamma_{\kappa l}}{\partial x_l} + g_{l\rho}\frac{\partial \gamma_{\kappa l}}{\partial x_\kappa}\right) + \frac{\partial \lg G}{\partial x_\rho} \cdot 2\frac{\partial \gamma_{\rho\alpha}}{\partial x_\alpha}$$

$$\frac{\partial \lg G}{\partial x_\sigma}\left(\frac{\partial \gamma_{\rho\alpha}}{\partial x_\alpha} + \frac{\partial \gamma_{\sigma\alpha}}{\partial x_\alpha}\right)$$

$$\underline{3\frac{\partial \lg G}{\partial x_\sigma}\frac{\partial \gamma_{\sigma\alpha}}{\partial x_\alpha} + \frac{\partial \lg G}{\partial x_\sigma}\frac{\partial \gamma_{\rho\alpha}}{\partial x_\alpha}}$$

$$\underbrace{+ 2\frac{\partial \gamma_{\rho\alpha}}{\partial x_\alpha}\left(g_{\kappa\rho}\frac{\partial \gamma_{\kappa l}}{\partial x_l} + g_{l\rho}\frac{\partial \gamma_{\kappa l}}{\partial x_\kappa}\right)}$$

$$\cancel{4\frac{\partial \gamma_{\rho\alpha}}{}} \qquad \cancel{4g_{\kappa\rho}\frac{\partial \gamma_{\kappa\alpha}}{\partial x_\alpha}\frac{\partial \gamma_{\rho\beta}}{\partial x_\beta}}$$

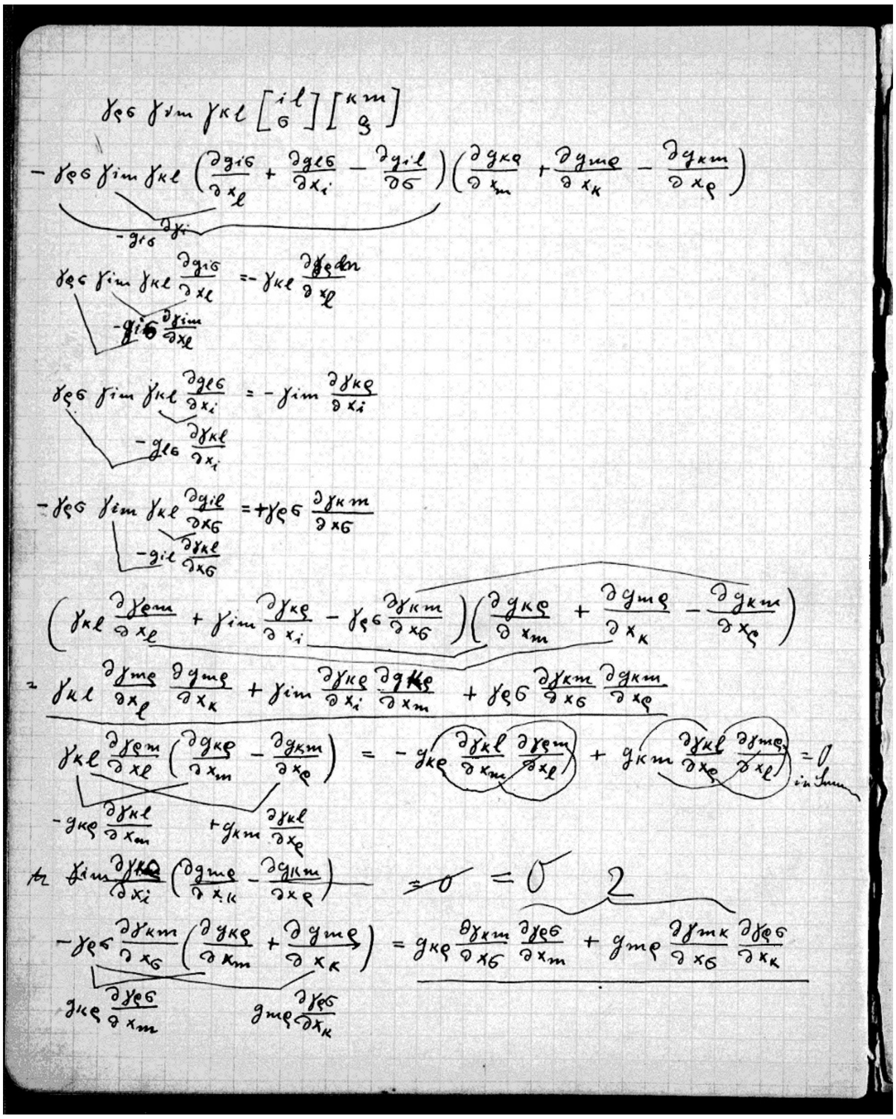
$l\ \rho\ \kappa\ \alpha$

$\kappa\ \rho\ \alpha\ \beta$

$$4\langle 2\rangle g_{\kappa\rho}\frac{\partial \gamma_{\kappa\alpha}}{\partial x_\alpha}\frac{\partial \gamma_{\rho\beta}}{\partial x_\beta} + 2\cancel{g_{\kappa\rho}}\frac{\partial \cancel{\gamma_{\kappa\alpha}}}{\partial x_\alpha} \;\Big|\; \cdot \frac{1}{4}$$

war richtig.

[p. 15 L]



$$\gamma_{\varrho\sigma}\gamma_{im}\gamma_{\kappa l}\begin{bmatrix} il \\ \sigma\end{bmatrix}\begin{bmatrix}\kappa m\\ \varrho\end{bmatrix}$$

$$-\underbrace{\gamma_{\varrho\sigma}\gamma_{im}\gamma_{\kappa l}\left(\frac{\partial g_{i\sigma}}{\partial x_l}+\frac{\partial g_{l\sigma}}{\partial x_i}-\frac{\partial g_{il}}{\partial \sigma}\right)}_{-g_{i\sigma}\partial\gamma_i}\left(\frac{\partial g_{\kappa\varrho}}{\partial x_m}+\frac{\partial g_{m\varrho}}{\partial x_\kappa}-\frac{\partial g_{\kappa m}}{\partial x_\varrho}\right)$$

$$\gamma_{\varrho\sigma}\gamma_{im}\gamma_{\kappa l}\frac{\partial g_{i\sigma}}{\partial x_l} = -\gamma_{\kappa l}\frac{\partial\gamma_{\varrho m}}{\partial x_l}$$

$$-g_{i\sigma}\frac{\partial\gamma_{im}}{\partial x_l}$$

$$\gamma_{\varrho\sigma}\gamma_{im}\gamma_{\kappa l}\frac{\partial g_{l\sigma}}{\partial x_i} = -\gamma_{im}\frac{\partial\gamma_{\kappa\varrho}}{\partial x_i}$$

$$-g_{l\sigma}\frac{\partial\gamma_{\kappa l}}{\partial x_i}$$

$$-\gamma_{\varrho\sigma}\gamma_{im}\gamma_{\kappa l}\frac{\partial g_{il}}{\partial x_\sigma} = +\gamma_{\varrho\sigma}\frac{\partial\gamma_{\kappa m}}{\partial x_\sigma}$$

$$-g_{il}\frac{\partial\gamma_{\kappa l}}{\partial x_\sigma}$$

$$\left(\gamma_{\kappa l}\frac{\partial\gamma_{\varrho m}}{\partial x_l}+\gamma_{im}\frac{\partial\gamma_{\kappa\varrho}}{\partial x_i}-\gamma_{\varrho\sigma}\frac{\partial\gamma_{\kappa m}}{\partial x_\sigma}\right)\left(\frac{\partial g_{\kappa\varrho}}{\partial x_m}+\frac{\partial g_{m\varrho}}{\partial x_\kappa}-\frac{\partial g_{\kappa m}}{\partial x_\varrho}\right)$$

$$= \gamma_{\kappa l}\frac{\partial\gamma_{m\varrho}}{\partial x_l}\frac{\partial g_{m\varrho}}{\partial x_\kappa}+\gamma_{im}\frac{\partial\gamma_{\kappa\varrho}}{\partial x_i}\frac{\partial g_{\kappa\varrho}}{\partial x_m}+\gamma_{\varrho\sigma}\frac{\partial\gamma_{\kappa m}}{\partial x_\sigma}\frac{\partial g_{\kappa m}}{\partial x_\varrho}$$

$$\gamma_{\kappa l}\frac{\partial\gamma_{\varrho m}}{\partial x_l}\left(\frac{\partial g_{\kappa\varrho}}{\partial x_m}-\frac{\partial g_{\kappa m}}{\partial x_\varrho}\right) = -g_{\kappa\varrho}\frac{\partial\gamma_{\kappa l}}{\partial x_m}\frac{\partial\gamma_{\varrho m}}{\partial x_l}+g_{\kappa m}\frac{\partial\gamma_{\kappa l}}{\partial x_\varrho}\frac{\partial\gamma_{m\varrho}}{\partial x_l} = 0 \text{ in Summe}$$

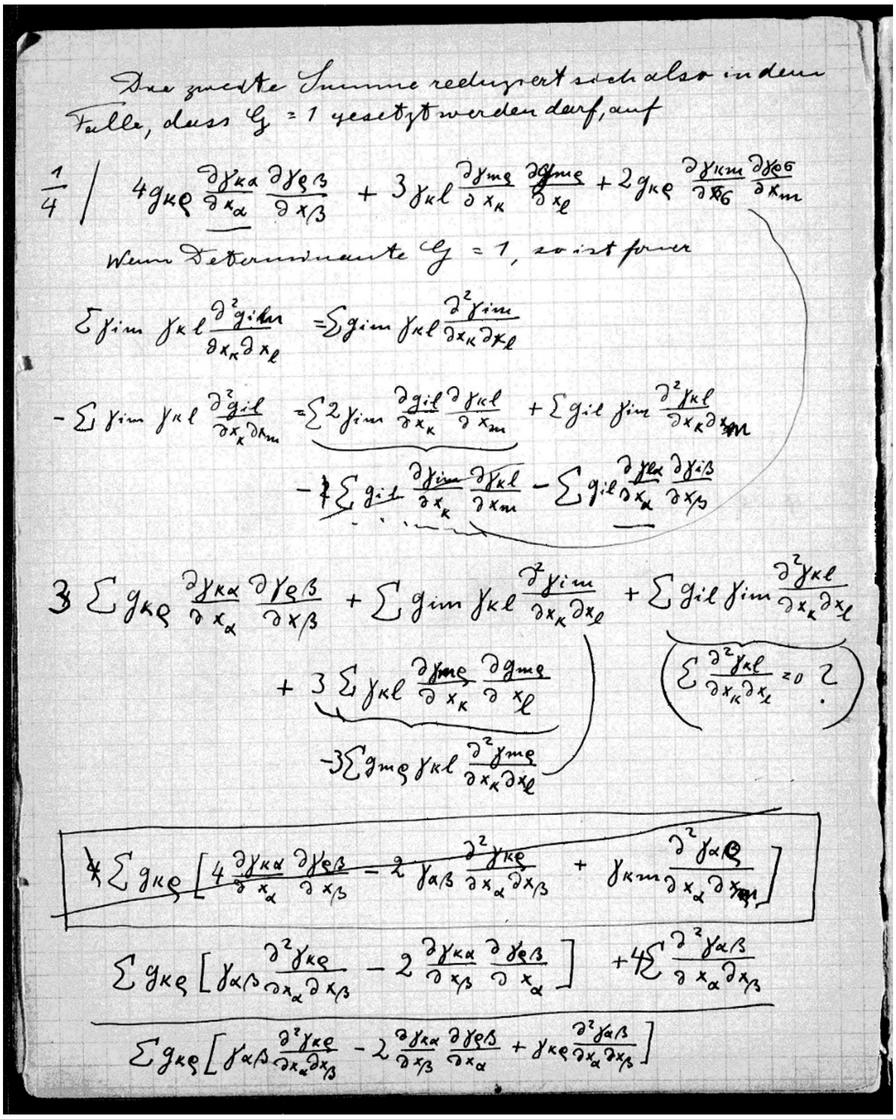
$$-g_{\kappa\varrho}\frac{\partial\gamma_{\kappa l}}{\partial x_m}\qquad +g_{\kappa m}\frac{\partial\gamma_{\kappa l}}{\partial x_\varrho}$$

ebenso $\gamma_{im}\frac{\partial\gamma_{\kappa\varrho}}{\partial x_i}\left(\frac{\partial g_{m\varrho}}{\partial x_\kappa}-\frac{\partial g_{\kappa m}}{\partial x_\varrho}\right) \cancel{=0} = 0 \quad 2$

$$-\gamma_{\varrho\sigma}\frac{\partial\gamma_{\kappa m}}{\partial x_\sigma}\left(\frac{\partial g_{\kappa\varrho}}{\partial x_m}+\frac{\partial g_{m\varrho}}{\partial x_\kappa}\right) = g_{\kappa\varrho}\frac{\partial\gamma_{\kappa m}}{\partial x_\sigma}\frac{\partial\gamma_{\varrho\sigma}}{\partial x_m}+g_{m\varrho}\frac{\partial\gamma_{m\kappa}}{\partial x_\sigma}\frac{\partial\gamma_{\varrho\sigma}}{\partial x_\kappa}$$

$$g_{\kappa\varrho}\frac{\partial\gamma_{\varrho\sigma}}{\partial x_m}\qquad g_{m\varrho}\frac{\partial\gamma_{\varrho\sigma}}{\partial x_\kappa}$$

[p. 15 R]



Die zweite Summe reduziert sich also in dem Falle, dass $G = 1$ gesetzt werden darf, auf

$$\frac{1}{4} \Big| \; 4 g_{\kappa\varrho} \frac{\partial \gamma_{\kappa\alpha}}{\partial x_\alpha} \frac{\partial \gamma_{\varrho\beta}}{\partial x_\beta} + 3 \gamma_{\kappa l} \frac{\partial \gamma_{m\varrho}}{\partial x_\kappa} \frac{\partial g_{m\varrho}}{\partial x_l} + 2 g_{\kappa\varrho} \frac{\partial \gamma_{\kappa m}}{\partial x_\sigma} \frac{\partial \gamma_{\varrho\sigma}}{\partial x_m}$$

Wenn Determinante $G = 1$, so ist ferner

$$\sum \gamma_{im} \gamma_{\kappa l} \frac{\partial^2 g_{im}}{\partial x_\kappa \partial x_l} = \sum g_{im} \gamma_{\kappa l} \frac{\partial^2 \gamma_{im}}{\partial x_\kappa \partial x_l}$$

$$- \sum \gamma_{im} \gamma_{\kappa l} \frac{\partial^2 g_{il}}{\partial x_\kappa \partial x_m} = \underbrace{\sum 2 \gamma_{im} \frac{\partial g_{il}}{\partial x_\kappa} \frac{\partial \gamma_{\kappa l}}{\partial x_m}}_{} + \sum g_{il} \gamma_{im} \frac{\partial^2 \gamma_{\kappa l}}{\partial x_\kappa \partial x_m}$$

$$- \cancel{2} \sum g_{il} \frac{\partial \gamma_{im}}{\partial x_\kappa} \frac{\partial \gamma_{\kappa l}}{\partial x_m} - \sum g_{il} \frac{\partial \gamma_{l\alpha}}{\partial x_\alpha} \frac{\partial \gamma_{i\beta}}{\partial x_\beta}$$

$$3 \sum g_{\kappa\varrho} \frac{\partial \gamma_{\kappa\alpha}}{\partial x_\alpha} \frac{\partial \gamma_{\varrho\beta}}{\partial x_\beta} + \sum g_{im} \gamma_{\kappa l} \frac{\partial^2 \gamma_{im}}{\partial x_\kappa \partial x_l} + \underbrace{\sum g_{il} \gamma_{im} \frac{\partial^2 \gamma_{\kappa l}}{\partial x_\kappa \partial x_l}}_{}$$

$$\left(\sum \frac{\partial^2 \gamma_{\kappa l}}{\partial x_\kappa \partial x_l} = 0 \; ? \right)$$

$$+ \underbrace{3 \sum \gamma_{\kappa l} \frac{\partial \gamma_{m\varrho}}{\partial x_\kappa} \frac{\partial g_{m\varrho}}{\partial x_l}}_{}$$

$$-3 \sum g_{m\varrho} \gamma_{\kappa l} \frac{\partial^2 \gamma_{m\varrho}}{\partial x_\kappa \partial x_l}$$

$$\boxed{\cancel{\tfrac{1}{4} \sum g_{\kappa\varrho} \left[4 \frac{\partial \gamma_{\kappa\alpha}}{\partial x_\alpha} \frac{\partial \gamma_{\varrho\beta}}{\partial x_\beta} - 2 \gamma_{\alpha\beta} \frac{\partial^2 \gamma_{\kappa\varrho}}{\partial x_\alpha \partial x_\beta} + \gamma_{\kappa m} \frac{\partial^2 \gamma_{\alpha\varrho}}{\partial x_\alpha \partial x_m} \right]}}$$

$$\sum g_{\kappa\varrho} \left[\gamma_{\alpha\beta} \frac{\partial^2 \gamma_{\kappa\varrho}}{\partial x_\alpha \partial x_\beta} - 2 \frac{\partial \gamma_{\kappa\alpha}}{\partial x_\beta} \frac{\partial \gamma_{\varrho\beta}}{\partial x_\alpha} \right] + 4 \sum \frac{\partial^2 \gamma_{\alpha\beta}}{\partial x_\alpha \partial x_\beta}$$

$$\sum g_{\kappa\varrho} \left[\gamma_{\alpha\beta} \frac{\partial^2 \gamma_{\kappa\varrho}}{\partial x_\alpha \partial x_\beta} - 2 \frac{\partial \gamma_{\kappa\alpha}}{\partial x_\beta} \frac{\partial \gamma_{\varrho\beta}}{\partial x_\alpha} + \gamma_{\kappa\varrho} \frac{\partial^2 \gamma_{\alpha\beta}}{\partial x_\alpha \partial x_\beta} \right]$$

[p. 15 L]

$$\gamma_{\rho\sigma}\gamma_{im}\gamma_{\kappa l}\begin{bmatrix} i\ l \\ \sigma \end{bmatrix}\begin{bmatrix} \kappa\ m \\ \rho \end{bmatrix}$$

$$-\gamma_{\rho\sigma}\gamma_{im}\gamma_{\kappa l}\left(\frac{\partial g_{i\sigma}}{\partial x_l}+\frac{\partial g_{l\sigma}}{\partial x_i}-\frac{\partial g_{il}}{\partial \sigma}\right)\left(\frac{\partial g_{\kappa\rho}}{\partial x_m}+\frac{\partial g_{m\rho}}{\partial x_\kappa}-\frac{\partial g_{\kappa m}}{\partial x_\rho}\right)$$

$$\partial\gamma_i$$

$$-g_{i\sigma}$$

$$\gamma_{\rho\sigma}\gamma_{im}\gamma_{\kappa l}\frac{\partial g_{i\sigma}}{\partial x_l} = -\gamma_{\kappa l}\frac{\partial \gamma_{\rho m}}{\partial x_l}$$

$$-g_{i\sigma}\frac{\partial \gamma_{im}}{\partial x_l}$$

$$\gamma_{\rho\sigma}\gamma_{im}\gamma_{\kappa l}\frac{\partial g_{l\sigma}}{\partial x_i} = -\gamma_{im}\frac{\partial \gamma_{\kappa\rho}}{\partial x_i}$$

$$-g_{l\sigma}\frac{\partial \gamma_{\kappa l}}{\partial x_i}$$

$$-\gamma_{\rho\sigma}\gamma_{im}\gamma_{\kappa l}\frac{\partial g_{il}}{\partial x_\sigma} = +\gamma_{\rho\sigma}\frac{\partial \gamma_{\kappa m}}{\partial x_\sigma}$$

$$-g_{il}\frac{\partial \gamma_{\kappa l}}{\partial x_\sigma}$$

$$\left(\gamma_{\kappa l}\frac{\partial \gamma_{\rho m}}{\partial x_l}+\gamma_{im}\frac{\partial \gamma_{\kappa\rho}}{\partial x_i}-\gamma_{\rho\sigma}\frac{\partial \gamma_{\kappa m}}{\partial x_\sigma}\right)\left(\frac{\partial g_{\kappa\rho}}{\partial x_m}+\frac{\partial g_{m\rho}}{\partial x_\kappa}-\frac{\partial g_{\kappa m}}{\partial x_\rho}\right)$$

$$\underline{= \gamma_{\kappa l}\frac{\partial \gamma_{m\rho}}{\partial x_l}\frac{\partial g_{m\rho}}{\partial x_\kappa}+\gamma_{im}\frac{\partial \gamma_{\kappa\rho}}{\partial x_i}\frac{\partial g_{\kappa\rho}}{\partial x_m}+\gamma_{\rho\sigma}\frac{\partial \gamma_{\kappa m}}{\partial x_\sigma}\frac{\partial g_{\kappa m}}{\partial x_\rho}}$$

$$\gamma_{\kappa l}\frac{\partial \gamma_{\rho m}}{\partial x_l}\left(\frac{\partial g_{\kappa\rho}}{\partial x_m}-\frac{\partial g_{\kappa m}}{\partial x_\rho}\right) = -g_{\kappa\rho}\left(\frac{\partial \gamma_{\kappa l}}{\partial x_m}\frac{\partial \gamma_{\rho m}}{\partial x_l}\right)+g_{\kappa m}\left(\frac{\partial \gamma_{\kappa l}}{\partial x_\rho}\frac{\partial \gamma_{m\rho}}{\partial x_l}\right) = 0$$

in Summe

$$-g_{\kappa\rho}\frac{\partial \gamma_{\kappa l}}{\partial x_m} \qquad +g_{\kappa m}\frac{\partial \gamma_{\kappa l}}{\partial x_\rho}$$

$$\gamma_{im}\frac{\partial \gamma_{\kappa\rho}}{\partial x_i}\left(\frac{\partial g_{m\rho}}{\partial x_\kappa}-\frac{\partial g_{\kappa m}}{\partial x_\rho}\right)\langle = 0\rangle = 0$$

$$-\gamma_{\rho\sigma}\frac{\partial \gamma_{\kappa m}}{\partial x_\sigma}\left(\frac{\partial g_{\kappa\rho}}{\partial x_m}+\frac{\partial g_{m\rho}}{\partial x_\kappa}\right) = \underline{g_{\kappa\rho}\frac{\partial \gamma_{\kappa m}}{\partial x_\sigma}\overbrace{\frac{\partial \gamma_{\rho\sigma}}{\partial x_m}+g_{m\rho}\frac{\partial \gamma_{m\kappa}}{\partial x_\sigma}}^{2}\frac{\partial \gamma_{\rho\sigma}}{\partial x_\kappa}}$$

$$g_{\kappa\rho}\frac{\partial \gamma_{\rho\sigma}}{\partial x_m} \qquad g_{m\rho}\frac{\partial \gamma_{\rho\sigma}}{\partial x_\kappa}$$

[p. 15 R]

Die zweite Summe reduziert sich also in dem Falle, dass G = 1 gesetzt werden darf, auf

$$\frac{1}{4}\ \Big|\ 4g_{\kappa\rho}\underline{\frac{\partial\gamma_{\kappa\alpha}}{\partial x_\alpha}}\frac{\partial\gamma_{\rho\beta}}{\partial x_\beta}+3\gamma_{\kappa l}\frac{\partial\gamma_{m\rho}}{\partial x_\kappa}\frac{\partial g_{m\rho}}{\partial x_l}+2g_{\kappa\rho}\frac{\partial\gamma_{\kappa m}}{\partial x_\sigma}\frac{\partial\gamma_{\rho\sigma}}{\partial x_m}$$

Wenn Determinante G = 1, so ist ferner

$$\sum\gamma_{im}\gamma_{\kappa l}\frac{\partial^2 g_{im}}{\partial x_\kappa\partial x_l}=\sum g_{im}\gamma_{\kappa l}\frac{\partial^2\gamma_{im}}{\partial x_\kappa\partial x_l}$$

$$-\sum\gamma_{im}\gamma_{\kappa l}\frac{\partial^2 g_{il}}{\partial x_\kappa\partial x_m}=\underbrace{\sum 2\gamma_{im}\frac{\partial g_{il}}{\partial x_\kappa}\frac{\partial\gamma_{\kappa l}}{\partial x_m}}+\sum g_{il}\gamma_{im}\frac{\partial^2\gamma_{\kappa l}}{\partial x_k\partial x_m}$$

$$-<2>\cancel{\sum g_{il}\frac{\partial\gamma_{im}}{\partial x_\kappa}\frac{\partial\gamma_{\kappa l}}{\partial x_m}}-\sum g_{il}\underline{\frac{\partial\gamma_{l\alpha}}{\partial x_\alpha}}\frac{\partial\gamma_{i\beta}}{\partial x_\beta}$$

$$3\sum g_{\kappa\rho}\frac{\partial\gamma_{\kappa\alpha}}{\partial x_\alpha}\frac{\partial\gamma_{\rho\beta}}{\partial x_\beta}+\sum g_{im}\gamma_{\kappa l}\frac{\partial^2\gamma_{im}}{\partial x_\kappa\partial x_l}+\underbrace{\sum g_{il}\gamma_{im}\frac{\partial^2\gamma_{\kappa l}}{\partial x_\kappa\partial x_l}}$$

$$\underbrace{+3\sum\gamma_{\kappa l}\frac{\partial\gamma_{m\rho}}{\partial x_\kappa}\frac{\partial g_{m\rho}}{\partial x_l}}\qquad\left(\sum\frac{\partial^2\gamma_{\kappa l}}{\partial x_\kappa\partial x_l}=0\quad ?\right)$$

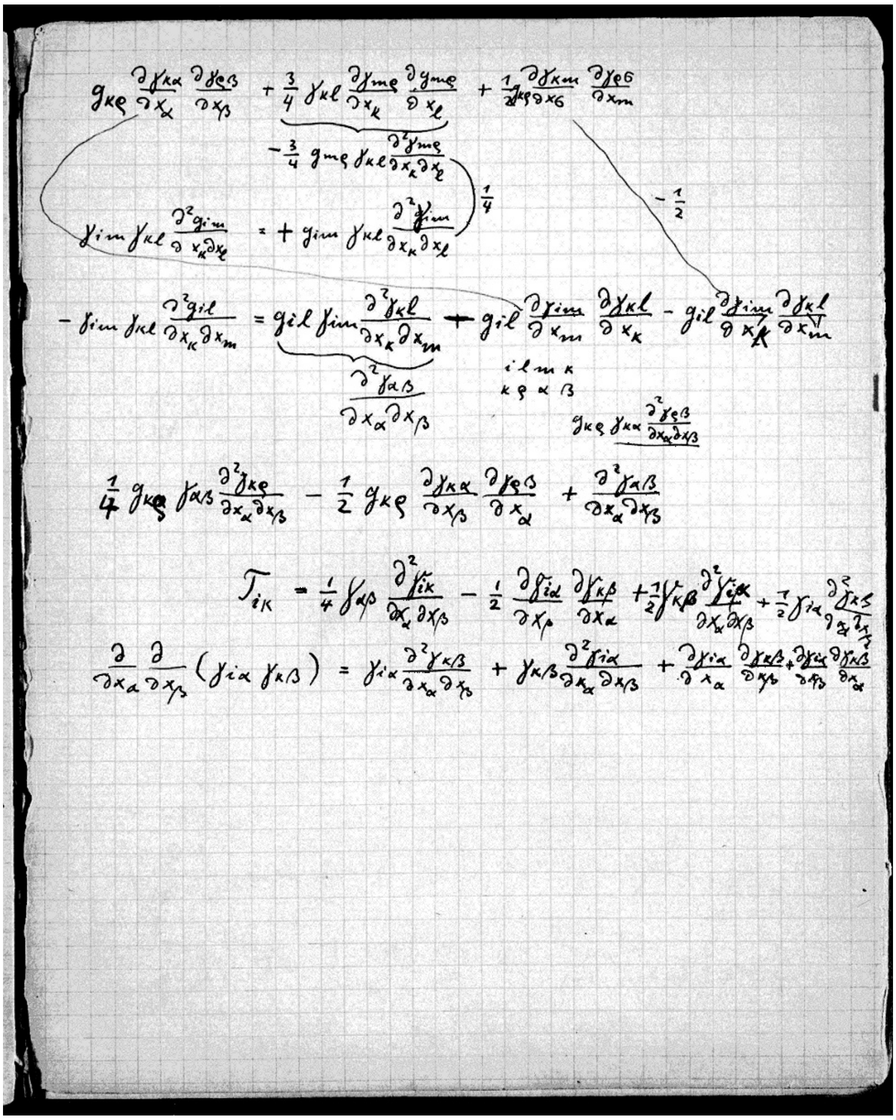
$$-3\sum g_{mp}\gamma_{\kappa l}\frac{\partial^2\gamma_{m\rho}}{\partial x_\kappa\partial x_l}$$

$$\boxed{\cancel{<4>\sum g_{\kappa\rho}\left[4\frac{\partial\gamma_{\kappa\alpha}}{\partial x_\alpha}\frac{\partial\gamma_{\rho\beta}}{\partial x_\beta}-2\gamma_{\alpha\beta}\frac{\partial^2\gamma_{\kappa\rho}}{\partial x_\alpha\partial x_\beta}+\gamma_{\kappa m}\frac{\partial^2\gamma_{\alpha\rho}}{\partial x_\alpha\partial x_m}\right]}}$$

$$\sum g_{\kappa\rho}\left[\gamma_{\alpha\beta}\frac{\partial^2\gamma_{\kappa\rho}}{\partial x_\alpha\partial x_\beta}-2\frac{\partial\gamma_{\kappa\alpha}}{\partial x_\beta}\frac{\partial\gamma_{\rho\beta}}{\partial x_\alpha}\right]+4\sum\frac{\partial^2\gamma_{\alpha\beta}}{\partial x_\alpha\partial x_\beta}$$

$$\sum g_{\kappa\rho}\left[\gamma_{\alpha\beta}\frac{\partial^2\gamma_{\kappa\rho}}{\partial x_\alpha\partial x_\beta}-2\frac{\partial\gamma_{\kappa\alpha}}{\partial x_\beta}\frac{\partial\gamma_{\rho\beta}}{\partial x_\alpha}+\gamma_{\kappa\rho}\frac{\partial^2\gamma_{\alpha\beta}}{\partial x_\alpha\partial x_\beta}\right]$$

[p. 16 L]



$$g_{\kappa\varrho}\frac{\partial\gamma_{\kappa\alpha}}{\partial x_\alpha}\frac{\partial\gamma_{\varrho\beta}}{\partial x_\beta} + \underbrace{\frac{3}{4}\gamma_{\kappa l}\frac{\partial\gamma_{m\varrho}}{\partial x_\kappa}\frac{\partial g_{m\varrho}}{\partial x_l}}_{} + \frac{1}{2}g_{\kappa\varrho}\frac{\partial\gamma_{\kappa m}}{\partial x_\sigma}\frac{\partial\gamma_{\varrho\sigma}}{\partial x_m}$$

$$-\frac{3}{4}g_{m\varrho}\gamma_{\kappa l}\frac{\partial^2\gamma_{m\varrho}}{\partial x_\kappa\partial x_l}$$

$$\frac{1}{4}$$

$$-\frac{1}{2}$$

$$\gamma_{im}\gamma_{\kappa l}\frac{\partial^2 g_{im}}{\partial x_\kappa\partial x_l} = +g_{im}\gamma_{\kappa l}\frac{\partial^2\gamma_{im}}{\partial x_\kappa\partial x_l}$$

$$-\gamma_{im}\gamma_{\kappa l}\frac{\partial^2 g_{il}}{\partial x_\kappa\partial x_m} = \underbrace{g_{il}\gamma_{im}\frac{\partial^2\gamma_{\kappa l}}{\partial x_\kappa\partial x_m}}_{\frac{\partial^2\gamma_{\alpha\beta}}{\partial x_\alpha\partial x_\beta}} + g_{il}\frac{\partial\gamma_{im}}{\partial x_m}\frac{\partial\gamma_{\kappa l}}{\partial x_\kappa} - g_{il}\frac{\partial\gamma_{im}}{\partial x_\kappa}\frac{\partial\gamma_{\kappa l}}{\partial x_m}$$

$i\ l\ m\ \kappa$

$\kappa\ \varrho\ \alpha\ \beta$

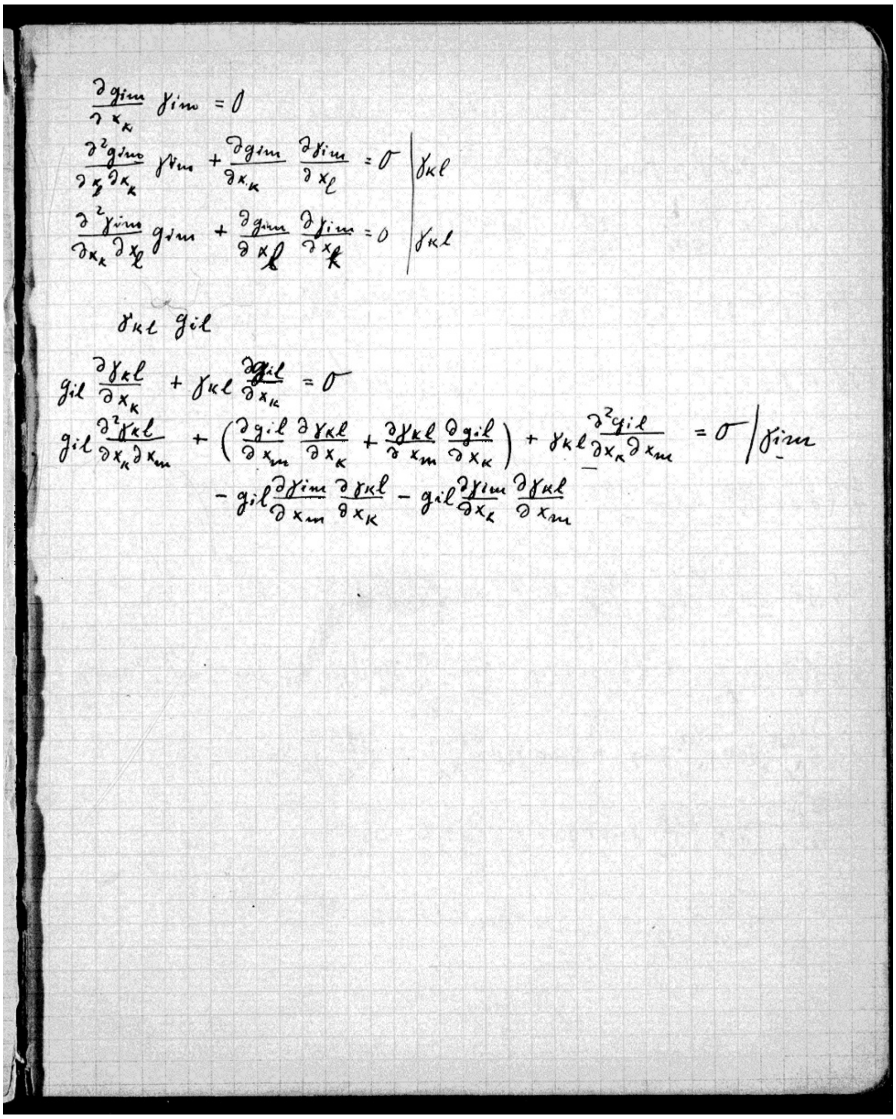
$$\underline{g_{\kappa\varrho}\gamma_{\kappa\alpha}\frac{\partial^2\gamma_{\varrho\beta}}{\partial x_\alpha\partial x_\beta}}$$

$$\frac{1}{4}g_{\kappa\varrho}\gamma_{\alpha\beta}\frac{\partial^2\gamma_{\kappa\varrho}}{\partial x_\alpha\partial x_\beta} - \frac{1}{2}g_{\kappa\varrho}\frac{\partial\gamma_{\kappa\alpha}}{\partial x_\beta}\frac{\partial\gamma_{\varrho\beta}}{\partial x_\alpha} + \frac{\partial^2\gamma_{\alpha\beta}}{\partial x_\alpha\partial x_\beta}$$

$$T_{i\kappa} = \frac{1}{4}\gamma_{\alpha\beta}\frac{\partial^2\gamma_{i\kappa}}{\partial x_\alpha\partial x_\beta} - \frac{1}{2}\frac{\partial\gamma_{i\alpha}}{\partial x_\beta}\frac{\partial\gamma_{\kappa\beta}}{\partial x_\alpha} + \frac{1}{2}\gamma_{\kappa\beta}\frac{\partial^2\gamma_{i\alpha}}{\partial x_\alpha\partial x_\beta} + \frac{1}{2}\gamma_{i\alpha}\frac{\partial^2\gamma_{\kappa\beta}}{\partial x_\alpha\partial x_\beta}$$

$$\frac{\partial}{\partial x_\alpha}\frac{\partial}{\partial x_\beta}(\gamma_{i\alpha}\gamma_{\kappa\beta}) = \gamma_{i\alpha}\frac{\partial^2\gamma_{\kappa\beta}}{\partial x_\alpha\partial x_\beta} + \gamma_{\kappa\beta}\frac{\partial^2\gamma_{i\alpha}}{\partial x_\alpha\partial x_\beta} + \frac{\partial\gamma_{i\alpha}}{\partial x_\alpha}\frac{\partial\gamma_{\kappa\beta}}{\partial x_\beta} + \frac{\partial\gamma_{i\alpha}}{\partial x_\beta}\frac{\partial\gamma_{\kappa\beta}}{\partial x_\alpha}$$

[p. 16 R]



$$\frac{\partial g_{im}}{\partial x_\kappa}\gamma_{im} = 0$$

$$\frac{\partial^2 g_{im}}{\partial x_l \partial x_\kappa}\gamma_{im} + \frac{\partial g_{im}}{\partial x_\kappa}\frac{\partial \gamma_{im}}{\partial x_l} = 0 \quad \Big| \gamma_{\kappa l}$$

$$\frac{\partial^2 \gamma_{im}}{\partial x_\kappa \partial x_l} g_{im} + \frac{\partial g_{im}}{\partial x_l}\frac{\partial \gamma_{im}}{\partial x_\kappa} = 0 \quad \Big| \gamma_{\kappa l}$$

$$\gamma_{\kappa l}\, g_{il}$$

$$g_{il}\frac{\partial \gamma_{\kappa l}}{\partial x_\kappa} + \gamma_{\kappa l}\frac{\partial g_{il}}{\partial x_\kappa} = 0$$

$$g_{il}\frac{\partial^2 \gamma_{\kappa l}}{\partial x_\kappa \partial x_m} + \left(\frac{\partial g_{il}}{\partial x_m}\frac{\partial \gamma_{\kappa l}}{\partial x_\kappa} + \frac{\partial \gamma_{\kappa l}}{\partial x_m}\frac{\partial g_{il}}{\partial x_\kappa}\right) + \gamma_{\kappa l}\frac{\partial^2 g_{il}}{\partial x_\kappa \partial x_m} = 0 \quad \Big| \gamma_{im}$$

$$- g_{il}\frac{\partial \gamma_{im}}{\partial x_m}\frac{\partial \gamma_{\kappa l}}{\partial x_\kappa} - g_{il}\frac{\partial \gamma_{im}}{\partial x_\kappa}\frac{\partial \gamma_{\kappa l}}{\partial x_m}$$

[p. 16 L]

$$g_{\kappa\rho}\frac{\partial\gamma_{\kappa\alpha}}{\partial x_\alpha}\frac{\partial\gamma_{\rho\beta}}{\partial x_\beta}+\underbrace{\frac{3}{4}\gamma_{\kappa l}\frac{\partial\gamma_{m\rho}}{\partial x_\kappa}\frac{\partial g_{m\rho}}{\partial x_l}}+\frac{1}{2}g_{\kappa\rho}\frac{\partial\gamma_{\kappa m}}{\partial x_\sigma}\frac{\partial\gamma_{\rho\sigma}}{\partial x_m}$$

$$-\frac{3}{4}g_{m\rho}\gamma_{\kappa l}\frac{\partial^2\gamma_{m\rho}}{\partial x_\kappa\partial x_l}$$

$$\frac{1}{4} \qquad -\frac{1}{2}$$

$$\gamma_{im}\gamma_{\kappa l}\frac{\partial^2 g_{im}}{\partial x_\kappa\partial x_l} = +\,g_{im}\gamma_{\kappa l}\frac{\partial^2\gamma_{im}}{\partial x_\kappa\partial x_l}$$

$$-\gamma_{im}\gamma_{\kappa l}\frac{\partial^2 g_{il}}{\partial x_\kappa\partial x_m} = \underbrace{g_{il}\gamma_{im}\frac{\partial^2\gamma_{\kappa l}}{\partial x_\kappa\partial x_m}}_{\frac{\partial^2\gamma_{\alpha\beta}}{\partial x_\alpha\partial x_\beta}} - g_{il}\frac{\partial\gamma_{im}}{\partial x_m}\frac{\partial\gamma_{\kappa l}}{\partial x_\kappa} - g_{il}\frac{\partial\gamma_{im}}{\partial x_\kappa}\frac{\partial\gamma_{\kappa l}}{\partial x_m}$$

$$\begin{array}{cccc} i & l & m & \kappa \\ \kappa & \rho & \alpha & \beta \end{array}$$

$$g_{\kappa\rho}\gamma_{\kappa\alpha}\underline{\frac{\partial^2\gamma_{\rho\beta}}{\partial x_\alpha\partial x_\beta}}$$

$$\frac{1}{4}g_{\kappa\rho}\gamma_{\alpha\beta}\frac{\partial^2\gamma_{\kappa\rho}}{\partial x_\alpha\partial x_\beta}-\frac{1}{2}g_{\kappa\rho}\frac{\partial\gamma_{\kappa\alpha}}{\partial x_\beta}\frac{\partial\gamma_{\rho\beta}}{\partial x_\alpha}+\frac{\partial^2\gamma_{\alpha\beta}}{\partial x_\alpha\partial x_\beta}$$

$$T_{i\kappa} = \frac{1}{4}\gamma_{\alpha\beta}\frac{\partial^2\gamma_{i\kappa}}{\partial x_\alpha\partial x_\beta}-\frac{1}{2}\frac{\partial\gamma_{i\alpha}}{\partial x_\beta}\frac{\partial\gamma_{\kappa\beta}}{\partial x_\alpha}+\frac{1}{2}\gamma_{\kappa\beta}\frac{\partial^2\gamma_{i\alpha}}{\partial x_\alpha\partial x_\beta}+\frac{1}{2}\gamma_{i\alpha}\frac{\partial^2\gamma_{\kappa\beta}}{\partial x_\alpha\partial x_\beta}$$

$$\frac{\partial}{\partial x_\alpha}\frac{\partial}{\partial x_\beta}(\gamma_{i\alpha}\gamma_{\kappa\beta}) = \gamma_{i\alpha}\frac{\partial^2\gamma_{\kappa\beta}}{\partial x_\alpha\partial x_\beta}+\gamma_{\kappa\beta}\frac{\partial^2\gamma_{i\alpha}}{\partial x_\alpha\partial x_\beta}+\frac{\partial\gamma_{i\alpha}}{\partial x_\alpha}\frac{\partial\gamma_{\kappa\beta}}{\partial x_\beta}+\frac{\partial\gamma_{i\alpha}}{\partial x_\beta}\frac{\partial\gamma_{\kappa\beta}}{\partial x_\alpha}$$

[p. 16 R]

$$\frac{\partial g_{im}}{\partial x_\kappa}\gamma_{im} = 0$$

$$\frac{\partial^2 g_{im}}{\partial x_l \partial x_\kappa}\gamma_{im} + \frac{\partial g_{im}}{\partial x_\kappa}\frac{\partial \gamma_{im}}{\partial x_l} = 0 \qquad \Big| \quad \gamma_{\kappa l}$$

$$\frac{\partial^2 \gamma_{im}}{\partial x_\kappa \partial x_l} g_{im} + \frac{\partial g_{im}}{\partial x_l}\frac{\partial \gamma_{im}}{\partial x_\kappa} = 0 \qquad \Big| \quad \gamma_{\kappa l}$$

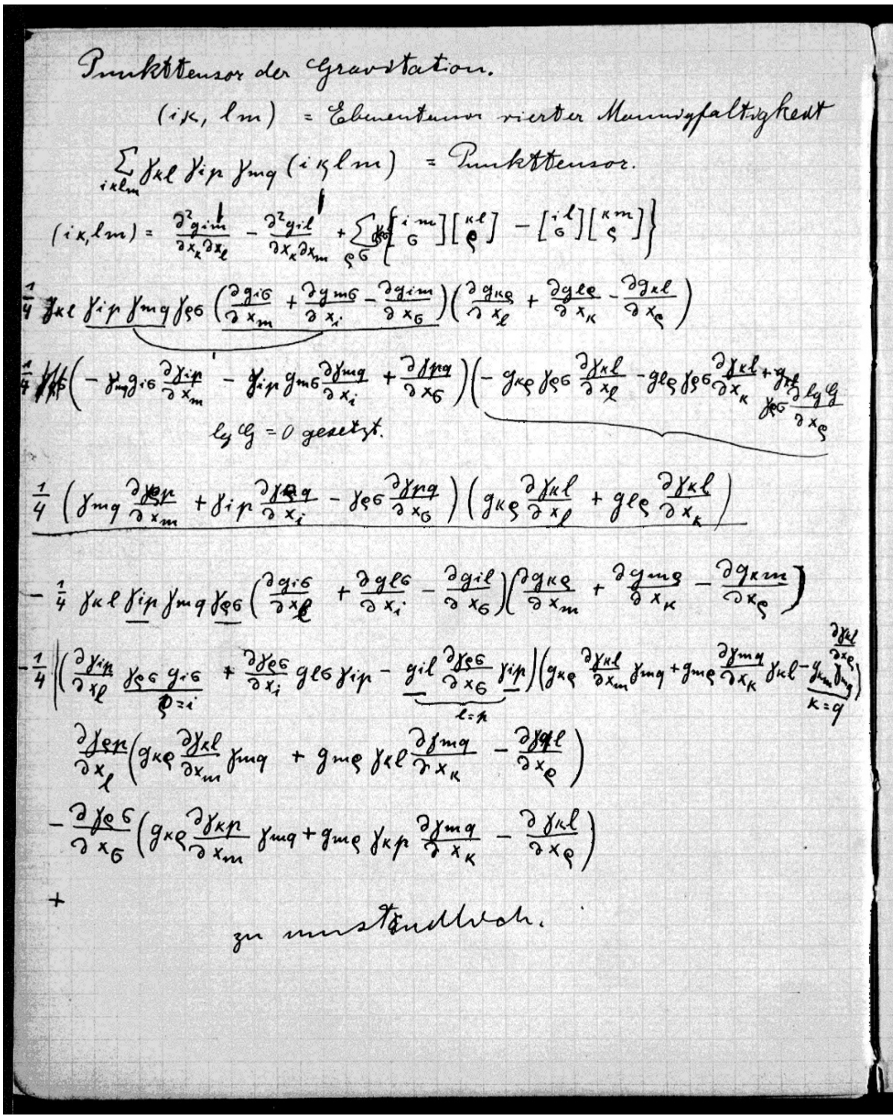
$$\gamma_{\kappa l} g_{il}$$

$$g_{il}\frac{\partial \gamma_{\kappa l}}{\partial x_\kappa} + \gamma_{\kappa l}\frac{\partial g_{il}}{\partial x_\kappa} = 0$$

$$g_{il}\frac{\partial^2 \gamma_{\kappa l}}{\partial x_\kappa \partial x_m} + \left(\frac{\partial g_{il}}{\partial x_m}\frac{\partial \gamma_{\kappa l}}{\partial x_\kappa} + \frac{\partial \gamma_{\kappa l}}{\partial x_m}\frac{\partial g_{il}}{\partial x_\kappa}\right) + \gamma_{\kappa l}\frac{\partial^2 g_{il}}{\underline{\partial} x_\kappa \partial x_m} = 0 \qquad \Big| \quad \gamma_{\underline{i}m}$$

$$-g_{il}\frac{\partial \gamma_{im}}{\partial x_m}\frac{\partial \gamma_{\kappa l}}{\partial x_\kappa} - g_{il}\frac{\partial \gamma_{im}}{\partial x_\kappa}\frac{\partial \gamma_{\kappa l}}{\partial x_m}$$

[p. 17 L]



Punkttensor der Gravitation.

(ik, lm) = Elemententensor vierter Mannigfaltigkeit

$\sum_{iklm} \gamma_{kl}\gamma_{ip}\gamma_{mq}(iklm)$ = Punkttensor.

$$(ik,lm) = \frac{\partial^2 g_{im}}{\partial x_k \partial x_l} - \frac{\partial^2 g_{il}}{\partial x_k \partial x_m} + \sum_{\rho\sigma} \left\{ \begin{bmatrix} im \\ \sigma \end{bmatrix} \begin{bmatrix} kl \\ \rho \end{bmatrix} - \begin{bmatrix} il \\ \sigma \end{bmatrix} \begin{bmatrix} km \\ \rho \end{bmatrix} \right\}$$

$$\frac{1}{4} \gamma_{kl}\gamma_{ip}\gamma_{mq}\gamma_{\rho\sigma} \left(\frac{\partial g_{i\sigma}}{\partial x_m} + \frac{\partial g_{m\sigma}}{\partial x_i} - \frac{\partial g_{im}}{\partial x_\sigma} \right) \left(\frac{\partial g_{k\rho}}{\partial x_l} + \frac{\partial g_{l\rho}}{\partial x_k} - \frac{\partial g_{kl}}{\partial x_\rho} \right)$$

$$\frac{1}{4} \left(-\gamma_{mq} g_{i\sigma} \frac{\partial \gamma_{ip}}{\partial x_m} - \gamma_{ip} g_{m\sigma} \frac{\partial \gamma_{mq}}{\partial x_i} + \frac{\partial \gamma_{pq}}{\partial x_\sigma} \right) \left(\underbrace{- g_{k\rho}\gamma_{\rho\sigma} \frac{\partial \gamma_{kl}}{\partial x_l} - g_{l\rho}\gamma_{\rho\sigma} \frac{\partial \gamma_{kl}}{\partial x_k} + \gamma_{\rho\sigma} \frac{\partial \lg G}{\partial x_\rho}}_{} \right)$$

$\lg G = 0$ gesetzt.

$$\frac{1}{4} \left(\gamma_{mq} \frac{\partial \gamma_{\sigma p}}{\partial x_m} + \gamma_{ip} \frac{\partial \gamma_{\sigma q}}{\partial x_i} - \gamma_{\rho\sigma} \frac{\partial \gamma_{pq}}{\partial x_\sigma} \right) \left(g_{k\rho} \frac{\partial \gamma_{kl}}{\partial x_l} + g_{l\rho} \frac{\partial \gamma_{kl}}{\partial x_k} \right)$$

$$- \frac{1}{4} \gamma_{kl}\gamma_{ip}\gamma_{mq}\gamma_{\rho\sigma} \left(\frac{\partial g_{i\sigma}}{\partial x_l} + \frac{\partial g_{l\sigma}}{\partial x_i} - \frac{\partial g_{il}}{\partial x_\sigma} \right) \left(\frac{\partial g_{k\rho}}{\partial x_m} + \frac{\partial g_{m\rho}}{\partial x_k} - \frac{\partial g_{km}}{\partial x_\rho} \right)$$

$$- \frac{1}{4} \left(\left(\frac{\partial \gamma_{ip}}{\partial x_l} \underbrace{\gamma_{\rho\sigma} g_{i\sigma}}_{\rho = i} + \frac{\partial \gamma_{\rho\sigma}}{\partial x_i} g_{l\sigma}\gamma_{ip} - \underbrace{g_{il} \frac{\partial \gamma_{\rho\sigma}}{\partial x_\sigma} \gamma_{ip}}_{l=p} \right) \left(g_{k\rho} \frac{\partial \gamma_{kl}}{\partial x_m} \gamma_{mq} + g_{m\rho} \frac{\partial \gamma_{mq}}{\partial x_k} \gamma_{kl} - \underbrace{\gamma_{km}\gamma_{mq} \frac{\partial \gamma_{kl}}{\partial x_\rho}}_{k=q} \right) \right.$$

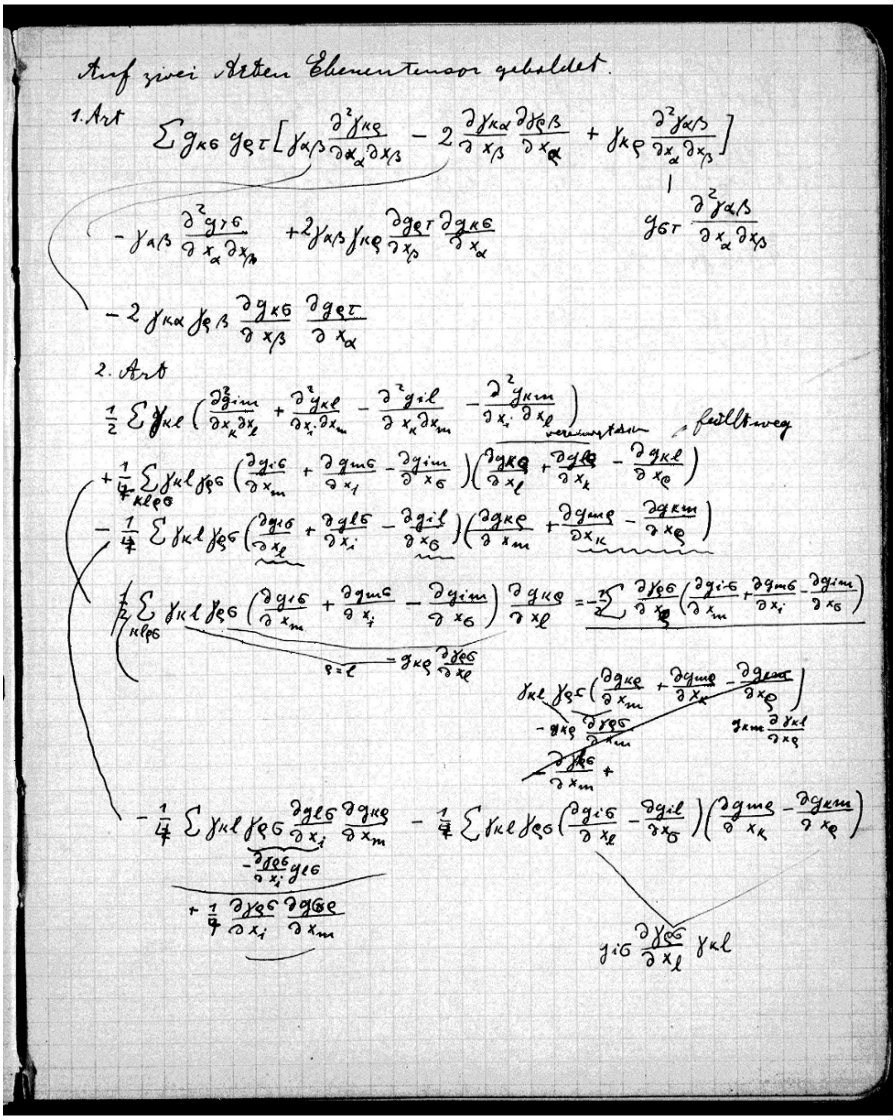
$$\frac{\partial \gamma_{\rho p}}{\partial x_l} \left(g_{k\rho} \frac{\partial \gamma_{kl}}{\partial x_m} \gamma_{mq} + g_{m\rho}\gamma_{kl} \frac{\partial \gamma_{mq}}{\partial x_k} - \frac{\partial \gamma_{ql}}{\partial x_\rho} \right)$$

$$- \frac{\partial \gamma_{\rho\sigma}}{\partial x_\sigma} \left(g_{k\rho} \frac{\partial \gamma_{kp}}{\partial x_m} \gamma_{mq} + g_{m\rho}\gamma_{kp} \frac{\partial \gamma_{mq}}{\partial x_k} - \frac{\partial \gamma_{kl}}{\partial x_\rho} \right)$$

$+$

zu umständlich.

[p. 17 R]



Auf zwei Arten Elementartensor gebildet.

1. Art

$$\sum g_{\kappa\sigma} g_{\varrho\tau} \left[\gamma_{\alpha\beta} \frac{\partial^2 \gamma_{\kappa\varrho}}{\partial x_\alpha \partial x_\beta} - 2 \frac{\partial \gamma_{\kappa\alpha}}{\partial x_\beta} \frac{\partial \gamma_{\varrho\beta}}{\partial x_\alpha} + \gamma_{\kappa\varrho} \frac{\partial^2 \gamma_{\alpha\beta}}{\partial x_\alpha \partial x_\beta} \right]$$

$$\gamma_{\sigma\tau} \frac{\partial^2 \gamma_{\alpha\beta}}{\partial x_\alpha \partial x_\beta}$$

$$- \gamma_{\alpha\beta} \frac{\partial^2 g_{\tau\sigma}}{\partial x_\alpha \partial x_\beta} + 2 \gamma_{\alpha\beta} \gamma_{\kappa\varrho} \frac{\partial g_{\varrho\tau}}{\partial x_\beta} \frac{\partial g_{\kappa\sigma}}{\partial x_\alpha}$$

$$- 2 \gamma_{\kappa\alpha} \gamma_{\varrho\beta} \frac{\partial g_{\kappa\sigma}}{\partial x_\beta} \frac{\partial g_{\varrho\tau}}{\partial x_\alpha}$$

2. Art

$$\frac{1}{2} \sum \gamma_{\kappa l} \left(\frac{\partial^2 g_{im}}{\partial x_\kappa \partial x_l} + \frac{\partial^2 \gamma_{\kappa l}}{\partial x_i \partial x_m} - \frac{\partial^2 g_{il}}{\partial x_\kappa \partial x_m} - \frac{\partial^2 g_{\kappa m}}{\partial x_i \partial x_l} \right)$$

verschwindet fällt weg

$$+ \frac{1}{4} \sum_{\kappa l \varrho\sigma} \gamma_{\kappa l} \gamma_{\varrho\sigma} \left(\frac{\partial g_{i\sigma}}{\partial x_m} + \frac{\partial g_{m\sigma}}{\partial x_i} - \frac{\partial g_{im}}{\partial x_\sigma} \right) \left(\frac{\partial g_{\kappa\varrho}}{\partial x_l} + \frac{\partial g_{l\varrho}}{\partial x_\kappa} - \frac{\partial g_{\kappa l}}{\partial x_\varrho} \right)$$

$$- \frac{1}{4} \sum \gamma_{\kappa l} \gamma_{\varrho\sigma} \left(\frac{\partial g_{i\sigma}}{\partial x_l} + \frac{\partial g_{l\sigma}}{\partial x_i} - \frac{\partial g_{il}}{\partial x_\sigma} \right) \left(\frac{\partial g_{\kappa\varrho}}{\partial x_m} + \frac{\partial g_{m\varrho}}{\partial x_\kappa} - \frac{\partial g_{\kappa m}}{\partial x_\varrho} \right)$$

$$\frac{1}{2} \sum_{\kappa l \varrho\sigma} \gamma_{\kappa l} \gamma_{\varrho\sigma} \left(\frac{\partial g_{i\sigma}}{\partial x_m} + \frac{\partial g_{m\sigma}}{\partial x_i} - \frac{\partial g_{im}}{\partial x_\sigma} \right) \frac{\partial g_{\kappa\varrho}}{\partial x_l} = -\frac{1}{2} \sum \frac{\partial \gamma_{\varrho\sigma}}{\partial x_\varrho} \left(\frac{\partial g_{i\sigma}}{\partial x_m} + \frac{\partial g_{m\sigma}}{\partial x_i} - \frac{\partial g_{im}}{\partial x_\sigma} \right)$$

$$\varrho = l \qquad - g_{\kappa\varrho} \frac{\partial \gamma_{\varrho\sigma}}{\partial x_l}$$

$$\gamma_{\kappa l} \gamma_{\varrho\sigma} \left(\frac{\partial g_{\kappa\varrho}}{\partial x_m} + \frac{\partial g_{m\varrho}}{\partial x_\kappa} - \frac{\partial g_{\kappa m}}{\partial x_\varrho} \right)$$

$$- g_{\kappa\varrho} \frac{\partial \gamma_{\varrho\sigma}}{\partial x_m} \qquad g_{\kappa m} \frac{\partial \gamma_{\kappa l}}{\partial x_\varrho}$$

$$\cancel{- \frac{\partial \gamma_{l\sigma}}{\partial x_m}} +$$

$$- \frac{1}{4} \sum \gamma_{\kappa l} \gamma_{\varrho\sigma} \frac{\partial g_{l\sigma}}{\partial x_i} \frac{\partial g_{\kappa\varrho}}{\partial x_m} - \frac{1}{4} \sum \gamma_{\kappa l} \gamma_{\varrho\sigma} \left(\frac{\partial g_{i\sigma}}{\partial x_l} - \frac{\partial g_{il}}{\partial x_\sigma} \right) \left(\frac{\partial g_{m\varrho}}{\partial x_\kappa} - \frac{\partial g_{\kappa m}}{\partial x_\varrho} \right)$$

$$- \frac{\partial \gamma_{\varrho\sigma}}{\partial x_i} g_{l\sigma}$$

$$+ \frac{1}{4} \frac{\partial \gamma_{\varrho\sigma}}{\partial x_i} \frac{\partial g_{\sigma\varrho}}{\partial x_m}$$

$$g_{i\sigma} \frac{\partial \gamma_{\varrho\sigma}}{\partial x_l} \gamma_{\kappa l}$$

[p. 17 L]

Punkttensor der Gravitation.

$(i\kappa, lm)$ = Ebenentensor vierter Mannigfaltigkeit

$$\sum_{i\kappa lm} \gamma_{\kappa l}\gamma_{ip}\gamma_{mq}\ (i\kappa, lm) = \text{Punkttensor}.$$

$$(i\ \kappa,\ lm) = \frac{\partial^2 g_{im}}{\partial x_\kappa \partial x_l} - \frac{\partial^2 g_{il}}{\partial x_\kappa \partial x_m} + \sum_{\rho\sigma} \gamma_{\rho\sigma} \left\{ \begin{bmatrix} i & m \\ & \sigma \end{bmatrix} \begin{bmatrix} \kappa & l \\ & \rho \end{bmatrix} - \begin{bmatrix} i & l \\ & \sigma \end{bmatrix} \begin{bmatrix} \kappa & m \\ & \rho \end{bmatrix} \right\}$$

$$\frac{1}{4}\gamma_{\kappa l}\underline{\gamma_{ip}\gamma_{mq}}\gamma_{\rho\sigma}\left(\underline{\frac{\partial g_{i\sigma}}{\partial x_m} + \frac{\partial g_{m\sigma}}{\partial x_i} - \frac{\partial g_{im}}{\partial x_\sigma}}\right)\left(\frac{\partial g_{\kappa\rho}}{\partial x_l} + \frac{\partial g_{l\rho}}{\partial x_\kappa} - \frac{\partial g_{\kappa l}}{\partial x_\rho}\right)$$

$$\frac{1}{4}\cancel{\gamma_{\rho\sigma}}\left(-\gamma_{mq}g_{i\sigma}\frac{\partial \gamma_{ip}}{\partial x_m} - \gamma_{ip}g_{m\sigma}\frac{\partial \gamma_{mq}}{\partial x_i} + \frac{\partial \gamma_{pq}}{\partial x_\sigma}\right)\underbrace{\left(-g_{\kappa\rho}\gamma_{\rho\sigma}\frac{\partial \gamma_{\kappa l}}{\partial x_l} - g_{l\rho}\gamma_{\rho\sigma}\frac{\partial \gamma_{\kappa l}}{\partial x_\kappa} + \cancel{g_{\kappa l}}\gamma_{\rho\sigma}\frac{\partial \lg G}{\partial x_\rho}\right.}$$

$\lg G = 0$ gesetzt.

$$\frac{1}{4}\left(\gamma_{mq}\frac{\partial \gamma_{\rho p}}{\partial x_m} + \gamma_{ip}\frac{\partial \gamma_{\rho q}}{\partial x_i} - \gamma_{\rho\sigma}\frac{\partial \gamma_{pq}}{\partial x_\sigma}\right)\left(g_{\kappa\rho}\frac{\partial \gamma_{\kappa l}}{\partial x_l} + g_{l\rho}\frac{\partial \gamma_{\kappa l}}{\partial x_\kappa}\right)$$

$$-\frac{1}{4}\gamma_{\kappa l}\underline{\gamma_{ip}}\gamma_{mq}\underline{\gamma_{\rho\sigma}}\left(\frac{\partial g_{i\sigma}}{\partial x_l} + \frac{\partial g_{l\sigma}}{\partial x_i} - \frac{\partial g_{il}}{\partial x_\sigma}\right)\left(\frac{\partial g_{\kappa\rho}}{\partial x_m} + \frac{\partial g_{m\rho}}{\partial x_\kappa} - \frac{\partial g_{\kappa m}}{\partial x_\rho}\right)$$

$$-\frac{1}{4}\left|\left(\frac{\partial \gamma_{ip}}{\partial x_l}\underbrace{\gamma_{\rho\sigma}g_{i\sigma}}_{\rho\,=\,i} + \frac{\partial \gamma_{\rho\sigma}}{\partial x_i}g_{l\sigma}\gamma_{ip} - \underbrace{\underline{g_{il}}\frac{\partial \gamma_{\rho\sigma}}{\partial x_\sigma}\underline{\gamma_{ip}}}_{l\,=\,p}\right)\left(g_{\kappa\rho}\frac{\partial \gamma_{\kappa l}}{\partial x_m}\gamma_{mq} + g_{m\rho}\frac{\partial \gamma_{mq}}{\partial x_\kappa}\gamma_{\kappa l} - \underbrace{g_{\kappa m}\gamma_{mq}}_{\kappa\,=\,q}\frac{\partial \gamma_{\kappa l}}{\partial x_\rho}\right)\right.$$

$$\frac{\partial \gamma_{\rho p}}{\partial x_l}\left(g_{\kappa\rho}\frac{\partial \gamma_{\kappa l}}{\partial x_m}\gamma_{mq} + g_{m\rho}\gamma_{\kappa l}\frac{\partial \gamma_{mq}}{\partial x_\kappa} - \frac{\partial \gamma_{ql}}{\partial x_\rho}\right)$$

$$-\frac{\partial \gamma_{\rho\sigma}}{\partial x_\sigma}\left(g_{\kappa\rho}\frac{\partial \gamma_{\kappa p}}{\partial x_m}\gamma_{mq} + g_{m\rho}\gamma_{\kappa p}\frac{\partial \gamma_{mq}}{\partial x_\kappa} - \frac{\partial \gamma_{\kappa l}}{\partial x_\rho}\right)$$

+

zu umständlich.

[p. 17 R]

Auf zwei Arten Ebenentensor gebildet.

1. Art

$$\sum g_{\kappa\sigma} g_{\rho\tau}\left[\gamma_{\alpha\beta}\frac{\partial^2\gamma_{\kappa\rho}}{\partial x_\alpha \partial x_\beta} - 2\frac{\partial\gamma_{\kappa\alpha}}{\partial x_\beta}\frac{\partial\gamma_{\rho\beta}}{\partial x_\alpha} + \gamma_{\kappa\rho}\frac{\partial^2\gamma_{\alpha\beta}}{\partial x_\alpha \partial x_\beta}\right]$$

$$-\gamma_{\alpha\beta}\frac{\partial^2 g_{\tau\sigma}}{\partial x_\alpha \partial x_\beta} + 2\gamma_{\alpha\beta}\gamma_{\kappa\rho}\frac{\partial g_{\rho\tau}}{\partial x_\beta}\frac{\partial g_{\kappa\sigma}}{\partial x_\alpha} \qquad g_{\sigma\tau}\frac{\partial^2\gamma_{\alpha\beta}}{\partial x_\alpha \partial x_\beta}$$

$$-2\gamma_{\kappa\alpha}\gamma_{\rho\beta}\frac{\partial g_{\kappa\sigma}}{\partial x_\beta}\frac{\partial g_{\rho\tau}}{\partial x_\alpha}$$

2. Art

$$\frac{1}{2}\sum \gamma \langle g\rangle_{\kappa l}\left(\frac{\partial^2 g_{im}}{\partial x_\kappa \partial x_l} + \frac{\partial^2 g_{\kappa l}}{\partial x_i \partial x_m} - \frac{\partial^2 g_{il}}{\partial x_\kappa \partial x_m} - \frac{\partial^2 g_{\kappa m}}{\partial x_i \partial x_l}\right)$$

$$+\frac{1}{4\langle 2\rangle}\sum_{\kappa l\rho\sigma}\gamma_{\kappa l}\gamma_{\rho\sigma}\left(\frac{\partial g_{i\sigma}}{\partial x_m} + \frac{\partial g_{m\sigma}}{\partial x_i} - \frac{\partial g_{im}}{\partial x_\sigma}\right)\left(\overbrace{\frac{\partial g_{\kappa\rho}}{\partial x_l} + \frac{\partial g_{l\rho}}{\partial x_\kappa}}^{\text{vereinigt sich}} - \underbrace{\frac{\partial g_{\kappa l}}{\partial x_\rho}}_{\text{fällt weg}}\right)$$

$$-\frac{1}{4\langle 2\rangle}\sum \gamma_{\kappa l}\gamma_{\rho\sigma}\left(\frac{\partial g_{i\sigma}}{\partial x_l} + \frac{\partial g_{l\sigma}}{\partial x_i} - \frac{\partial g_{il}}{\partial x_\sigma}\right)\left(\frac{\partial g_{\kappa\rho}}{\partial x_m} + \frac{\partial g_{m\rho}}{\partial x_\kappa} - \frac{\partial g_{\kappa m}}{\partial x_\rho}\right)$$

$$\left\langle\frac{1}{2}\right\rangle\sum_{\kappa l\rho\sigma}\gamma_{\kappa l}\gamma_{\rho\sigma}\left(\frac{\partial g_{i\sigma}}{\partial x_m} + \frac{\partial g_{m\sigma}}{\partial x_i} - \frac{\partial g_{im}}{\partial x_\sigma}\right)\frac{\partial g_{\kappa\rho}}{\partial x_l} = \underline{-\frac{1}{2}\sum\frac{\partial\gamma_{\rho\sigma}}{\partial x_\rho}\left(\frac{\partial g_{i\sigma}}{\partial x_m} + \frac{\partial g_{m\sigma}}{\partial x_i} - \frac{\partial g_{im}}{\partial x_\sigma}\right)}$$

$$\rho = l \qquad = g_{\kappa\rho}\frac{\partial\gamma_{\rho\sigma}}{\partial x_l}$$

$$\gamma_{\kappa l}\gamma_{\rho\sigma}\left(\frac{\partial g_{\kappa\rho}}{\partial x_m} + \frac{\partial g_{m\rho}}{\partial x_\kappa} - \frac{\partial g_{\kappa m}}{\partial x_\rho}\right)$$

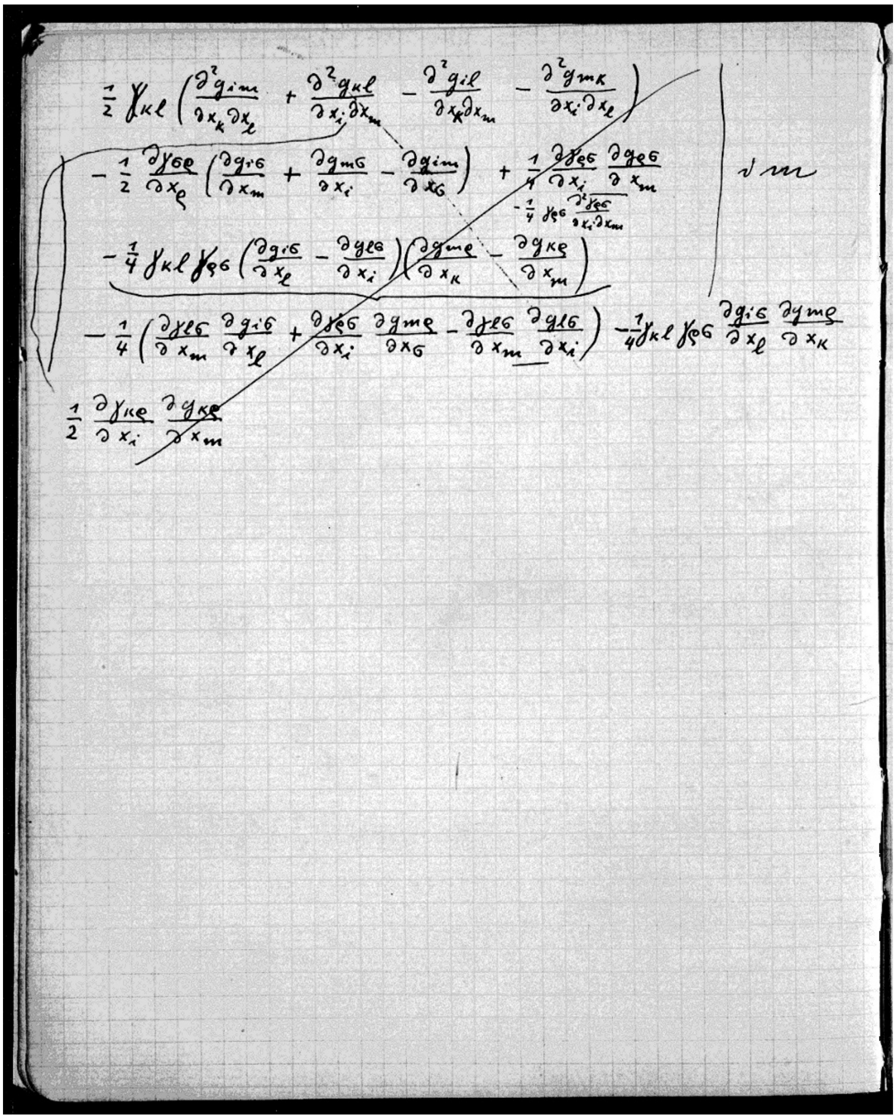
$$-g_{\kappa\rho}\frac{\partial\gamma_{\rho\sigma}}{\partial x_m} \qquad g_{\kappa m}\frac{\partial\gamma_{\kappa l}}{\partial x_\rho}$$

$$\cancel{-\frac{\partial\gamma_{l\sigma}}{\partial x_m}+}$$

$$-\frac{1}{4}\sum\gamma_{\kappa l}\underbrace{\gamma_{\rho\sigma}\frac{\partial g_{l\sigma}}{\partial x_i}}_{-\frac{\partial\gamma_{\rho\sigma}}{\partial x_i}g_{l\sigma}}\frac{\partial g_{\kappa\rho}}{\partial x_m} - \frac{1}{4}\sum\gamma_{\kappa l}\gamma_{\rho\sigma}\left(\frac{\partial g_{i\sigma}}{\partial x_l} - \frac{\partial g_{il}}{\partial x_\sigma}\right)\left(\frac{\partial g_{m\rho}}{\partial x_\kappa} - \frac{\partial g_{\kappa m}}{\partial x_\rho}\right)$$

$$+\frac{1}{4}\frac{\partial\gamma_{\rho\sigma}}{\partial x_i}\frac{\partial g_{\sigma\rho}}{\partial x_m} \qquad g_{i\sigma}\frac{\partial\gamma_{\rho\sigma}}{\partial x_l}\gamma_{\kappa l}$$

[p. 18 L]



$$\frac{1}{2}\gamma_{\kappa l}\left(\frac{\partial^2 g_{im}}{\partial x_\kappa \partial x_l}+\frac{\partial^2 g_{\kappa l}}{\partial x_i \partial x_m}-\frac{\partial^2 g_{il}}{\partial x_\kappa \partial x_m}-\frac{\partial^2 g_{m\kappa}}{\partial x_i \partial x_l}\right)$$

$$-\frac{1}{2}\frac{\partial \gamma_{\sigma\rho}}{\partial x_\rho}\left(\frac{\partial g_{i\sigma}}{\partial x_m}+\frac{\partial g_{m\sigma}}{\partial x_i}-\frac{\partial g_{im}}{\partial x_\sigma}\right)+\frac{1}{4}\frac{\partial \gamma_{\rho\sigma}}{\partial x_i}\frac{\partial g_{\rho\sigma}}{\partial x_m} \qquad \partial m$$

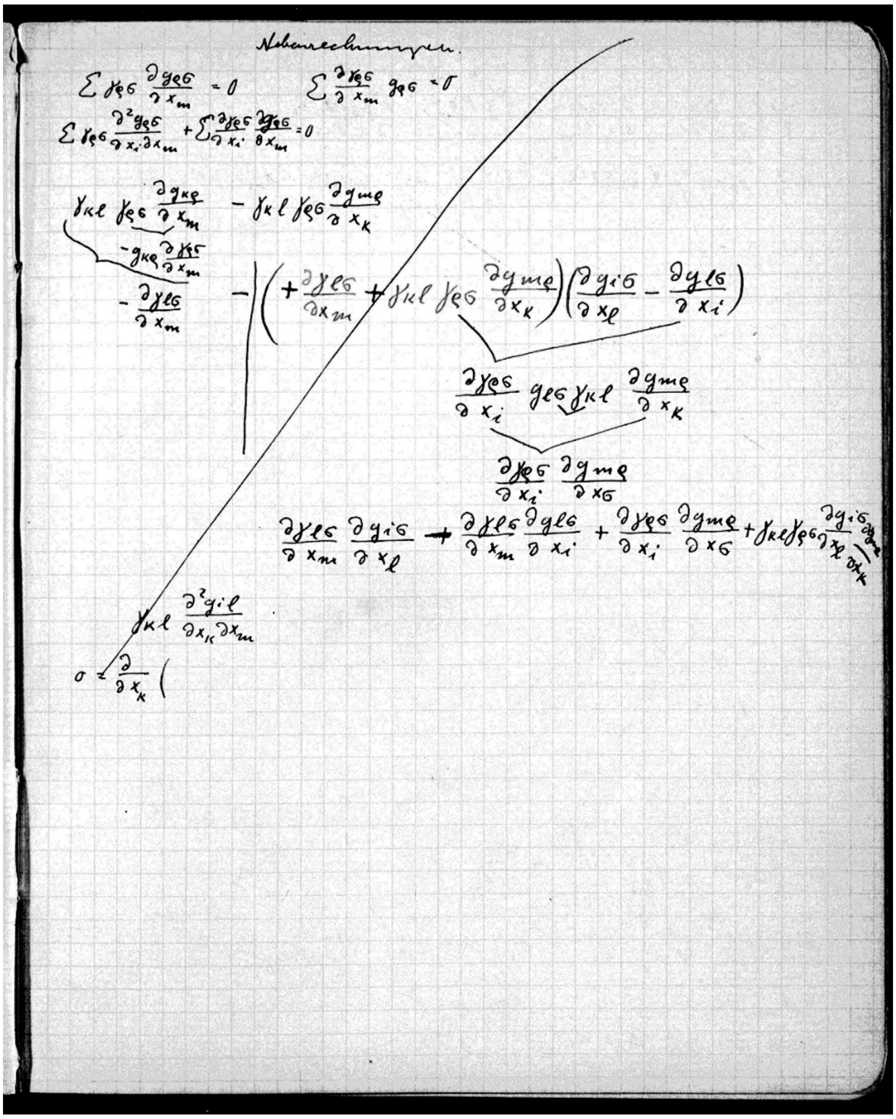
$$-\frac{1}{4}\gamma_{\rho\sigma}\frac{\partial^2 g_{\rho\sigma}}{\partial x_i \partial x_m}$$

$$-\frac{1}{4}\gamma_{\kappa l}\gamma_{\rho\sigma}\left(\frac{\partial g_{i\sigma}}{\partial x_l}-\frac{\partial g_{l\sigma}}{\partial x_i}\right)\left(\frac{\partial g_{m\rho}}{\partial x_\kappa}-\frac{\partial g_{\kappa\rho}}{\partial x_m}\right)$$

$$-\frac{1}{4}\left(\frac{\partial \gamma_{l\sigma}}{\partial x_m}\frac{\partial g_{i\sigma}}{\partial x_l}+\frac{\partial \gamma_{\rho\sigma}}{\partial x_i}\frac{\partial g_{m\rho}}{\partial x_\sigma}-\frac{\partial \gamma_{l\sigma}}{\partial x_m}\frac{\partial g_{l\sigma}}{\partial x_i}\right)-\frac{1}{4}\gamma_{\kappa l}\gamma_{\rho\sigma}\frac{\partial g_{i\sigma}}{\partial x_l}\frac{\partial g_{m\rho}}{\partial x_\kappa}$$

$$\frac{1}{2}\frac{\partial \gamma_{\kappa\rho}}{\partial x_i}\frac{\partial g_{\kappa\rho}}{\partial x_m}$$

[p. 18R]



Nebenrechnungen.

$$\sum \gamma_{\varrho\sigma}\frac{\partial g_{\varrho\sigma}}{\partial x_m} = 0 \qquad \sum \frac{\partial \gamma_{\varrho\sigma}}{\partial x_m} g_{\varrho\sigma} = 0$$

$$\sum \gamma_{\varrho\sigma}\frac{\partial^2 g_{\varrho\sigma}}{\partial x_i \partial x_m} + \sum \frac{\partial \gamma_{\varrho\sigma}}{\partial x_i}\frac{\partial g_{\varrho\sigma}}{\partial x_m} = 0$$

$$\underbrace{\gamma_{\kappa l}\gamma_{\varrho\sigma}\frac{\partial g_{\kappa\varrho}}{\partial x_m}}_{\substack{-g_{\kappa\varrho}\frac{\partial \gamma_{\varrho\sigma}}{\partial x_m} \\ -\frac{\partial \gamma_{l\sigma}}{\partial x_m}}} - \gamma_{\kappa l}\gamma_{\varrho\sigma}\frac{\partial g_{m\varrho}}{\partial x_\kappa}$$

$$-\left(+\frac{\partial \gamma_{l\sigma}}{\partial x_m} + \gamma_{\kappa l}\gamma_{\varrho\sigma}\frac{\partial g_{m\varrho}}{\partial x_\kappa}\right)\left(\frac{\partial g_{i\sigma}}{\partial x_l} - \frac{\partial g_{l\sigma}}{\partial x_i}\right)$$

$$\frac{\partial \gamma_{\varrho\sigma}}{\partial x_i} g_{l\sigma}\gamma_{\kappa l}\frac{\partial g_{m\varrho}}{\partial x_\kappa}$$

$$\frac{\partial \gamma_{\varrho\sigma}}{\partial x_i}\frac{\partial g_{m\varrho}}{\partial x_\sigma}$$

$$\frac{\partial \gamma_{l\sigma}}{\partial x_m}\frac{\partial g_{i\sigma}}{\partial x_l} + \frac{\partial \gamma_{l\sigma}}{\partial x_m}\frac{\partial g_{l\sigma}}{\partial x_i} + \frac{\partial \gamma_{\varrho\sigma}}{\partial x_i}\frac{\partial g_{m\varrho}}{\partial x_\sigma} + \gamma_{\kappa l}\gamma_{\varrho\sigma}\frac{\partial g_{i\sigma}}{\partial x_l}\frac{\partial g_{m\varrho}}{\partial x_\kappa}$$

$$\gamma_{\kappa l}\frac{\partial^2 g_{il}}{\partial x_\kappa \partial x_m}.$$

$$0 = \frac{\partial}{\partial x_\kappa}\Big($$

[p. 18 L]

$$\frac{1}{2}\gamma_{\kappa l}\left(\frac{\partial^2 g_{im}}{\partial x_\kappa \partial x_l} + \frac{\partial^2 g_{\kappa l}}{\partial x_i \partial x_m} - \frac{\partial^2 g_{il}}{\partial x_\kappa \partial x_m} - \frac{\partial^2 g_{m\kappa}}{\partial x_i \partial x_l}\right)$$

$$-\frac{1}{2}\frac{\partial \gamma_{\sigma\rho}}{\partial x_\rho}\left(\frac{\partial g_{i\sigma}}{\partial x_m} + \frac{\partial g_{m\sigma}}{\partial x_i} - \frac{\partial g_{im}}{\partial x_\sigma}\right) + \frac{1}{4}\frac{\partial \gamma_{\rho\sigma}}{\partial x_i}\frac{\partial g_{\rho\sigma}}{\partial x_m} \qquad i \; m$$

$$-\frac{1}{4}\gamma_{\rho\sigma}\frac{\partial^2 \gamma_{\rho\sigma}}{\partial x_i \partial x_m}$$

$$\underbrace{-\frac{1}{4}\gamma_{\kappa l}\gamma_{\rho\sigma}\left(\frac{\partial g_{i\sigma}}{\partial x_l} - \frac{\partial g_{l\sigma}}{\partial x_i}\right)\left(\frac{\partial g_{m\rho}}{\partial x_\kappa} - \frac{\partial g_{\kappa\rho}}{\partial x_m}\right)}$$

$$-\frac{1}{4}\left(\frac{\partial \gamma_{l\sigma}}{\partial x_m}\frac{\partial g_{i\sigma}}{\partial x_l} + \frac{\partial \gamma_{\rho\sigma}}{\partial x_i}\frac{\partial g_{m\rho}}{\partial x_\sigma} - \frac{\partial \gamma_{l\sigma}}{\partial x_m}\frac{\partial g_{l\sigma}}{\partial x_i}\right) - \frac{1}{4}\gamma_{\kappa l}\gamma_{\rho\sigma}\frac{\partial g_{i\sigma}}{\partial x_l}\frac{\partial g_{m\rho}}{\partial x_\kappa} \quad \text{[eq. 132]}$$

$$\frac{1}{2}\frac{\partial \gamma_{\kappa\rho}}{\partial x_i}\frac{\partial g_{\kappa\rho}}{\partial x_m}$$

[p. 18 R]

Nebenrechnungen.

$$\sum \gamma_{\rho\sigma} \frac{\partial g_{\rho\sigma}}{\partial x_m} = 0 \qquad \sum \frac{\partial \gamma_{\rho\sigma}}{\partial x_m} g_{\rho\sigma} = 0$$

$$\sum \gamma_{\rho\sigma} \frac{\partial^2 g_{\rho\sigma}}{\partial x_i \partial x_m} + \sum \frac{\partial \gamma_{\rho\sigma}}{\partial x_i} \frac{\partial g_{\rho\sigma}}{\partial x_m} = 0$$

$$\gamma_{\kappa l} \gamma_{\rho\sigma} \frac{\partial g_{\kappa\rho}}{\partial x_m} - \gamma_{\kappa l} \gamma_{\rho\sigma} \frac{\partial g_{m\rho}}{\partial x_\kappa}$$

$$-g_{\kappa\rho} \frac{\partial \gamma_{\rho\sigma}}{\partial x_m}$$

$$-\frac{\partial \gamma_{l\sigma}}{\partial x_m}$$

$$-\left(+ \frac{\partial \gamma_{l\sigma}}{\partial x_m} + \gamma_{\kappa l} \gamma_{\rho\sigma} \frac{\partial g_{m\rho}}{\partial x_\kappa}\right)\left(\frac{\partial g_{i\sigma}}{\partial x_l} - \frac{\partial g_{l\sigma}}{\partial x_i}\right)$$

$$\frac{\partial \gamma_{\rho\sigma}}{\partial x_i} g_{l\sigma} \gamma_{\kappa l} \frac{\partial g_{m\rho}}{\partial x_\kappa}$$

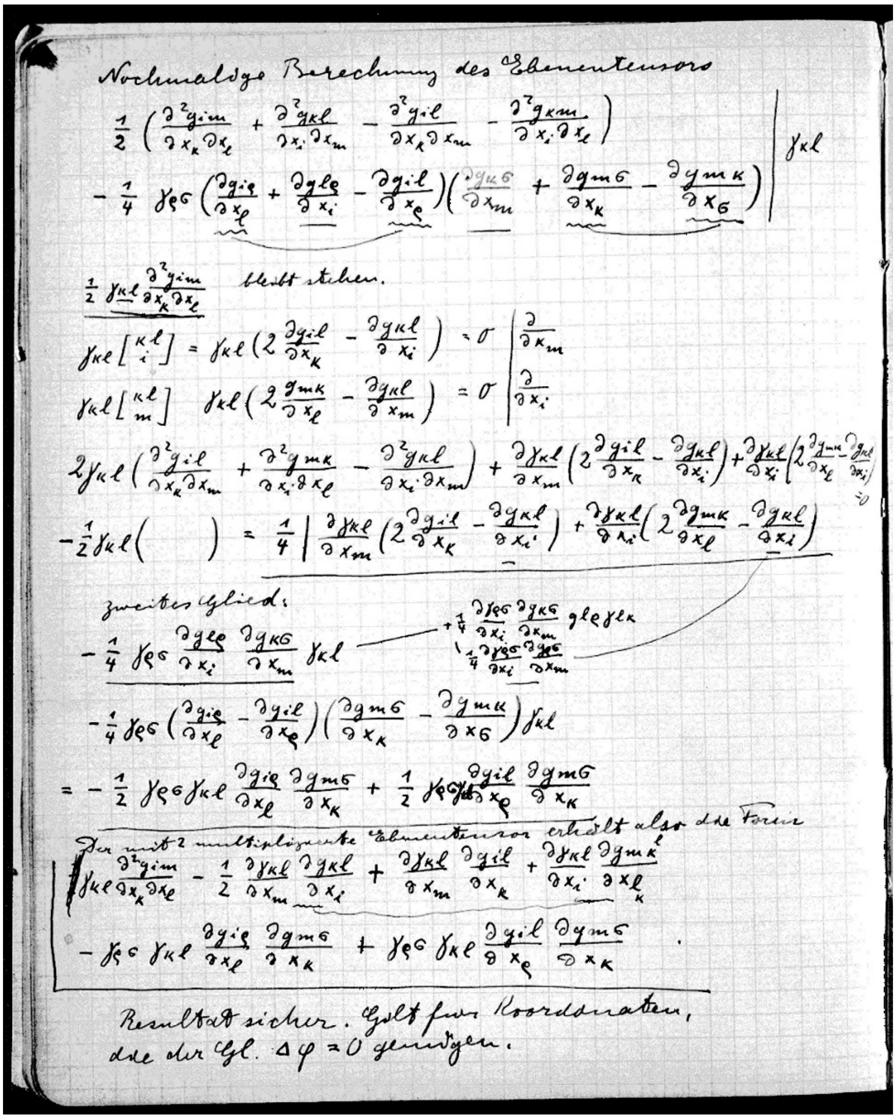
$$\frac{\partial \gamma_{\rho\sigma}}{\partial x_i} \frac{\partial g_{m\rho}}{\partial x_\sigma}$$

$$\frac{\partial \gamma_{l\sigma}}{\partial x_m} \frac{\partial g_{i\sigma}}{\partial x_l} \langle + \rangle - \frac{\partial \gamma_{l\sigma}}{\partial x_m} \frac{\partial g_{l\sigma}}{\partial x_i} + \frac{\partial \gamma_{\rho\sigma}}{\partial x_i} \frac{\partial g_{m\rho}}{\partial x_\sigma} + \gamma_{\kappa l} \gamma_{\rho\sigma} \frac{\partial g_{i\sigma}}{\partial x_l} \frac{\partial g_{m\rho}}{\partial x_\kappa}$$

$$\gamma_{\kappa l} \frac{\partial^2 g_{il}}{\partial x_\kappa \partial x_m}$$

$$0 = \frac{\partial}{\partial x_\kappa}\Big($$

[p. 19 L]



Nochmalige Berechnung des Elemententensors

$$\left.\begin{array}{l}\frac{1}{2}\left(\frac{\partial^2 g_{im}}{\partial x_\kappa \partial x_l} + \frac{\partial^2 g_{\kappa l}}{\partial x_i \partial x_m} - \frac{\partial^2 g_{il}}{\partial x_\kappa \partial x_m} - \frac{\partial^2 g_{\kappa m}}{\partial x_i \partial x_l}\right) \\ - \frac{1}{4}\gamma_{\varrho\sigma}\left(\frac{\partial g_{i\varrho}}{\partial x_l} + \frac{\partial g_{l\varrho}}{\partial x_i} - \frac{\partial g_{il}}{\partial x_\varrho}\right)\left(\frac{\partial g_{\kappa\sigma}}{\partial x_m} + \frac{\partial g_{m\sigma}}{\partial x_\kappa} - \frac{\partial g_{m\kappa}}{\partial x_\sigma}\right)\end{array}\right| \gamma_{\kappa l}$$

$\frac{1}{2}\gamma_{\kappa l}\frac{\partial^2 g_{im}}{\partial x_\kappa \partial x_l}$ bleibt stehen.

$$\gamma_{\kappa l}\begin{bmatrix}\kappa l \\ i\end{bmatrix} = \gamma_{\kappa l}\left(2\frac{\partial g_{il}}{\partial x_\kappa} - \frac{\partial g_{\kappa l}}{\partial x_i}\right) = 0 \quad \Big| \frac{\partial}{\partial x_m}$$

$$\gamma_{\kappa l}\begin{bmatrix}\kappa l \\ m\end{bmatrix} \quad \gamma_{\kappa l}\left(2\frac{\partial g_{m\kappa}}{\partial x_l} - \frac{\partial g_{\kappa l}}{\partial x_m}\right) = 0 \quad \Big| \frac{\partial}{\partial x_i}$$

$$2\gamma_{\kappa l}\left(\frac{\partial^2 g_{il}}{\partial x_\kappa \partial x_m} + \frac{\partial^2 g_{m\kappa}}{\partial x_i \partial x_l} - \frac{\partial^2 g_{\kappa l}}{\partial x_i \partial x_m}\right) + \frac{\partial \gamma_{\kappa l}}{\partial x_m}\left(2\frac{\partial g_{il}}{\partial x_\kappa} - \frac{\partial g_{\kappa l}}{\partial x_i}\right) + \frac{\partial \gamma_{\kappa l}}{\partial x_i}\left(2\frac{\partial g_{m\kappa}}{\partial x_l} - \frac{\partial g_{\kappa l}}{\partial x_i}\right) = 0$$

$$-\frac{1}{2}\gamma_{\kappa l}\left(\quad\right) = \frac{1}{4}\left|\frac{\partial \gamma_{\kappa l}}{\partial x_m}\left(2\frac{\partial g_{il}}{\partial x_\kappa} - \frac{\partial g_{\kappa l}}{\partial x_i}\right) + \frac{\partial \gamma_{\kappa l}}{\partial x_i}\left(2\frac{\partial g_{m\kappa}}{\partial x_l} - \frac{\partial g_{\kappa l}}{\partial x_i}\right)\right.$$

zweites Glied:

$$-\frac{1}{4}\gamma_{\varrho\sigma}\frac{\partial g_{l\varrho}}{\partial x_i}\frac{\partial g_{\kappa\sigma}}{\partial x_m}\gamma_{\kappa l} \qquad +\frac{1}{4}\frac{\partial \gamma_{\varrho\sigma}}{\partial x_i}\frac{\partial g_{\kappa\sigma}}{\partial x_m}g_{l\varrho}\gamma_{l\kappa} \qquad \frac{1}{4}\frac{\partial \gamma_{\varrho\sigma}}{\partial x_i}\frac{\partial g_{\varrho\sigma}}{\partial x_m}$$

$$-\frac{1}{4}\gamma_{\varrho\sigma}\left(\frac{\partial g_{i\varrho}}{\partial x_l} - \frac{\partial g_{il}}{\partial x_\varrho}\right)\left(\frac{\partial g_{m\sigma}}{\partial x_\kappa} - \frac{\partial g_{m\kappa}}{\partial x_\sigma}\right)\gamma_{\kappa l}$$

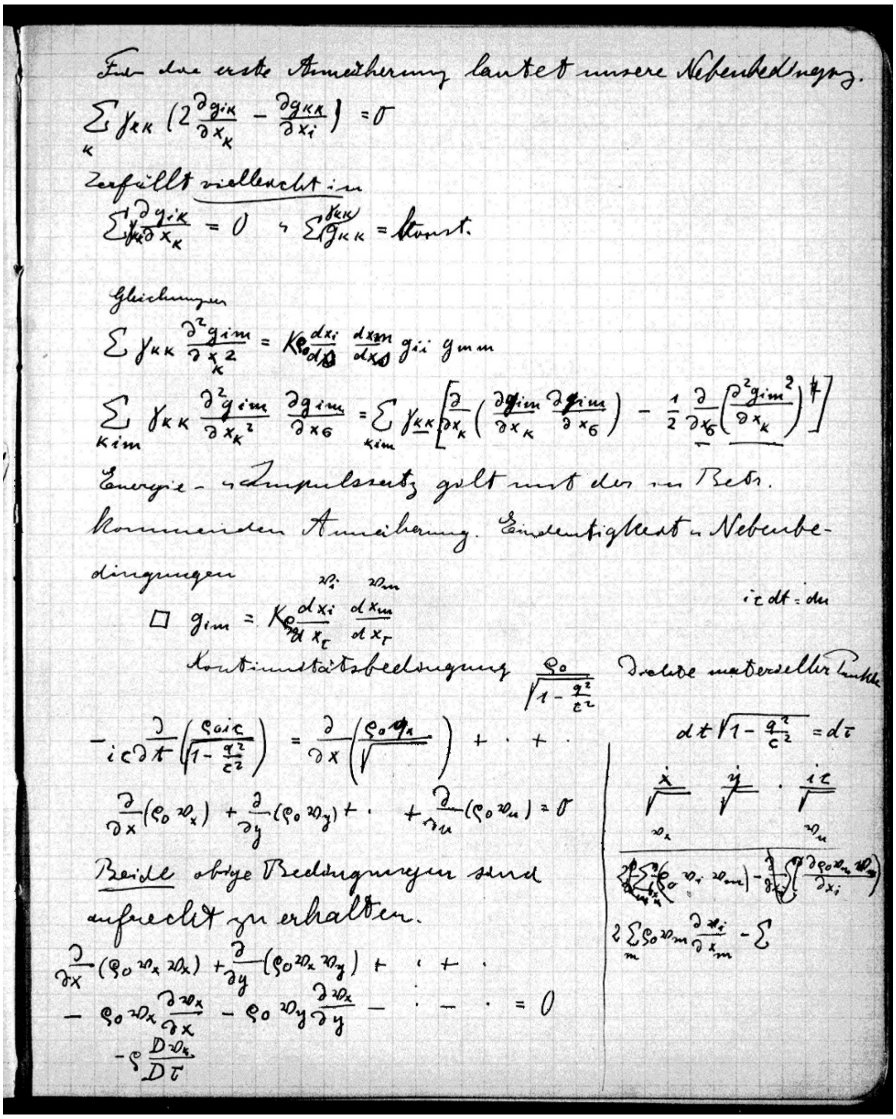
$$= -\frac{1}{2}\gamma_{\varrho\sigma}\gamma_{\kappa l}\frac{\partial g_{i\varrho}}{\partial x_l}\frac{\partial g_{m\sigma}}{\partial x_\kappa} + \frac{1}{2}\gamma_{\varrho\sigma}\gamma_{\kappa l}\frac{\partial g_{il}}{\partial x_\varrho}\frac{\partial g_{m\sigma}}{\partial x_\kappa}$$

Der mit 2 multiplizierte Elemententensor erhält also die Form

$$\boxed{\begin{array}{l}\gamma_{\kappa l}\frac{\partial^2 g_{im}}{\partial x_\kappa \partial x_l} - \frac{1}{2}\frac{\partial \gamma_{\kappa l}}{\partial x_m}\frac{\partial g_{\kappa l}}{\partial x_i} + \frac{\partial \gamma_{\kappa l}}{\partial x_m}\frac{\partial g_{il}}{\partial x_\kappa} + \frac{\partial \gamma_{\kappa l}}{\partial x_i}\frac{\partial g_{m\kappa}}{\partial x_l} \\ - \gamma_{\varrho\sigma}\gamma_{\kappa l}\frac{\partial g_{i\varrho}}{\partial x_l}\frac{\partial g_{m\sigma}}{\partial x_\kappa} + \gamma_{\varrho\sigma}\gamma_{\kappa l}\frac{\partial g_{il}}{\partial x_\varrho}\frac{\partial g_{m\sigma}}{\partial x_\kappa}\,.\end{array}}$$

Resultat sicher. Gilt für Koordinaten, die der Gl. $\Delta\varphi = 0$ genügen.

[p. 19 R]



Für die erste Annäherung lautet unsere Nebenbedingung.

$$\sum_\kappa \gamma_{\kappa\kappa}\left(2\frac{\partial g_{i\kappa}}{\partial x_\kappa} - \frac{\partial g_{\kappa\kappa}}{\partial x_i}\right) = 0$$

Zerfällt vielleicht in

$$\sum \gamma_{\kappa\kappa}\frac{\partial g_{i\kappa}}{\partial x_\kappa} = 0 \quad \text{u.} \quad \sum \gamma_{\kappa\kappa} g_{\kappa\kappa} = \text{Konst.}$$

Gleichungen

$$\sum \gamma_{\kappa\kappa}\frac{\partial^2 g_{im}}{\partial x_\kappa^2} = K\varrho_0 \frac{dx_i}{d\tau}\frac{dx_m}{d\tau} g_{ii} g_{mm}$$

$$\sum_{\kappa i m} \gamma_{\kappa\kappa}\frac{\partial^2 g_{im}}{\partial x_\kappa^2}\frac{\partial g_{im}}{\partial x_\sigma} = \sum_{\kappa i m}\gamma_{\kappa\kappa}\left[\frac{\partial}{\partial x_\kappa}\left(\frac{\partial g_{im}}{\partial x_\kappa}\frac{\partial g_{im}}{\partial x_\sigma}\right) - \frac{1}{2}\frac{\partial}{\partial x_\sigma}\left(\frac{\partial g_{im}}{\partial x_\kappa}\right)^2\right]$$

Energie- u. Impulssatz gilt mit der in Betr. kommenden Annäherung. Eindeutigkeit u. Nebenbedingungen

$$\square g_{im} = K\varrho_0 \frac{dx_i}{d\tau}\frac{dx_m}{d\tau} \qquad ic\,dt = du$$

Kontinuitätsbedingung $\dfrac{\varrho_0}{\sqrt{1-\frac{q^2}{c^2}}}$ Dichte materieller Punkte

$$-\frac{\partial}{ic\,\partial t}\left(\frac{\varrho_0 ic}{\sqrt{1-\frac{q^2}{c^2}}}\right) = \frac{\partial}{\partial x}\left(\frac{\varrho_0 q_x}{\sqrt{\ }}\right) + \cdot + \cdot$$

$$\frac{\partial}{\partial x}(\varrho_0 \mathfrak{v}_x) + \frac{\partial}{\partial y}(\varrho_0 \mathfrak{v}_y) + \cdot + \frac{\partial}{\partial u}(\varrho_0 \mathfrak{v}_u) = 0$$

Beide obige Bedingungen sind aufrecht zu erhalten.

$$\frac{\partial}{\partial x}(\varrho_0 \mathfrak{v}_x \mathfrak{v}_x) + \frac{\partial}{\partial y}(\varrho_0 \mathfrak{v}_x \mathfrak{v}_y) + \cdot + \cdot$$

$$-\varrho_0 \mathfrak{v}_x \frac{\partial \mathfrak{v}_x}{\partial x} - \varrho_0 \mathfrak{v}_y \frac{\partial \mathfrak{v}_x}{\partial y} - \cdot - \cdot = 0$$

$$-\varrho\frac{D\mathfrak{v}_x}{D\tau}$$

$$dt\sqrt{1-\frac{q^2}{c^2}} = d\tau$$

$$\frac{\dot{x}}{\sqrt{\ }} \quad \frac{\dot{y}}{\sqrt{\ }} \cdot \frac{ic}{\sqrt{\ }}$$

$$\mathfrak{v}_x \qquad \mathfrak{v}_u$$

$$\sum \frac{\partial}{\partial x_m}(\varrho_0 \mathfrak{v}_i \mathfrak{v}_m) - \frac{\partial}{\partial x_i}\sum\left(\frac{\partial \varrho_0 \mathfrak{v}_m \mathfrak{v}_m}{\partial x_i}\right)$$

$$2\sum_m \varrho_0 \mathfrak{v}_m \frac{\partial \mathfrak{v}_i}{\partial x_m} - \sum$$

[p. 19 L]

Nochmalige Berechnung des Ebenentensors

$$\left.\begin{array}{l}\frac{1}{2}\left(\frac{\partial^2 g_{im}}{\partial x_\kappa \partial x_l} + \frac{\partial^2 g_{\kappa l}}{\partial x_i \partial x_m} - \frac{\partial^2 g_{il}}{\partial x_\kappa \partial x_m} - \frac{\partial^2 g_{\kappa m}}{\partial x_i \partial x_l}\right) \\ -\frac{1}{4}\gamma_{\rho\sigma}\left(\frac{\partial g_{i\rho}}{\partial x_l} + \underline{\frac{\partial g_{l\rho}}{\partial x_i}} - \frac{\partial g_{il}}{\partial x_\rho}\right)\left(\underline{\frac{\partial g_{\kappa\sigma}}{\partial x_m}} + \frac{\partial g_{m\sigma}}{\partial x_\kappa} - \frac{\partial g_{m\kappa}}{\partial x_\sigma}\right)\end{array}\right| \gamma_{\kappa l}$$

$$\underline{\frac{1}{2}\gamma_{\kappa l}\frac{\partial^2 g_{im}}{\partial x_\kappa \partial x_l}} \quad \text{bleibt stehen.}$$

$$\left.\begin{array}{l}\gamma_{\kappa l}\begin{bmatrix}\kappa & l \\ & i\end{bmatrix} = \gamma_{\kappa l}\left(2\frac{\partial g_{il}}{\partial x_\kappa} - \frac{\partial g_{\kappa l}}{\partial x_i}\right) = 0 \\ \gamma_{\kappa l}\begin{bmatrix}\kappa & l \\ & m\end{bmatrix} \quad \gamma_{\kappa l}\left(2\frac{\partial g_{m\kappa}}{\partial x_l} - \frac{\partial g_{\kappa l}}{\partial x_m}\right) = 0\end{array}\right| \begin{array}{l}\frac{\partial}{\partial x_m} \\ \frac{\partial}{\partial x_i}\end{array}$$

$$2\gamma_{\kappa l}\left(\frac{\partial^2 g_{il}}{\partial x_\kappa \partial x_m} + \frac{\partial^2 g_{m\kappa}}{\partial x_i \partial x_l} - \frac{\partial^2 g_{\kappa l}}{\partial x_i \partial x_m}\right) + \frac{\partial \gamma_{\kappa l}}{\partial x_m}\left(2\frac{\partial g_{il}}{\partial x_\kappa} - \frac{\partial g_{\kappa l}}{\partial x_i}\right) + \frac{\partial \gamma_{\kappa l}}{\partial x_i}\left(2\frac{\partial g_{m\kappa}}{\partial x_l} - \frac{\partial g_{\kappa l}}{\partial x_i}\right) = 0$$

$$-\frac{1}{2}\gamma_{\kappa l}(\quad) = \frac{1}{4}\left|\, \frac{\partial \gamma_{\kappa l}}{\partial x_m}\left(2\frac{\partial g_{il}}{\partial x_\kappa} - \underline{\frac{\partial g_{\kappa l}}{\partial x_i}}\right) + \frac{\partial \gamma_{\kappa l}}{\partial x_i}\left(2\frac{\partial g_{m\kappa}}{\partial x_l} - \underline{\frac{\partial g_{\kappa l}}{\partial x_i}}\right)\right.$$

zweites Glied:

$$\underline{-\frac{1}{4}\gamma_{\rho\sigma}\frac{\partial g_{l\rho}}{\partial x_i}\frac{\partial g_{\kappa\sigma}}{\partial x_m}\gamma_{\kappa l}} \qquad +\frac{1}{4}\frac{\partial \gamma_{\rho\sigma}}{\partial x_i}\frac{\partial g_{\kappa\sigma}}{\partial x_m}g_{l\rho}\gamma_{l\kappa}$$

$$\underline{\frac{1}{4}\frac{\partial \gamma_{\rho\sigma}}{\partial x_i}\frac{\partial g_{\kappa\sigma}}{\partial x_m}}$$

$$-\frac{1}{4}\gamma_{\rho\sigma}\left(\frac{\partial g_{i\rho}}{\partial x_l} - \frac{\partial g_{il}}{\partial x_\rho}\right)\left(\frac{\partial g_{m\sigma}}{\partial x_\kappa} - \frac{\partial g_{m\kappa}}{\partial x_\sigma}\right)\gamma_{\kappa l}$$

$$\underline{= -\frac{1}{2}\gamma_{\rho\sigma}\gamma_{\kappa l}\frac{\partial g_{i\rho}}{\partial x_l}\frac{\partial g_{m\sigma}}{\partial x_\kappa} + \frac{1}{2}\gamma_{\rho\sigma}\gamma_{\kappa l}\frac{\partial g_{il}}{\partial x_\rho}\frac{\partial g_{m\sigma}}{\partial x_\kappa}}$$

Der mit 2 multiplizierte Ebenentensor erhält also die Form

$$\left|\begin{array}{l}\langle\frac{1}{2}\rangle\gamma_{\kappa l}\frac{\partial^2 g_{im}}{\partial x_\kappa \partial x_l} - \frac{1}{2}\frac{\partial \gamma_{\kappa l}}{\partial x_m}\frac{\partial g_{\kappa l}}{\partial x_i} + \frac{\partial \gamma_{\kappa l}}{\partial x_m}\frac{\partial g_{il}}{\partial x_\kappa} + \frac{\partial \gamma_{\kappa l}}{\partial x_i}\frac{\partial g_{m\kappa}{}^{l}}{\partial x_{l}{}_{\kappa}} \\ -\gamma_{\rho\sigma}\gamma_{\kappa l}\frac{\partial g_{i\rho}}{\partial x_l}\frac{\partial g_{m\sigma}}{\partial x_\kappa} + \gamma_{\rho\sigma}\gamma_{\kappa l}\frac{\partial g_{il}}{\partial x_\rho}\frac{\partial g_{m\sigma}}{\partial x_\kappa}\end{array}\right.$$

Resultat sicher. Gilt für Koordinaten, die der Gl. $\Delta\varphi = 0$ genügen.

[p. 19 R]

Für die erste Annäherung lautet unsere Nebenbedingung.

$$\sum_{\kappa}\gamma_{\kappa\kappa}\left(2\frac{\partial g_{i\kappa}}{\partial x_\kappa}-\frac{\partial g_{\kappa\kappa}}{\partial x_i}\right) = 0$$

Zerfällt vielleicht in

$$\sum\gamma_{\kappa\kappa}\frac{\partial g_{i\kappa}}{\partial x_\kappa} = 0 \qquad \text{u} \qquad \sum \overset{\gamma_{\kappa\kappa}}{g_{\kappa\kappa}} = \text{konst.}$$

Gleichungen

$$\sum\gamma_{\kappa\kappa}\frac{\partial^2 g_{im}}{\partial x_\kappa^2} = K\rho_0\frac{dx_i}{ds}\frac{dx_m}{ds}g_{ii}g_{mm}$$

$$\sum_{\kappa i m}\gamma_{\kappa\kappa}\frac{\partial^2 g_{im}}{\partial x_\kappa^2}\frac{\partial g_{im}}{\partial x_\sigma} = \sum_{\kappa i m}\underline{\gamma_{\kappa\kappa}}\left[\frac{\partial}{\partial x_\kappa}\left(\frac{\partial g_{im}}{\partial x_\kappa}\frac{\partial g_{im}}{\partial x_\sigma}\right)-\frac{1}{2}\frac{\partial}{\partial \underline{x_\sigma}}\left(\frac{\partial^2 g_{im}{}^2}{\underline{\partial x_\kappa}}\right)^{\cancel{2}}\right]$$

Energie- u Impulssatz gilt mit der in Betr. kommenden Annäherung.
Eindeutigkeit u Nebenbedingungen

$$\Box g_{im} = K\rho_0\overset{w_i}{\frac{dx_i}{dx_\tau}}\overset{w_m}{\frac{dx_m}{dx_\tau}} \qquad\qquad icdt = du$$

Kontinuitätsbedingung $\dfrac{\rho_0}{\sqrt{1-\frac{q^2}{c^2}}}$ Dichte materieller Punkte

$$-\frac{\partial}{ic\,\partial t}\left(\frac{\rho_0 ic}{\sqrt{1-\frac{q^2}{c^2}}}\right) = \frac{\partial}{\partial x}\left(\frac{\rho_0 q_x}{\sqrt{}}\right) + \cdot + \cdot$$

$$\frac{\partial}{\partial x}(\rho_0 w_x)+\frac{\partial}{\partial y}(\rho_0 w_y)+ \cdot + \frac{\partial}{\partial u}(\rho_0 w_u) = 0$$

Beide obige Bedingungen sind aufrecht zu erhalten.

$$\frac{\partial}{\partial x}(\rho_0 w_x w_x)+\frac{\partial}{\partial y}(\rho_0 w_x w_y)+ \cdot + \cdot$$

$$-\rho_0 w_x\frac{\partial w_x}{\partial x}-\rho_0 w_y\frac{\partial w_x}{\partial y} - \cdot - \cdot = 0$$

$$-\rho\frac{Dw_x}{D\tau}$$

$$dt\sqrt{1-\frac{q^2}{c^2}} = d\tau$$

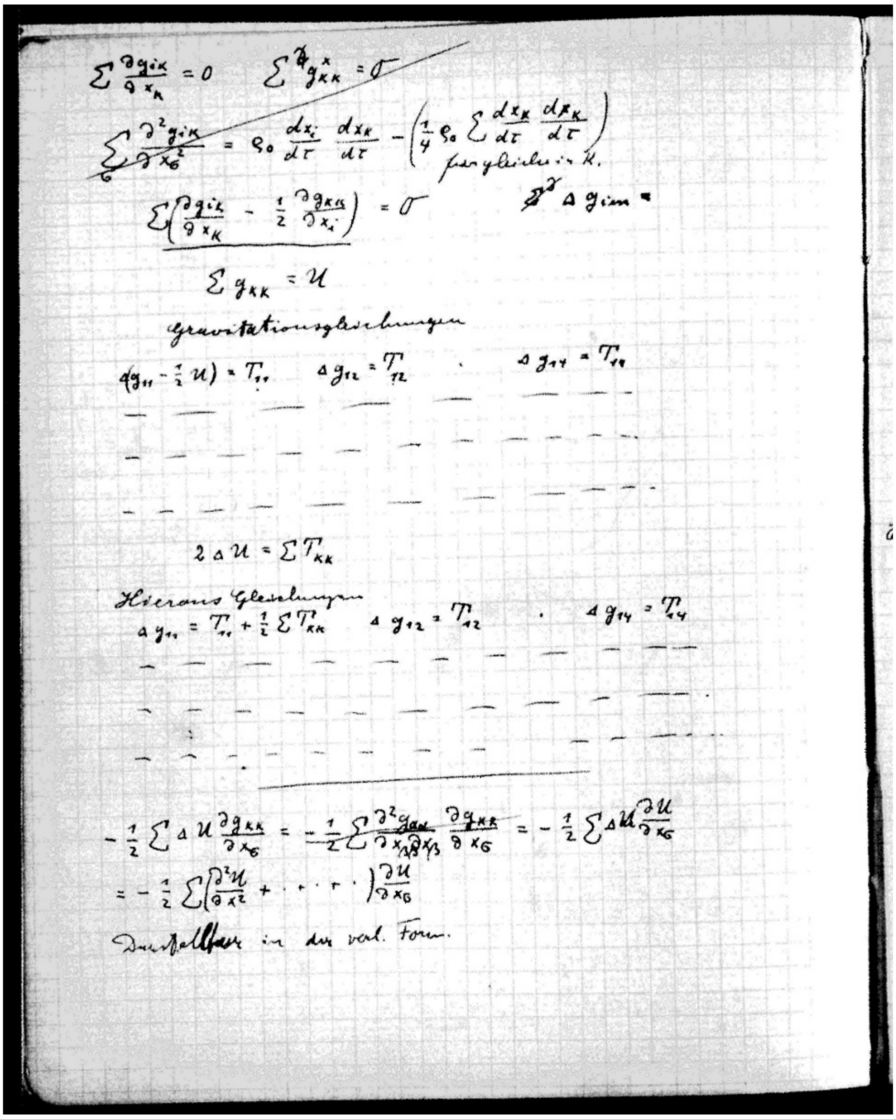
$$\underset{w_x}{\frac{\dot x}{\sqrt{}}}\quad \frac{\dot y}{\sqrt{}}\quad \cdot \quad \underset{w_u}{\frac{ic}{\sqrt{}}}$$

$$2<\frac{\partial}{\partial x_m}>\sum\frac{\partial}{\partial x_m}(\rho_0 w_i w_m)$$

$$-\cancel{\frac{\partial}{\partial x_i}}\sum\left(\frac{\partial\rho_0 w_m w_m}{\partial x_i}\right)$$

$$2\sum_m \rho_0 w_m\frac{\partial w_i}{\partial x_m}-\sum$$

[p. 20 L]



$$\sum \frac{\partial g_{ik}}{\partial x_k} = 0 \qquad \sum \frac{\partial g_{\kappa\kappa}^{x}}{} = 0$$

$$\sum_\sigma \frac{\partial^2 g_{ik}}{\partial x_\sigma^2} = \varrho_0 \frac{dx_i}{d\tau}\frac{dx_k}{d\tau} - \left(\frac{1}{4} \varrho_0 \sum \frac{dx_\kappa}{d\tau}\frac{dx_\kappa}{d\tau}\right)$$

für gleiche i u. k.

$$\sum\left(\frac{\partial g_{ik}}{\partial x_k} - \frac{1}{2} \frac{\partial g_{\kappa\kappa}}{\partial x_i}\right) = 0 \qquad \Delta g_{im} =$$

$$\sum g_{\kappa\kappa} = \mathfrak{U}$$

Gravitationsgleichungen

$\Delta(g_{11} - \frac{1}{2}\mathfrak{U}) = T_{11}$ $\quad \Delta g_{12} = T_{12}$ $\quad . \quad \Delta g_{14} = T_{14}$

— — — — — — — —

— — — — — — — —

— — — — — — — —

$$2\Delta\mathfrak{U} = \sum T_{\kappa\kappa}$$

Hieraus Gleichungen

$\Delta g_{11} = T_{11} + \frac{1}{2}\sum T_{\kappa\kappa}$ $\quad \Delta g_{12} = T_{12}$ $\quad . \quad \Delta g_{14} = T_{14}$

— — — — — — — —

— — — — — — — —

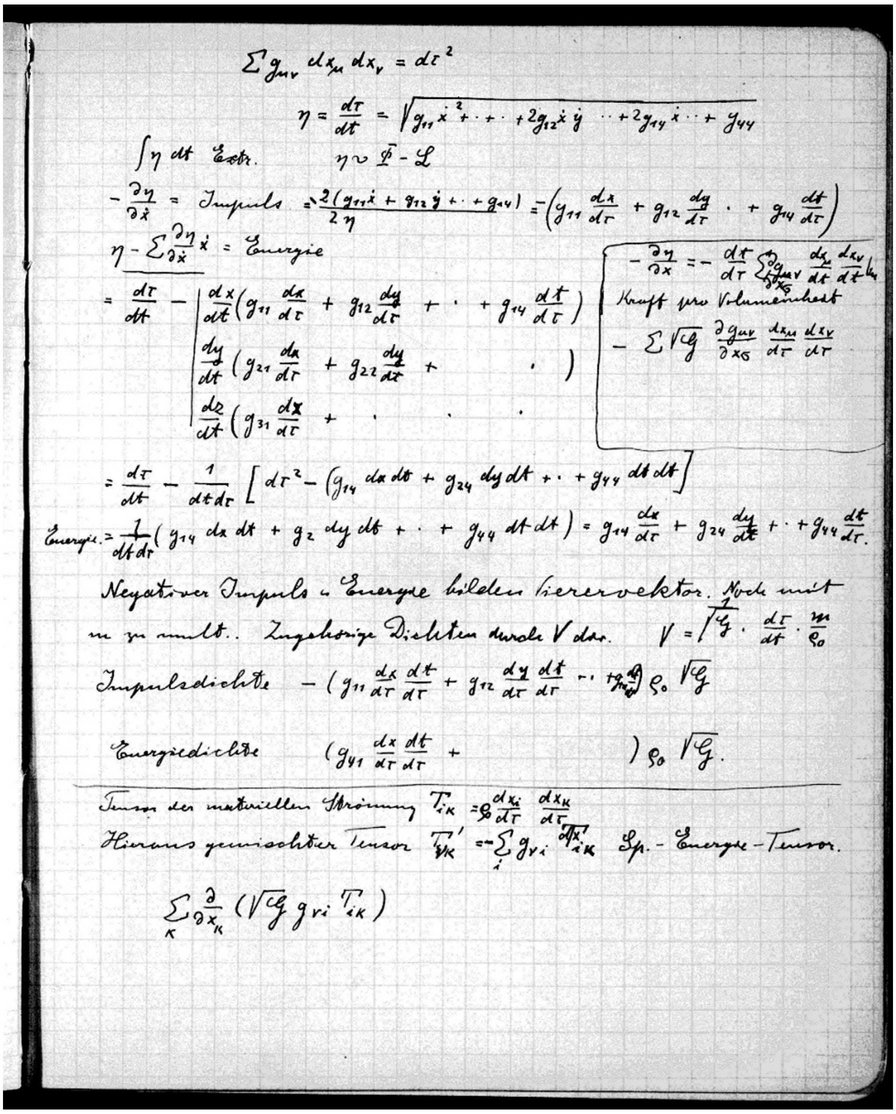
— — — — — — — —

$$-\frac{1}{2}\sum \Delta\mathfrak{U} \frac{\partial g_{\kappa\kappa}}{\partial x_\sigma} = -\frac{1}{2} \sum \frac{\partial^2 g_{\alpha\alpha}}{\partial x_\beta \partial x_\beta}\frac{\partial g_{\kappa\kappa}}{\partial x_\sigma} = -\frac{1}{2}\sum \Delta\mathfrak{U} \frac{\partial \mathfrak{U}}{\partial x_\sigma}$$

$$= -\frac{1}{2} \sum\left(\frac{\partial^2 \mathfrak{U}}{\partial x^2}+ \cdot \cdot \cdot \cdot\right)\frac{\partial \mathfrak{U}}{\partial x_\sigma}$$

Darstellbar in der verl. Form.

[p. 20 R]



$$\sum g_{\mu\nu}\, dx_\mu\, dx_\nu = d\tau^2$$

$$\eta = \frac{d\tau}{dt} = \sqrt{g_{11}\dot{x}^2 + \cdot + \cdot + 2g_{12}\dot{x}\dot{y} \cdots + 2g_{14}\dot{x} \cdots + g_{44}}$$

$\int \eta\, dt$ Extr. $\qquad \eta \sim \Phi - \mathfrak{L}$

$$-\frac{\partial\eta}{\partial\dot{x}} = \text{Impuls} = \frac{2(g_{11}\dot{x} + g_{12}\dot{y} + \cdot + g_{14})}{2\eta} = -\left(g_{11}\frac{dx}{d\tau} + g_{12}\frac{dy}{d\tau} \cdot + g_{14}\frac{dt}{d\tau}\right)$$

$$\eta - \sum \frac{\partial\eta}{\partial\dot{x}}\dot{x} = \text{Energie}$$

$$= \frac{d\tau}{dt} - \left| \begin{array}{l} \frac{dx}{dt}\left(g_{11}\frac{dx}{d\tau} + g_{12}\frac{dy}{d\tau} + \cdot + g_{14}\frac{dt}{d\tau}\right) \\ \frac{dy}{dt}\left(g_{21}\frac{dx}{d\tau} + g_{22}\frac{dy}{d\tau} + \quad \cdot\right) \\ \frac{dz}{dt}\left(g_{31}\frac{dx}{d\tau} + \quad \cdot \quad \cdot \quad \cdot\right. \end{array}\right.$$

$$-\frac{\partial\eta}{\partial x} = -\frac{dt}{d\tau}\sum \frac{\partial g_{\mu\nu}}{\partial x_\sigma}\frac{dx_\mu}{dt}\frac{dx_\nu}{dt}$$

Kraft pro Volumeneinheit

$$-\sum \sqrt{G}\,\frac{\partial g_{\mu\nu}}{\partial x_\sigma}\frac{dx_\mu}{d\tau}\frac{dx_\nu}{d\tau}$$

$$= \frac{d\tau}{dt} - \frac{1}{dt\, d\tau}\left[d\tau^2 - \left(g_{14}\, dx\, dt + g_{24}\, dy\, dt + \cdot + g_{44}\, dt\, dt\right]\right.$$

$$\text{Energie} = \frac{1}{dt\, d\tau}\left(g_{14}\, dx\, dt + g_2\, dy\, dt + \cdot + g_{44}\, dt\, dt\right) = g_{14}\frac{dx}{d\tau} + g_{24}\frac{dy}{d\tau} + \cdot + g_{44}\frac{dt}{d\tau}.$$

Negativer Impuls u. Energie bilden Vierervektor. Noch mit m zu mult.. Zugehörige Dichten durch V div. $\quad V = \frac{1}{\sqrt{G}} \cdot \frac{d\tau}{dt} \cdot \frac{m}{\rho_0}$

Impulsdichte $\quad -\left(g_{11}\frac{dx}{d\tau}\frac{dt}{d\tau} + g_{12}\frac{dy}{d\tau}\frac{dt}{d\tau} \cdot\cdot + g_{14}\frac{dt}{d\tau}\frac{dt}{d\tau}\right)\rho_0\sqrt{G}$

Energiedichte $\quad \left(g_{41}\frac{dx}{d\tau}\frac{dt}{d\tau} + \qquad\qquad \right)\rho_0\sqrt{G}.$

Tensor der materiellen Strömung $T_{ik} = \rho_0\frac{dx_i}{d\tau}\frac{dx_k}{d\tau}$

Hieraus gemischter Tensor $\mathfrak{T}_{\nu k}' = -\sum_i g_{\nu i}\, T_{ik}$ Sp.- Energie-Tensor.

$$\sum_k \frac{\partial}{\partial x_k}\left(\sqrt{G}\, g_{\nu i}\, T_{ik}\right)$$

[p. 20 L]

$$\sum \frac{\partial g_{i\kappa}}{\partial x_\kappa} = 0 \qquad \cancel{\sum g^x_{\kappa\kappa} = 0}$$

$$\cancel{\sum_\sigma \frac{\partial^2 g_{i\kappa}}{\partial x_\sigma^2} = \rho_0 \frac{dx_i}{d\tau}\frac{dx_\kappa}{d\tau} - \left(\frac{1}{4}\rho_0 \sum \frac{dx_\kappa}{d\tau}\frac{dx_\kappa}{d\tau}\right)}$$

für gleiche i u κ.

$$\sum \underline{\left(\frac{\partial g_{i\kappa}}{\partial x_\kappa} - \frac{1}{2}\frac{\partial g_{\kappa\kappa}}{\partial x_i}\right)} = 0 \qquad \cancel{\sum}\cancel{\partial}\,\Delta g_{im} =$$

$$\sum g_{\kappa\kappa} = U$$

Gravitationsgleichungen

$$\Delta\left(g_{11} - \frac{1}{2}U\right) = T_{11} \quad \Delta g_{12} = T_{12} \quad \cdot \quad \Delta g_{14} = T_{14}$$

— — — — — — — — — — —

— — — — — — — — — — —

— — — — — — — — — — —

$$2\Delta U = \sum T_{\kappa\kappa}$$

Hieraus Gleichungen

$$\Delta g_{11} = T_{11} + \frac{1}{2}\sum T_{\kappa\kappa} \quad \Delta g_{12} = T_{12} \quad \cdot \quad \Delta g_{14} = T_{14}$$

— — — — — — — — — — —

— — — — — — — — — — —

— — — — — — — — — — —

$$-\frac{1}{2}\sum \Delta U \frac{\partial g_{\kappa\kappa}}{\partial x_\sigma} = \cancel{-\frac{1}{2}\sum \frac{\partial^2 g_{\alpha\alpha}}{\partial x_\beta \partial x_\beta}\frac{\partial g_{\kappa\kappa}}{\partial x_\sigma}} = -\frac{1}{2}\sum \Delta U \frac{\partial U}{\partial x_\sigma}$$

$$= -\frac{1}{2}\sum \left[\frac{\partial^2 U}{\partial x^2} + \cdot + \cdot + \cdot\right]\frac{\partial U}{\partial x_\sigma}$$

Darstellbar in der verl. Form.

[p. 20 R]

$$\sum g_{\mu\nu} dx_\mu dx_\nu = d\tau^2$$

$$\eta = \frac{d\tau}{dt} = \sqrt{g_{11}\dot{x}^2 + \cdot + \cdot + 2g_{12}\dot{x}\dot{y} \cdot\cdot + 2g_{14}\dot{x} \cdot\cdot + g_{44}}$$

$\int \eta \; dt$ Extr. $\qquad \eta \sim \Phi - L$

$$-\frac{\partial\eta}{\partial\dot{x}} = \text{Impuls} = -\frac{2(g_{11}\dot{x} + g_{12}\dot{y} + \cdot + g_{14})}{2\eta} = -\left(g_{11}\frac{dx}{d\tau} + g_{12}\frac{dy}{d\tau} \cdot + g_{14}\frac{dt}{d\tau}\right)$$

$$\eta - \sum \frac{\partial\eta}{\partial\dot{x}}\dot{x} = \text{Energie}$$

$$= \frac{d\tau}{dt} - \left| \begin{array}{l} \frac{dx}{dt}\left(g_{11}\frac{dx}{d\tau} + g_{12}\frac{dy}{d\tau} + \cdot + g_{14}\frac{dt}{d\tau}\right) \\ \frac{dy}{dt}\left(g_{21}\frac{dx}{d\tau} + g_{22}\frac{dy}{d\tau} + \qquad \cdot \qquad \right) \\ \frac{dz}{dt}\left(g_{31}\frac{dx}{d\tau} + \quad \cdot \quad \cdot \quad \cdot \right. \end{array} \right.$$

$$-\frac{\partial\eta}{\partial x} = -\frac{dt}{d\tau}\sum \frac{\partial g_{\mu\nu}}{\partial x_\sigma}\frac{dx_\mu}{dt}\frac{dx_\nu}{dt}$$

Kraft pro Volumeneinheit

$$-\sum \sqrt{G}\frac{\partial g_{\mu\nu}}{\partial x_\sigma}\frac{dx_\mu}{d\tau}\frac{dx_\nu}{d\tau}$$

$$= \frac{d\tau}{dt} - \frac{1}{dtd\tau}[d\tau^2 - (g_{14}dxdt + g_{24}dydt + \cdot + g_{44}dtdt]$$

$$\text{Energie.} = \frac{1}{dtd\tau}(g_{14}dxdt + g_2 dydt + \cdot + g_{44}dtdt) = g_{14}\frac{dx}{d\tau} + g_{24}\frac{dy}{d\tau} + \cdot + g_{44}\frac{dt}{d\tau}.$$

Negativer Impuls u Energie bilden Vierervektor. Noch mit m zu mult..
Zugehörige Dichten durch V div. $\quad V = \frac{1}{\sqrt{G}} \cdot \frac{d\tau}{dt} \cdot \frac{m}{\rho_0}$

Impulsdichte $-\left(g_{11}\frac{dx}{d\tau}\frac{dt}{d\tau} + g_{12}\frac{dy}{d\tau}\frac{dt}{d\tau} + \cdot + g_{14}\frac{dt}{d\tau}\right)\rho_0\sqrt{G}$

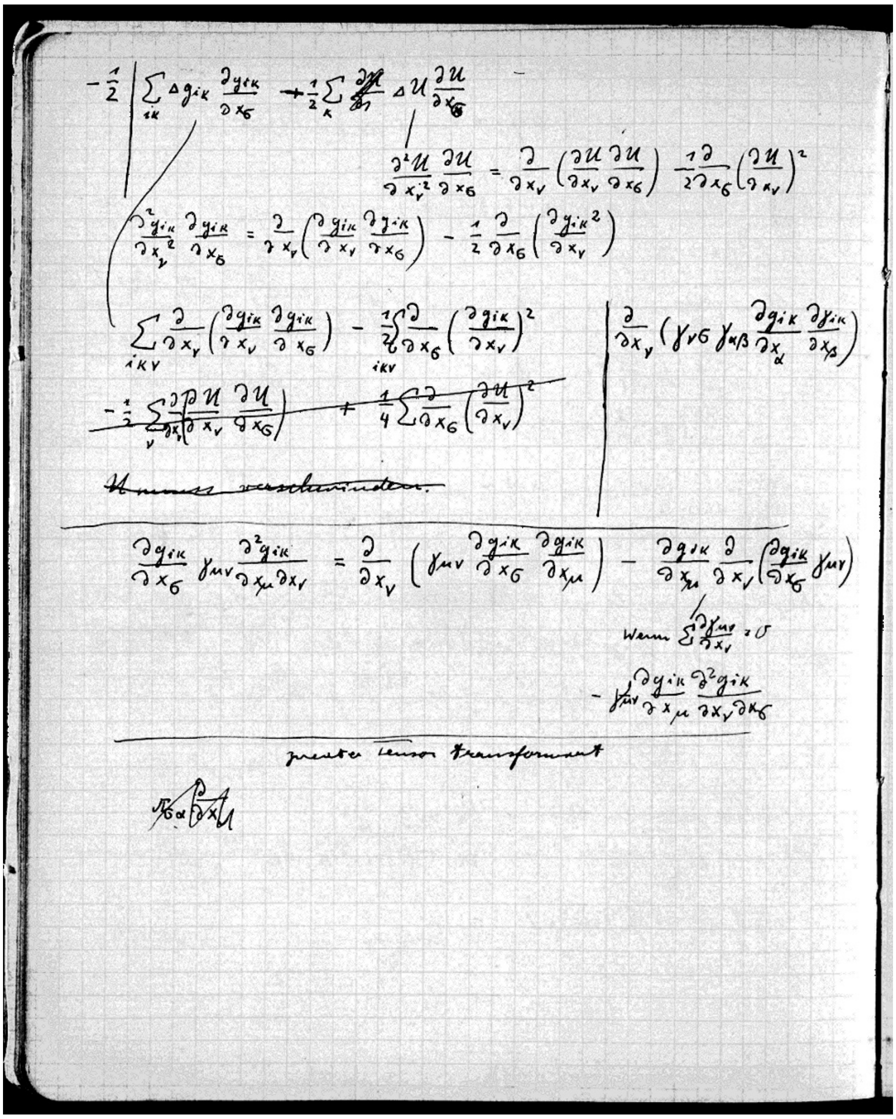
Energiedichte $\left(g_{41}\frac{dx}{d\tau}\frac{dt}{d\tau} + \qquad\qquad \right)\rho_0\sqrt{G}.$

Tensor der materiellen Strömung $T_{i\kappa} = \rho_0 \frac{dx_i}{d\tau}\frac{dx_\kappa}{d\tau}$

Hieraus gemischter Tensor $\quad T'_{\nu\kappa} = -\sum_i g_{\nu i}\langle dx \rangle T_{i\kappa}$ Sp. - Energie - Tensor.

$$\sum_\kappa \frac{\partial}{\partial x_\kappa}(\sqrt{G} g_{\nu i} T_{i\kappa})$$

[p. 21 L]



$$-\frac{1}{2}\left|\sum_{ik}\Delta g_{ik}\frac{\partial g_{ik}}{\partial x_\sigma} + \frac{1}{2}\sum_{\kappa}\cancel{\frac{\partial \mathfrak{U}}{\partial}}\Delta\mathfrak{U}\frac{\partial\mathfrak{U}}{\partial x_\sigma}\right.$$

$$\frac{\partial^2\mathfrak{U}}{\partial x_\nu^2}\frac{\partial\mathfrak{U}}{\partial x_\sigma} = \frac{\partial}{\partial x_\nu}\left(\frac{\partial\mathfrak{U}}{\partial x_\nu}\frac{\partial\mathfrak{U}}{\partial x_\sigma}\right) - \frac{1}{2}\frac{\partial}{\partial x_\sigma}\left(\frac{\partial\mathfrak{U}}{\partial x_\nu}\right)^2$$

$$\frac{\partial^2 g_{ik}}{\partial x_\nu^2}\frac{\partial g_{ik}}{\partial x_\sigma} = \frac{\partial}{\partial x_\nu}\left(\frac{\partial g_{ik}}{\partial x_\nu}\frac{\partial g_{ik}}{\partial x_\sigma}\right) - \frac{1}{2}\frac{\partial}{\partial x_\sigma}\left(\frac{\partial g_{ik}}{\partial x_\nu}^2\right)$$

$$\sum_{ik\nu}\frac{\partial}{\partial x_\nu}\left(\frac{\partial g_{ik}}{\partial x_\nu}\frac{\partial g_{ik}}{\partial x_\sigma}\right) - \frac{1}{2}\sum_{ik\nu}\frac{\partial}{\partial x_\sigma}\left(\frac{\partial g_{ik}}{\partial x_\nu}\right)^2 \qquad \frac{\partial}{\partial x_\nu}\left(\gamma_{\nu\sigma}\gamma_{\alpha\beta}\frac{\partial g_{ik}}{\partial x_\alpha}\frac{\partial g_{ik}}{\partial x_\beta}\right)$$

$$\cancel{-\frac{1}{2}\sum_{\nu}\frac{\partial}{\partial x_\nu}\left(\frac{\partial\mathfrak{U}}{\partial x_\nu}\frac{\partial\mathfrak{U}}{\partial x_\sigma}\right) + \frac{1}{4}\sum\frac{\partial}{\partial x_\sigma}\left(\frac{\partial\mathfrak{U}}{\partial x_\nu}\right)^2}$$

~~U muss verschwinden.~~

$$\frac{\partial g_{ik}}{\partial x_\sigma}\gamma_{\mu\nu}\frac{\partial^2 g_{ik}}{\partial x_\mu\partial x_\nu} = \frac{\partial}{\partial x_\nu}\left(\gamma_{\mu\nu}\frac{\partial g_{ik}}{\partial x_\sigma}\frac{\partial g_{ik}}{\partial x_\mu}\right) - \frac{\partial g_{ik}}{\partial x_\mu}\frac{\partial}{\partial x_\nu}\left(\frac{\partial g_{ik}}{\partial x_\sigma}\gamma_{\mu\nu}\right)$$

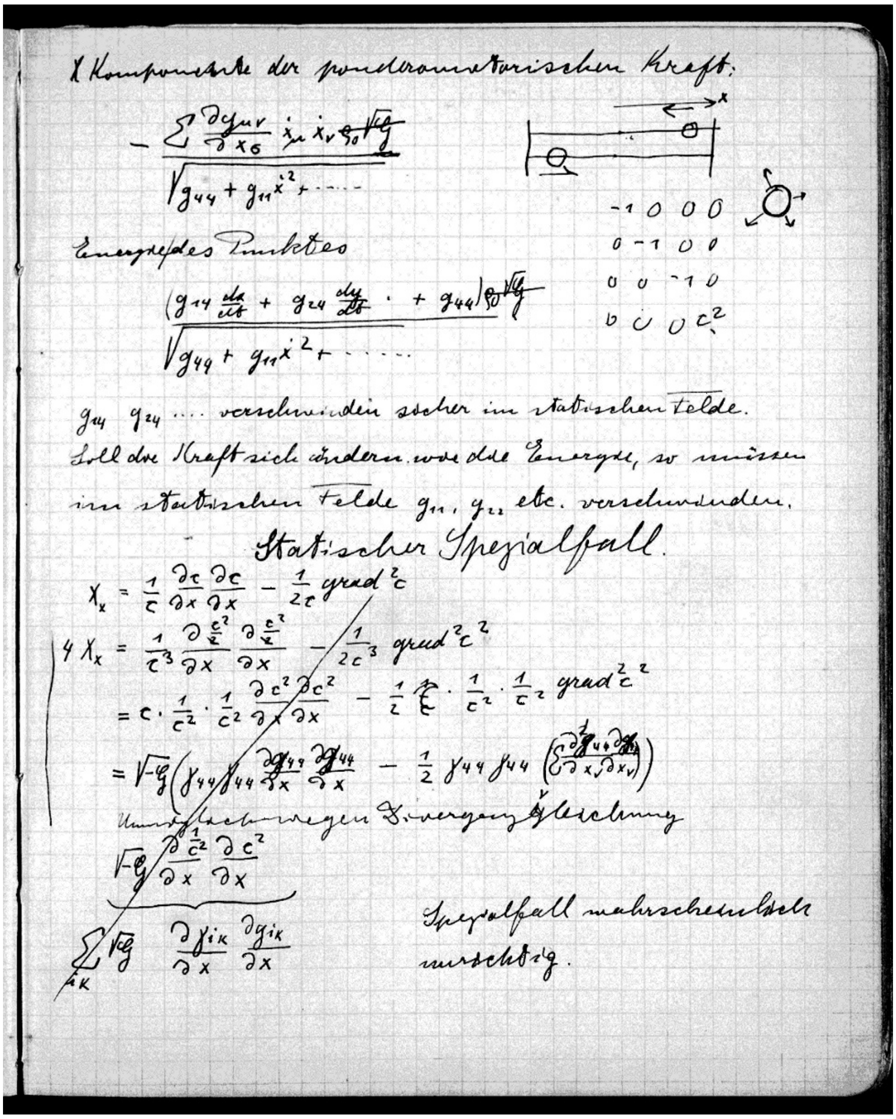
Wenn $\sum\frac{\partial\gamma_{\mu\nu}}{\partial x_\nu} = 0$

$$-\gamma_{\mu\nu}\frac{\partial g_{ik}}{\partial x_\mu}\frac{\partial^2 g_{ik}}{\partial x_\nu\partial x_\sigma}$$

zweiter Tensor transformiert

$$\cancel{\gamma_{\sigma\alpha}\frac{\partial}{\partial x_\alpha}\mathfrak{U}}$$

[p. 21 R]



X Komponente der ponderomotorischen Kraft.

$$-\frac{\sum \frac{\partial g_{\mu\nu}}{\partial x_\sigma}\dot{x}_\mu \dot{x}_\nu \varrho_0 \sqrt{g}}{\sqrt{g_{44}+g_{11}\dot{x}^2+\cdots}}$$

$$\begin{matrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & c^2 \end{matrix}$$

Energie des Punktes

$$\frac{\left(g_{14}\frac{dx}{dt} + g_{24}\frac{dy}{dt}\cdot + g_{44}\right)\varrho_0\sqrt{g}}{\sqrt{g_{44}+g_{11}\dot{x}^2+\cdots}}$$

g_{14} g_{24} .... verschwinden sicher im statischen Felde.
Soll die Kraft sich ändern wie die Energie, so müssen im statischen Felde g_{11}, g_{22} etc. verschwinden.

Statischer Spezialfall.

$$X_x = \frac{1}{c}\frac{\partial c}{\partial x}\frac{\partial c}{\partial x} - \frac{1}{2c}\,\mathrm{grad}^2 c$$

$$4X_x = \frac{1}{c^3}\frac{\partial c^2}{\partial x}\frac{\partial c^2}{\partial x} - \frac{1}{2c^3}\,\mathrm{grad}^2 c^2$$

$$= c\cdot\frac{1}{c^2}\cdot\frac{1}{c^2}\frac{\partial c^2}{\partial x}\frac{\partial c^2}{\partial x} - \frac{1}{2}c\cdot\frac{1}{c^2}\cdot\frac{1}{c^2}\,\mathrm{grad}^2 c^2$$

$$= \sqrt{-g}\left(\gamma_{44}\gamma_{44}\frac{\partial g_{44}}{\partial x}\frac{\partial g_{44}}{\partial x} - \frac{1}{2}\gamma_{44}\gamma_{44}\left(\sum \frac{\partial g_{44}}{\partial x_\nu}\frac{\partial g_{44}}{\partial x_\nu}\right)\right)$$

Unmöglich wegen Divergenz Gleichung

$$\underbrace{\sqrt{-g}\,\frac{\partial \frac{1}{c^2}}{\partial x}\frac{\partial c^2}{\partial x}}$$

$$\sum_{ik}\sqrt{g}\,\frac{\partial \gamma_{ik}}{\partial x}\frac{\partial g_{ik}}{\partial x}$$

Spezialfall wahrscheinlich unrichtig.

[p. 21 L]

$$-\frac{1}{2}\left|\ \sum_{i\kappa}\Delta g_{i\kappa}\frac{\partial g_{i\kappa}}{\partial x_\sigma}\langle+\rangle-\frac{1}{2}\sum_{\kappa}\langle\frac{\partial U}{\partial}\rangle\Delta U\frac{\partial U}{\partial x_\sigma}\right.$$

$$\frac{\partial^2 U}{\partial x_\nu^2}\frac{\partial U}{\partial x_\sigma} = \frac{\partial}{\partial x_\nu}\left(\frac{\partial U}{\partial x_\nu}\frac{\partial U}{\partial x_\sigma}\right)-\frac{1}{2}\frac{\partial}{\partial x_\sigma}\left(\frac{\partial U}{\partial x_\nu}\right)^2$$

$$\frac{\partial^2 g_{i\kappa}}{\partial x_\nu^2}\frac{\partial g_{i\kappa}}{\partial x_\sigma} = \frac{\partial}{\partial x_\nu}\left(\frac{\partial g_{i\kappa}}{\partial x_\nu}\frac{\partial g_{i\kappa}}{\partial x_\sigma}\right)-\frac{1}{2}\frac{\partial}{\partial x_\sigma}\left(\frac{\partial g_{i\kappa^2}}{\partial x_\nu}\right)$$

$$\sum_{i\kappa\nu}\frac{\partial}{\partial x_\nu}\left(\frac{\partial g_{i\kappa}}{\partial x_\nu}\frac{\partial g_{i\kappa}}{\partial x_\sigma}\right)-\frac{1}{2}\sum_{i\kappa\nu}\frac{\partial}{\partial x_\sigma}\left(\frac{\partial g_{i\kappa}}{\partial x_\nu}\right)^2$$

$$\cancel{-\frac{1}{2}\sum_{\nu}\frac{\partial}{\partial x_\nu}\left(\frac{\partial U}{\partial x_\nu}\frac{\partial U}{\partial x_\sigma}\right)+\frac{1}{4}\sum\frac{\partial}{\partial x_\sigma}\left(\frac{\partial U}{\partial x_\nu}\right)^2}$$

< U muss verschwinden. >

$$\frac{\partial}{\partial x_\nu}\left(\gamma_{\nu\sigma}\gamma_{\alpha\beta}\frac{\partial g_{i\kappa}}{\partial x_\alpha}\frac{\partial \gamma_{i\kappa}}{\partial x_\beta}\right)$$

$$\frac{\partial g_{i\kappa}}{\partial x_\sigma}\gamma_{\mu\nu}\frac{\partial^2 g_{i\kappa}}{\partial x_\mu \partial x_\nu} = \frac{\partial}{\partial x_\nu}\left(\gamma_{\mu\nu}\frac{\partial g_{i\kappa}}{\partial x_\sigma}\frac{\partial g_{i\kappa}}{\partial x_\mu}\right)-\frac{\partial g_{i\kappa}}{\partial x_\mu}\frac{\partial}{\partial x_\nu}\left(\frac{\partial g_{i\kappa}}{\partial x_\sigma}\gamma_{\mu\nu}\right)$$

Wenn $\sum\frac{\partial \gamma_{\mu\nu}}{\partial x_\nu} = 0$

$$-\gamma_{\mu\nu}\frac{\partial g_{i\kappa}}{\partial x_\mu}\frac{\partial^2 g_{i\kappa}}{\partial x_\nu \partial x_\sigma}$$

---

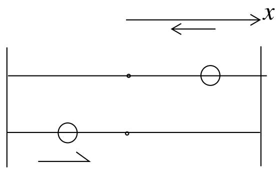
zweiter Tensor transformiert

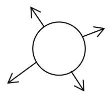
$$\langle\pi_{\sigma\alpha}\frac{\partial}{\partial x_\alpha}\rangle$$

[p. 21 R]

X Komponente der ponderomotischen Kraft:

$$-\frac{\sum \frac{\partial g_{\mu\nu}}{\partial x_\sigma}\dot{x}_\mu \dot{x}_\nu \langle \rho_0 \sqrt{G}\rangle}{\sqrt{g_{44}+g_{11}\dot{x}^2+\ \cdot\ \cdot\ \cdot}}$$





Energie des Punktes

$$\frac{\left(g_{14}\frac{dx}{dt}+g_{24}\frac{dy}{dt}\cdot\ +g_{44}\right)\langle \rho_0 \sqrt{G}\rangle}{\sqrt{g_{44}+g_{11}\dot{x}^2+\ \cdot\ \cdot\ \cdot}}$$

$$\begin{matrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & c^2 \end{matrix}$$

g_{14} g_{24} .... verschwinden sicher im statischen Felde. Soll die Kraft sich ändern wie die Energie, so müssen im statischen Felde g_{11}, g_{22} etc. verschwinden.

Statischer Spezialfall.

$$X_x = \frac{1}{c}\frac{\partial c}{\partial x}\frac{\partial c}{\partial x} - \frac{1}{2c}\mathrm{grad}^2 c$$

$$4X_x = \frac{1}{c^3}\frac{\partial \frac{c^2}{\langle 2\rangle}}{\partial x}\frac{\partial \frac{c^2}{\langle 2\rangle}}{\partial x} - \frac{1}{2c^3}\mathrm{grad}^2 c^2$$

$$= c\cdot\frac{1}{c^2}\cdot\frac{1}{c^2}\frac{\partial c^2}{\partial x}\frac{\partial c^2}{\partial x} - \frac{1}{2}\left\langle\frac{1}{c}\right\rangle c\cdot\frac{1}{c^2}\cdot\frac{1}{c^2}\mathrm{grad}^2 c^2$$

$$= \sqrt{-G}\left(\gamma_{44}\gamma_{44}\frac{\partial\langle\gamma\rangle g_{44}}{\partial x}\frac{\partial\langle\gamma\rangle g_{44}}{\partial x} - \frac{1}{2}\gamma_{44}\gamma_{44}\left(\sum_\nu \frac{\partial^{\langle 2\rangle}\langle\gamma\rangle g_{44}}{\partial x_\nu}\frac{\langle\gamma\rangle\partial g_{44}}{\partial x_\nu}\right)\right)$$

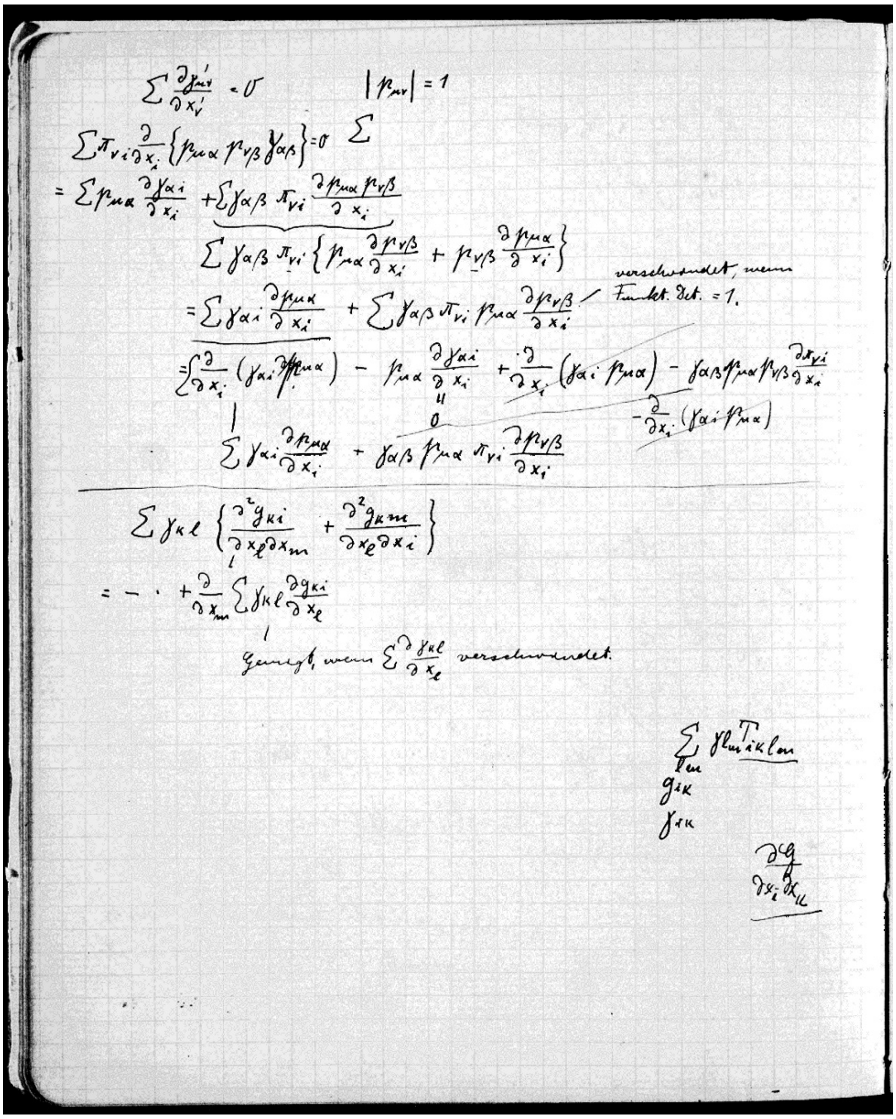
Unmöglich wegen Divergenzgleichung

$$\underbrace{\sqrt{-G}\frac{\partial\frac{1}{c^2}}{\partial x}\frac{\partial c^2}{\partial x}}$$

$$\sum_{i\kappa}\sqrt{G}\quad\frac{\partial\gamma_{i\kappa}}{\partial x}\frac{\partial g_{i\kappa}}{\partial x}$$

Spezialfall wahrscheinlich unrichtig.

[p. 22 L]



$$\sum \frac{\partial \gamma'_{\mu\nu}}{\partial x'_\nu} = 0 \qquad |p_{\mu\nu}| = 1$$

$$\sum \pi_{\nu i} \frac{\partial}{\partial x_i} \{ p_{\mu\alpha} p_{\nu\beta} \gamma_{\alpha\beta} \} = 0 \quad \sum$$

$$= \sum p_{\mu\alpha} \frac{\partial \gamma_{\alpha i}}{\partial x_i} + \underbrace{\sum \gamma_{\alpha\beta} \pi_{\nu i} \frac{\partial p_{\mu\alpha} p_{\nu\beta}}{\partial x_i}}$$

$$\sum \gamma_{\alpha\beta} \pi_{\nu i} \left\{ p_{\mu\alpha} \frac{\partial p_{\nu\beta}}{\partial x_i} + p_{\nu\beta} \frac{\partial p_{\mu\alpha}}{\partial x_i} \right\}$$

$$= \sum \gamma_{\alpha i} \frac{\partial p_{\mu\alpha}}{\partial x_i} + \sum \gamma_{\alpha\beta} \pi_{\nu i} p_{\mu\alpha} \frac{\partial p_{\nu\beta}}{\partial x_i}$$ / verschwindet, wenn Funkt. Det. = 1.

$$= \sum \frac{\partial}{\partial x_i} (\gamma_{\alpha i} p_{\mu\alpha}) - p_{\mu\alpha} \frac{\partial \gamma_{\alpha i}}{\partial x_i} \cancel{+ \frac{\partial}{\partial x_i} (\gamma_{\alpha i} p_{\mu\alpha}) - \gamma_{\alpha\beta} p_{\mu\alpha} p_{\nu\beta} \frac{\partial \pi_{\nu i}}{\partial x_i}}$$

$$\frac{\partial \gamma_{\alpha i}}{\partial x_i} = 0 \qquad \cancel{- \frac{\partial}{\partial x_i} (\gamma_{\alpha i} p_{\mu\alpha})}$$

$$\sum \gamma_{\alpha i} \frac{\partial p_{\mu\alpha}}{\partial x_i} + \gamma_{\alpha\beta} p_{\mu\alpha} \pi_{\nu i} \frac{\partial p_{\nu\beta}}{\partial x_i}$$

$$\sum \gamma_{\kappa l} \left\{ \frac{\partial^2 g_{\kappa i}}{\partial x_l \partial x_m} + \frac{\partial^2 g_{\kappa m}}{\partial x_l \partial x_i} \right\}$$

$$= - \cdot + \frac{\partial}{\partial x_m} \sum \gamma_{\kappa l} \frac{\partial g_{\kappa i}}{\partial x_l}$$

Genügt, wenn $\sum \frac{\partial \gamma_{\kappa l}}{\partial x_l}$ verschwindet.

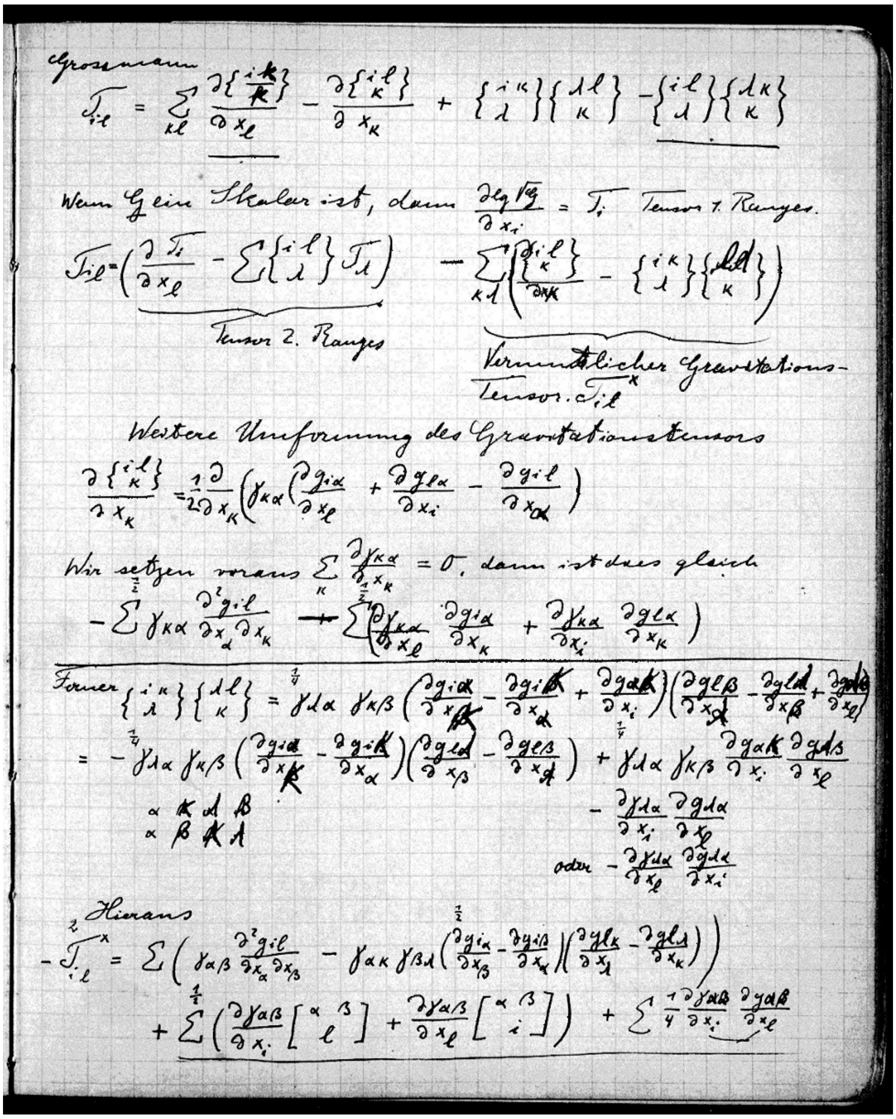
$$\sum_{lm} \gamma_{lm} T_{iklm}$$

g_{ik}

γ_{ik}

$$\frac{\partial g}{\partial x_i \partial x_k}$$

[p. 22 R]



Grossmann

$$\mathcal{T}_{il} = \sum_{\kappa l} \frac{\partial \begin{Bmatrix} i\,\kappa \\ \kappa \end{Bmatrix}}{\partial x_l} - \frac{\partial \begin{Bmatrix} i\,l \\ \kappa \end{Bmatrix}}{\partial x_\kappa} + \begin{Bmatrix} i\,\kappa \\ \lambda \end{Bmatrix}\begin{Bmatrix} \lambda\,l \\ \kappa \end{Bmatrix} - \begin{Bmatrix} i\,l \\ \lambda \end{Bmatrix}\begin{Bmatrix} \lambda\,\kappa \\ \kappa \end{Bmatrix}$$

Wenn G ein Skalar ist, dann $\dfrac{\partial \lg \sqrt{G}}{\partial x_i} = \mathcal{T}_i$ Tensor 1. Ranges.

$$\mathcal{T}_{il} = \underbrace{\left(\frac{\partial \mathcal{T}_i}{\partial x_l} - \sum \begin{Bmatrix} i\,l \\ \lambda \end{Bmatrix} \mathcal{T}_\lambda \right)}_{\text{Tensor 2. Ranges}} - \underbrace{\sum_{\kappa\lambda} \left(\frac{\partial \begin{Bmatrix} i\,l \\ \kappa \end{Bmatrix}}{\partial x_\kappa} - \begin{Bmatrix} i\,\kappa \\ \lambda \end{Bmatrix}\begin{Bmatrix} \lambda\,l \\ \kappa \end{Bmatrix} \right)}_{\text{Vermutlicher Gravitations-Tensor } \mathcal{T}_{il}^{x}}$$

Weitere Umformung des Gravitationstensors

$$\frac{\partial \begin{Bmatrix} i\,l \\ \kappa \end{Bmatrix}}{\partial x_\kappa} = \frac{1}{2} \frac{\partial}{\partial x_\kappa} \left(\gamma_{\kappa\alpha} \left(\frac{\partial g_{i\alpha}}{\partial x_l} + \frac{\partial g_{l\alpha}}{\partial x_i} - \frac{\partial g_{il}}{\partial x_\alpha} \right) \right)$$

Wir setzen voraus $\sum_\kappa \dfrac{\partial \gamma_{\kappa\alpha}}{\partial x_\kappa} = 0$, dann ist dies gleich

$$-\overset{\frac{1}{2}}{\sum} \gamma_{\kappa\alpha} \frac{\partial^2 g_{il}}{\partial x_\alpha \partial x_\kappa} + \overset{\frac{1}{2}}{\sum} \left(\frac{\partial \gamma_{\kappa\alpha}}{\partial x_l} \frac{\partial g_{i\alpha}}{\partial x_\kappa} + \frac{\partial \gamma_{\kappa\alpha}}{\partial x_i} \frac{\partial g_{l\alpha}}{\partial x_\kappa} \right)$$

Ferner

$$\begin{Bmatrix} i\,\kappa \\ \lambda \end{Bmatrix}\begin{Bmatrix} \lambda\,l \\ \kappa \end{Bmatrix} = \overset{\frac{1}{4}}{} \gamma_{\lambda\alpha} \gamma_{\kappa\beta} \left(\frac{\partial g_{i\alpha}}{\partial x_\kappa} - \frac{\partial g_{i\kappa}}{\partial x_\alpha} + \frac{\partial g_{\alpha\kappa}}{\partial x_i} \right) \left(\frac{\partial g_{l\beta}}{\partial x_\lambda} - \frac{\partial g_{l\lambda}}{\partial x_\beta} + \frac{\partial g_{\beta\lambda}}{\partial x_l} \right)$$

$$= -\overset{\frac{1}{4}}{} \gamma_{\lambda\alpha} \gamma_{\kappa\beta} \left(\frac{\partial g_{i\alpha}}{\partial x_\kappa} - \frac{\partial g_{i\kappa}}{\partial x_\alpha} \right) \left(\frac{\partial g_{l\lambda}}{\partial x_\beta} - \frac{\partial g_{l\beta}}{\partial x_\lambda} \right) + \overset{\frac{1}{4}}{} \gamma_{\lambda\alpha} \gamma_{\kappa\beta} \frac{\partial g_{\alpha\kappa}}{\partial x_i} \frac{\partial g_{\lambda\beta}}{\partial x_l}$$

$\alpha \ \kappa \ \lambda \ \beta$

$\alpha \ \beta \ \kappa \ \lambda$

$$- \frac{\partial \gamma_{\lambda\alpha}}{\partial x_i} \frac{\partial g_{\lambda\alpha}}{\partial x_l}$$

oder $- \dfrac{\partial \gamma_{\lambda\alpha}}{\partial x_l} \dfrac{\partial g_{\lambda\alpha}}{\partial x_i}$

Hieraus

$$-\overset{2}{}\mathcal{T}_{il}^{x} = \sum \left(\gamma_{\alpha\beta} \frac{\partial^2 g_{il}}{\partial x_\alpha \partial x_\beta} - \overset{\frac{1}{2}}{} \gamma_{\alpha\kappa} \gamma_{\beta\lambda} \left(\frac{\partial g_{i\alpha}}{\partial x_\beta} - \frac{\partial g_{i\beta}}{\partial x_\alpha} \right) \left(\frac{\partial g_{l\kappa}}{\partial x_\lambda} - \frac{\partial g_{l\lambda}}{\partial x_\kappa} \right) \right)$$

$$+ \overset{\frac{1}{2}}{\sum} \left(\frac{\partial \gamma_{\alpha\beta}}{\partial x_i} \begin{bmatrix} \alpha\,\beta \\ l \end{bmatrix} + \frac{\partial \gamma_{\alpha\beta}}{\partial x_l} \begin{bmatrix} \alpha\,\beta \\ i \end{bmatrix} \right) + \sum \frac{1}{4} \frac{\partial \gamma_{\alpha\beta}}{\partial x_i} \frac{\partial g_{\alpha\beta}}{\partial x_l}$$

[p. 22 L]

$$\sum \frac{\partial \gamma_{\mu\nu}'}{\partial x_\nu'} = 0 \qquad |p_{\mu\nu}| = 1$$

$$\sum$$

$$\sum \pi_{\nu i} \frac{\partial}{\partial x_i} \{ p_{\mu\alpha} p_{\nu\beta} \gamma_{\alpha\beta} \} = 0$$

$$= \sum p_{\mu\alpha} \frac{\partial \gamma_{\alpha i}}{\partial x_i} + \underbrace{\sum \gamma_{\alpha\beta} \pi_{\nu i} \frac{\partial p_{\mu\alpha} p_{\nu\beta}}{\partial x_i}}$$

$$\sum \gamma_{\alpha\beta} \pi_{\underline{\nu} i} \left\{ p_{\mu\alpha} \frac{\partial p_{\nu\beta}}{\partial x_i} + p_{\underline{\nu}\beta} \frac{\partial p_{\mu\alpha}}{\partial x_i} \right\}$$

$$= \underline{\sum \gamma_{\alpha i} \frac{\partial p_{\mu\alpha}}{\partial x_i}} + \sum \gamma_{\alpha\beta} \pi_{\nu i} p_{\mu\alpha} \frac{\partial p_{\nu\beta}}{\partial x_i}$$ verschwindet, wenn Funkt. Det. = 1.

$$= \sum \frac{\partial}{\partial x_i} (\gamma_{\alpha i} p_{\mu\alpha}) - \underset{\substack{\shortparallel \\ 0}}{p_{\mu\alpha} \frac{\partial \gamma_{\alpha i}}{\partial x_i}} + \cancel{\frac{\partial}{\partial x_i} (\gamma_{\alpha i} p_{\mu\alpha})} - \gamma_{\alpha\beta} p_{\mu\alpha} p_{\nu\beta} \frac{\partial \pi_{\nu i}}{\partial x_i}$$

$$\cancel{- \frac{\partial}{\partial x_i} (\gamma_{\alpha i} p_{\mu\alpha})}$$

$$\sum \gamma_{\alpha i} \frac{\partial p_{\mu\alpha}}{\partial x_i} + \cancel{\gamma_{\alpha\beta} p_{\mu\alpha} \pi_{\nu i} \frac{\partial p_{\nu\beta}}{\partial x_i}}$$

$$\sum \gamma_{\kappa l} \left\{ \frac{\partial^2 g_{\kappa i}}{\partial x_l \partial x_m} + \frac{\partial^2 g_{\kappa m}}{\partial x_l \partial x_i} \right\}$$

$$= - \cdot + \frac{\partial}{\partial x_m} \sum \gamma_{\kappa l} \frac{\partial g_{\kappa i}}{\partial x_l}$$

Genugt, wenn $\sum \frac{\partial \gamma_{\kappa l}}{\partial x_l}$ verschwindet.

$$\sum_{lm} \gamma_{lm} \underline{T_{i\kappa lm}}$$

$g_{i\kappa}$

$\gamma_{i\kappa}$

$$\underline{\frac{\partial G}{\partial x_i \partial x_\kappa}}$$

[p. 22 R]

Grossmann

$$T_{il} = \sum_{\kappa l} \frac{\partial \left\{ {i\,\kappa \atop \kappa} \right\}}{\underline{\partial x_l}} - \frac{\partial \left\{ {i\,l \atop \kappa} \right\}}{\partial x_\kappa} + \left\{ {i\,\kappa \atop \lambda} \right\}\left\{ {\lambda\, l \atop \kappa} \right\} - \underline{\left\{ {i\,l \atop \lambda} \right\}\left\{ {\lambda\,\kappa \atop \kappa} \right\}}$$

Wenn $\underline{G}$ ein Skalar ist, dann $\frac{\partial \lg\sqrt{G}}{\partial x_i} = T_i$ Tensor 1. Ranges.

$$T_{il} = \underbrace{\left(\frac{\partial T_i}{\partial x_l} - \sum \left\{ {i\,l \atop \lambda} \right\} T_\lambda\right)}_{\text{Tensor 2. Ranges}} \quad \underbrace{- \sum_{\kappa l}\left(\frac{\partial \left\{ {i\,l \atop \kappa} \right\}}{\partial x_\kappa} - \left\{ {i\,\kappa \atop \lambda} \right\}\left\{ {l\,\lambda \atop \kappa} \right\}\right)}_{\text{Vermutlicher Gravitations-Tensor. } T_{il}^x}$$

Weitere Umformung des Gravitationstensors

$$\frac{\partial \left\{ {i\,l \atop \kappa} \right\}}{\partial x_\kappa} = \frac{1}{2} \frac{\partial}{\partial x_\kappa}\left(\gamma_{\kappa\alpha}\left(\frac{\partial g_{i\alpha}}{\partial x_l} + \frac{\partial g_{l\alpha}}{\partial x_i} - \frac{\partial g_{il}}{\partial x_\alpha}\right)\right)$$

Wir setzen voraus $\sum_\kappa \frac{\partial \gamma_{\kappa\alpha}}{\partial x_\kappa} = 0$, dann ist dies gleich

$$-\overset{\frac{1}{2}}{\sum}\gamma_{\kappa\alpha}\frac{\partial^2 g_{il}}{\partial x_\alpha \partial x_\kappa} - \overset{\frac{1}{2}}{\sum}\left(\frac{\partial \gamma_{\kappa\alpha}}{\partial x_l}\frac{\partial g_{i\alpha}}{\partial x_\kappa} + \frac{\partial \gamma_{\kappa\alpha}}{\partial x_i}\frac{\partial g_{l\alpha}}{\partial x_\kappa}\right)$$

Ferner

$$\left\{ {i\,\kappa \atop \lambda} \right\}\left\{ {\lambda\, l \atop \kappa} \right\} = \overset{\frac{1}{4}}{}\gamma_{\lambda\alpha}\gamma_{\kappa\beta}\left(\frac{\partial g_{i\alpha}}{\partial x_\kappa} - \frac{\partial g_{i\kappa}}{\partial x_\alpha} + \frac{\partial g_{\alpha\kappa}}{\partial x_i}\right)\left(\frac{\partial g_{l\beta}}{\partial x_\lambda} - \frac{\partial g_{l\lambda}}{\partial x_\beta} + \frac{\partial g_{\lambda\beta}}{\partial x_l}\right)$$

$$= -\overset{\frac{1}{4}}{}\gamma_{\lambda\alpha}\gamma_{\kappa\beta}\left(\frac{\partial g_{i\alpha}}{\partial x_\kappa} - \frac{\partial g_{i\kappa}}{\partial x_\alpha}\right)\left(\frac{\partial g_{l\lambda}}{\partial x_\beta} - \frac{\partial g_{l\beta}}{\partial x_\lambda}\right) + \overset{\frac{1}{4}}{}\gamma_{\lambda\alpha}\gamma_{\kappa\beta}\frac{\partial g_{\alpha\kappa}}{\partial x_i}\frac{\partial g_{\lambda\beta}}{\partial x_l}$$

$\alpha\ \kappa\ \lambda\ \beta$

$\alpha\ \beta\ \kappa\ \lambda$

$$-\frac{\partial \gamma_{\lambda\alpha}}{\partial x_i}\frac{\partial g_{\lambda\alpha}}{\partial x_l}$$

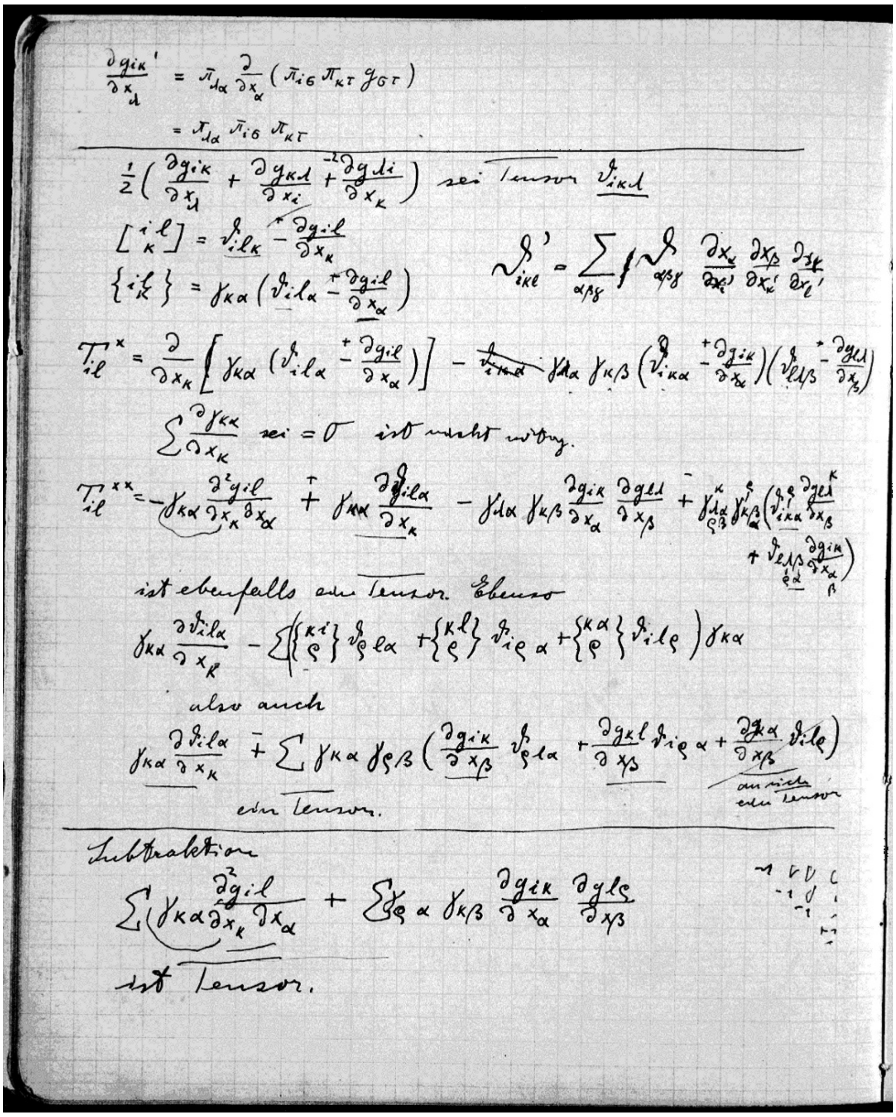
oder $-\frac{\partial \gamma_{\lambda\alpha}}{\partial x_l}\frac{\partial g_{\lambda\alpha}}{\partial x_i}$

Hieraus

$$-{}^{2}T_{il}^{x} = \sum\left(\gamma_{\alpha\beta}\frac{\partial^2 g_{il}}{\partial x_\alpha \partial x_\beta} - \gamma_{\alpha\kappa}\gamma_{\beta\lambda}\overset{\frac{1}{2}}{}\left(\frac{\partial g_{i\alpha}}{\partial x_\beta} - \frac{\partial g_{i\beta}}{\partial x_\alpha}\right)\left(\frac{\partial g_{l\kappa}}{\partial x_\lambda} - \frac{\partial g_{l\lambda}}{\partial x_k}\right)\right)$$

$$\underline{+\overset{\frac{1}{2}}{\sum}\left(\frac{\partial \gamma_{\alpha\beta}}{\partial x_i}\begin{bmatrix}\alpha\ \beta\\ l\end{bmatrix} + \frac{\partial \gamma_{\alpha\beta}}{\partial x_l}\begin{bmatrix}\alpha\ \beta\\ i\end{bmatrix}\right) + \sum\frac{1}{4}\frac{\partial \gamma_{\alpha\beta}}{\partial x_i}\frac{\partial g_{\alpha\beta}}{\partial x_l}}$$

[p. 23 L]



$$\frac{\partial g_{ik}'}{\partial x_\lambda} = \pi_{\lambda\alpha} \frac{\partial}{\partial x_\alpha}(\pi_{i\sigma}\pi_{\kappa\tau} g_{\sigma\tau})$$

$$= \pi_{\lambda\alpha}\pi_{i\sigma}\pi_{\kappa\tau}$$

$$\frac{1}{2}\left(\frac{\partial g_{i\kappa}}{\partial x_\lambda} + \frac{\partial g_{\kappa\lambda}}{\partial x_i} \mp \frac{\partial g_{\lambda i}}{\partial x_\kappa}\right) \text{ sei Tensor } \vartheta_{i\kappa\lambda}$$

$$\begin{bmatrix} il \\ \kappa \end{bmatrix} = \vartheta_{il\kappa} \mp \frac{\partial g_{il}}{\partial x_\kappa}$$

$$\vartheta_{ikl}' = \sum_{\alpha\beta\gamma} \vartheta_{\alpha\beta\gamma} \frac{\partial x_\alpha}{\partial x_i'} \frac{\partial x_\beta}{\partial x_\kappa'} \frac{\partial x_\gamma}{\partial x_l'}$$

$$\left\{ \begin{matrix} il \\ \kappa \end{matrix} \right\} = \gamma_{\kappa\alpha}\left(\vartheta_{il\alpha} \mp \frac{\partial g_{il}}{\partial x_\alpha}\right)$$

$$T_{il}^{x} = \frac{\partial}{\partial x_\kappa}\left[\gamma_{\kappa\alpha}\left(\vartheta_{il\alpha} \mp \frac{\partial g_{il}}{\partial x_\alpha}\right)\right] - \cancel{\vartheta_{i\kappa\alpha}}\, \gamma_{\lambda\alpha}\gamma_{\kappa\beta}\left(\vartheta_{i\kappa\alpha} \mp \frac{\partial g_{i\kappa}}{\partial x_\alpha}\right)\left(\vartheta_{l\lambda\beta} \mp \frac{\partial g_{l\lambda}}{\partial x_\beta}\right)$$

$$\sum \frac{\partial \gamma_{\kappa\alpha}}{\partial x_\kappa} \text{ sei } = 0 \text{ ist nicht nötig.}$$

$$T_{il}^{xx} = -\gamma_{\kappa\alpha}\frac{\partial^2 g_{il}}{\partial x_\kappa \partial x_\alpha} \pm \gamma_{\kappa\alpha}\frac{\partial \vartheta_{il\alpha}}{\partial x_\kappa} - \gamma_{\lambda\alpha}\gamma_{\kappa\beta}\frac{\partial g_{i\kappa}}{\partial x_\alpha}\frac{\partial g_{l\lambda}}{\partial x_\beta} \mp \gamma_{\lambda\alpha}\gamma_{\kappa\beta}\left(\vartheta_{i\kappa\alpha}\frac{\partial g_{l\lambda}}{\partial x_\beta} + \vartheta_{l\lambda\beta}\frac{\partial g_{i\kappa}}{\partial x_\alpha}\right)$$

ist ebenfalls ein Tensor. Ebenso

$$\gamma_{\kappa\alpha}\frac{\partial \vartheta_{il\alpha}}{\partial x_\kappa} - \sum\left(\left\{\begin{matrix}\kappa i\\ \varrho\end{matrix}\right\}\vartheta_{\varrho l\alpha} + \left\{\begin{matrix}\kappa l\\ \varrho\end{matrix}\right\}\vartheta_{i\varrho\alpha} + \left\{\begin{matrix}\kappa\alpha\\ \varrho\end{matrix}\right\}\vartheta_{il\varrho}\right)\gamma_{\kappa\alpha}$$

also auch

$$\gamma_{\kappa\alpha}\frac{\partial \vartheta_{il\alpha}}{\partial x_\kappa} \mp \sum \gamma_{\kappa\alpha}\gamma_{\varrho\beta}\left(\frac{\partial g_{i\kappa}}{\partial x_\beta}\vartheta_{\varrho l\alpha} + \frac{\partial g_{\kappa l}}{\partial x_\beta}\vartheta_{i\varrho\alpha} + \frac{\partial g_{\kappa\alpha}}{\partial x_\beta}\vartheta_{il\varrho}\right)$$

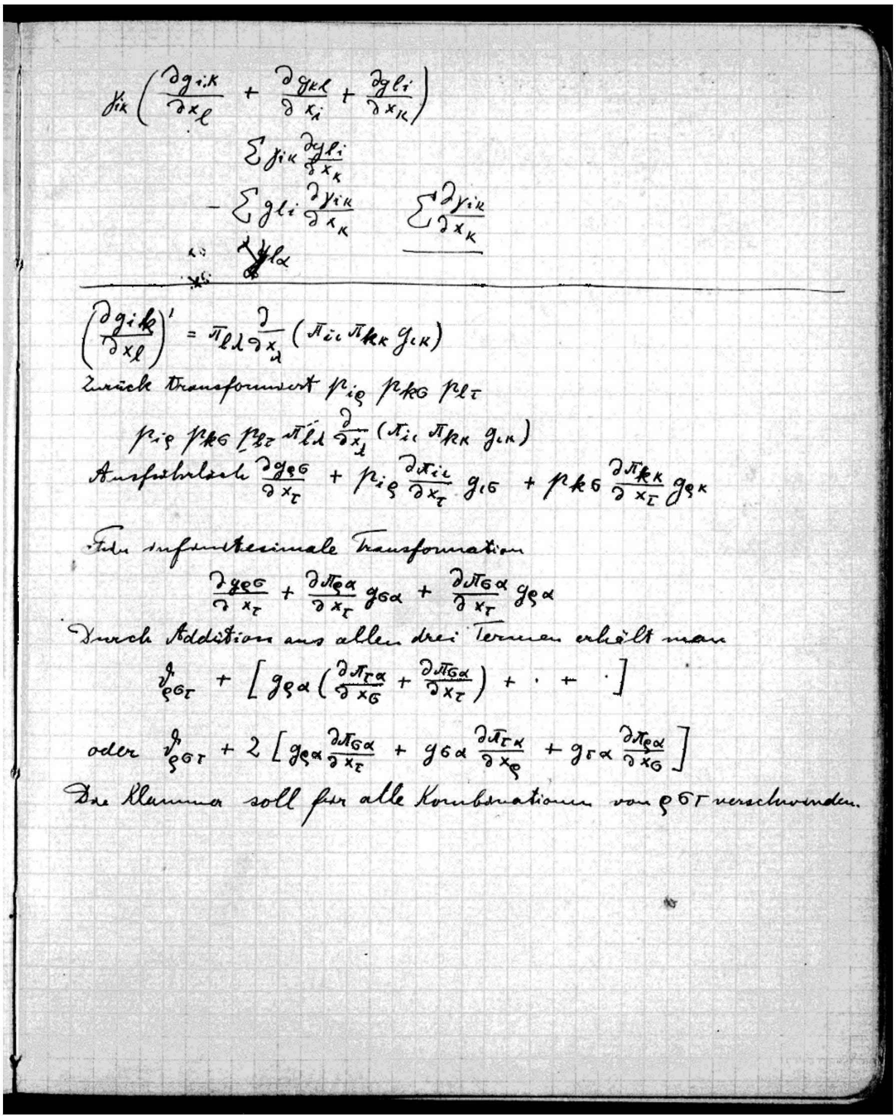
ein Tensor. an sich ein Tensor

Subtraktion

$$\sum \gamma_{\kappa\alpha}\frac{\partial^2 g_{il}}{\partial x_\kappa \partial x_\alpha} + \sum \gamma_{\varrho\alpha}\gamma_{\kappa\beta}\frac{\partial g_{i\kappa}}{\partial x_\alpha}\frac{\partial g_{l\varrho}}{\partial x_\beta}$$

ist Tensor.

[p. 23 R]



$$\gamma_{ik}\left(\frac{\partial g_{ik}}{\partial x_l} + \frac{\partial g_{kl}}{\partial x_i} + \frac{\partial g_{li}}{\partial x_k}\right)$$

$$\sum \gamma_{ik}\frac{\partial g_{li}}{\partial x_k}$$

$$-\sum g_{li}\frac{\partial \gamma_{ik}}{\partial x_k} \qquad \sum \frac{\partial \gamma_{ik}}{\partial x_k}$$

$$\gamma_{l\alpha}$$

$$\left(\frac{\partial g_{ik}}{\partial x_l}\right)' = \pi_{l\lambda}\frac{\partial}{\partial x_\lambda}(\pi_{i\iota}\pi_{k\kappa}g_{\iota\kappa})$$

Zurück transformiert $p_{i\varrho}\, p_{k\sigma}\, p_{l\tau}$

$$p_{i\varrho}\,p_{k\sigma}\,p_{l\tau}\,\pi_{l\lambda}\frac{\partial}{\partial x_\lambda}(\pi_{i\iota}\pi_{k\kappa}g_{\iota\kappa})$$

Ausführlich $\frac{\partial g_{\varrho\sigma}}{\partial x_\tau} + p_{i\varrho}\frac{\partial \pi_{i\iota}}{\partial x_\tau}g_{\iota\sigma} + p_{k\sigma}\frac{\partial \pi_{k\kappa}}{\partial x_\tau}g_{\varrho\kappa}$

Für infinitesimale Transformation

$$\frac{\partial g_{\varrho\sigma}}{\partial x_\tau} + \frac{\partial \pi_{\varrho\alpha}}{\partial x_\tau}g_{\sigma\alpha} + \frac{\partial \pi_{\sigma\alpha}}{\partial x_\tau}g_{\varrho\alpha}$$

Durch Addition aus allen drei Termen erhält man

$$\gamma_{\varrho\sigma\tau} + \left[g_{\varrho\alpha}\left(\frac{\partial \pi_{\tau\alpha}}{\partial x_\sigma} + \frac{\partial \pi_{\sigma\alpha}}{\partial x_\tau}\right) + \cdot + \cdot\right]$$

oder $\gamma_{\varrho\sigma\tau} + 2\left[g_{\varrho\alpha}\frac{\partial \pi_{\sigma\alpha}}{\partial x_\tau} + g_{\sigma\alpha}\frac{\partial \pi_{\tau\kappa}}{\partial x_\varrho} + g_{\tau\alpha}\frac{\partial \pi_{\varrho\alpha}}{\partial x_\sigma}\right]$

Die Klammer soll für alle Kombinationen von $\varrho\sigma\tau$ verschwinden.

[p. 23 L]

$$\frac{\partial g_{i\kappa}'}{\partial x_\lambda} = \pi_{\lambda\alpha}\frac{\partial}{\partial x_\alpha}(\pi_{i\sigma}\pi_{\kappa\tau}g_{\sigma\tau})$$
$$= \pi_{\lambda\alpha}\pi_{i\sigma}\pi_{\kappa\tau}$$

$$\frac{1}{2}\left(\frac{\partial g_{i\kappa}}{\partial x_\lambda} + \frac{\partial g_{\kappa\lambda}}{\partial x_i} \overset{-2}{+} \frac{\partial g_{\lambda i}}{\partial x_\kappa}\right) \quad \text{sei Tensor } \underline{\vartheta_{i\kappa\lambda}}$$

$$\begin{bmatrix} i & l \\ & \kappa \end{bmatrix} = \vartheta_{\underline{il\kappa}} \overset{\langle + \rangle}{-} \frac{\partial g_{il}}{\partial x_\kappa}$$

$$\vartheta'_{i\kappa l} = \sum_{\alpha\beta\gamma}\vartheta_{\alpha\beta\gamma}\frac{\partial x_\alpha}{\partial x_i'}\frac{\partial x_\beta}{\partial x_\kappa'}\frac{\partial x_\gamma}{\partial x_l'}$$

$$\begin{Bmatrix} i & l \\ & \kappa \end{Bmatrix} = \gamma_{\kappa\alpha}\left(\underline{\vartheta_{il\alpha}} \overset{\langle + \rangle}{-} \frac{\partial g_{il}}{\underline{\partial x_\alpha}}\right)$$

$$T_{il}^{x} = \frac{\partial}{\partial x_\kappa}\left[\gamma_{\kappa\alpha}\left(\vartheta_{il\alpha} \overset{+}{-} \frac{\partial g_{il}}{\partial x_\alpha}\right)\right] - \langle\vartheta_{i\kappa\alpha}\rangle\gamma_{\lambda\alpha}\gamma_{\kappa\beta}\left(\vartheta_{i\kappa\alpha} \overset{+}{-} \frac{\partial g_{i\kappa}}{\partial x_\alpha}\right)\left(\vartheta_{l\lambda\beta} \overset{+}{-} \frac{\partial g_{l\lambda}}{\partial x_\beta}\right)$$

$$\sum\frac{\partial\gamma_{\kappa\alpha}}{\partial x_\kappa} \quad \text{sei} = 0 \quad \text{ist nicht nötig.}$$

$$T_{il}^{xx} = -\gamma_{\kappa\alpha}\frac{\partial^2 g_{il}}{\partial x_\kappa\partial x_\alpha} \overset{+}{+} \underline{\gamma_{\kappa\alpha}\frac{\partial\vartheta_{il\alpha}}{\partial x_\kappa}} - \gamma_{\lambda\alpha}\gamma_{\kappa\beta}\frac{\partial g_{i\kappa}}{\partial x_\alpha}\frac{\partial g_{l\lambda}}{\partial x_\beta} \overset{-}{+} \gamma_{\underset{\rho\beta}{\lambda\alpha}}^{\kappa}\gamma_{\underset{\alpha}{\kappa\beta}}^{\rho}\left(\vartheta^{\rho}_{\underline{i\kappa\alpha}}\frac{\partial g_{l\lambda}^{\kappa}}{\partial x_\beta} + \vartheta_{\underset{\underline{\rho\alpha}}{l\lambda\beta}}\frac{\partial g_{i\kappa}}{\partial x_{\underset{\beta}{\alpha}}}\right)$$

ist ebenfalls ein Tensor. Ebenso

$$\gamma_{\kappa\alpha}\frac{\partial\vartheta_{il\alpha}}{\partial x_\kappa} - \sum\left(\begin{Bmatrix}\kappa & i \\ \rho\end{Bmatrix}\vartheta_{\rho l\alpha} + \begin{Bmatrix}\kappa & l \\ \rho\end{Bmatrix}\vartheta_{i\rho\alpha} + \begin{Bmatrix}\kappa & \alpha \\ \rho\end{Bmatrix}\vartheta_{il\rho}\right)\gamma_{\kappa\alpha}$$

also auch

$$\underline{\gamma_{\kappa\alpha}\frac{\partial\vartheta_{il\alpha}}{\partial x_\kappa}} \overset{-}{+} \sum\gamma_{\kappa\alpha}\gamma_{\rho\beta}\left(\frac{\partial g_{i\kappa}}{\underline{\partial x_\beta}}\vartheta_{\rho l\alpha} + \frac{\partial g_{\kappa l}}{\underline{\partial x_\beta}}\vartheta_{i\rho\alpha} + \cancel{\frac{\partial g_{\kappa\alpha}}{\underline{\partial x_\beta}}\vartheta_{il\rho}}\right)$$

ein Tensor. an sich ein Tensor

Subtraktion

$$\sum\gamma_{\kappa\alpha}\frac{\partial^2 g_{il}}{\partial x_\kappa\underline{\partial x_\alpha}} + \sum\gamma_{\rho\alpha}\gamma_{\kappa\beta}\frac{\partial g_{i\kappa}}{\partial x_\alpha}\frac{\partial g_{l\rho}}{\partial x_\beta}$$

$$\begin{matrix} -1 & 0 & 0 & 0 \\ & -1 & 0 & \cdot \\ & & -1 & \cdot \\ & & & +1 \end{matrix}$$

~~ist Tensor.~~

[p. 23 R]

$$\gamma_{i\kappa}\left(\frac{\partial g_{i\kappa}}{\partial x_l} + \frac{\partial g_{\kappa l}}{\partial x_i} + \frac{\partial g_{li}}{\partial x_\kappa}\right)$$

$$\sum \gamma_{i\kappa} \frac{\partial g_{li}}{\partial x_\kappa}$$

$$-\sum g_{li} \frac{\partial \gamma_{i\kappa}}{\partial x_\kappa} \qquad \underline{\sum \frac{\partial \gamma_{i\kappa}}{\partial x_\kappa}}$$

$$\langle g \rangle_{l\alpha}$$

$$\left(\frac{\partial g_{i\kappa}}{\partial x_l}\right)' = \pi_{l\lambda} \frac{\partial}{\partial x_\lambda}(\pi_{i\iota}\pi_{k\kappa} g_{\iota\kappa})$$

Zurück transformiert $p_{i\rho}p_{k\sigma}p_{l\tau}$

$$p_{i\rho}p_{k\sigma}\cancel{p_{l\tau}\pi_{l\lambda}} \frac{\partial}{\partial x_\lambda}(\pi_{i\iota}\pi_{k\kappa} g_{\iota\kappa})$$

Ausführlich $\dfrac{\partial g_{\rho\sigma}}{\partial x_\tau} + p_{i\rho}\dfrac{\partial \pi_{i\iota}}{\partial x_\tau} g_{\iota\sigma} + p_{k\sigma}\dfrac{\partial \pi_{k\kappa}}{\partial x_\tau} g_{\rho\kappa}$

Für infinitesimale Transformation

$$\frac{\partial g_{\rho\sigma}}{\partial x_\tau} + \frac{\partial \pi_{\rho\alpha}}{\partial x_\tau} g_{\sigma\alpha} + \frac{\partial \pi_{\sigma\alpha}}{\partial x_\tau} g_{\rho\alpha}$$

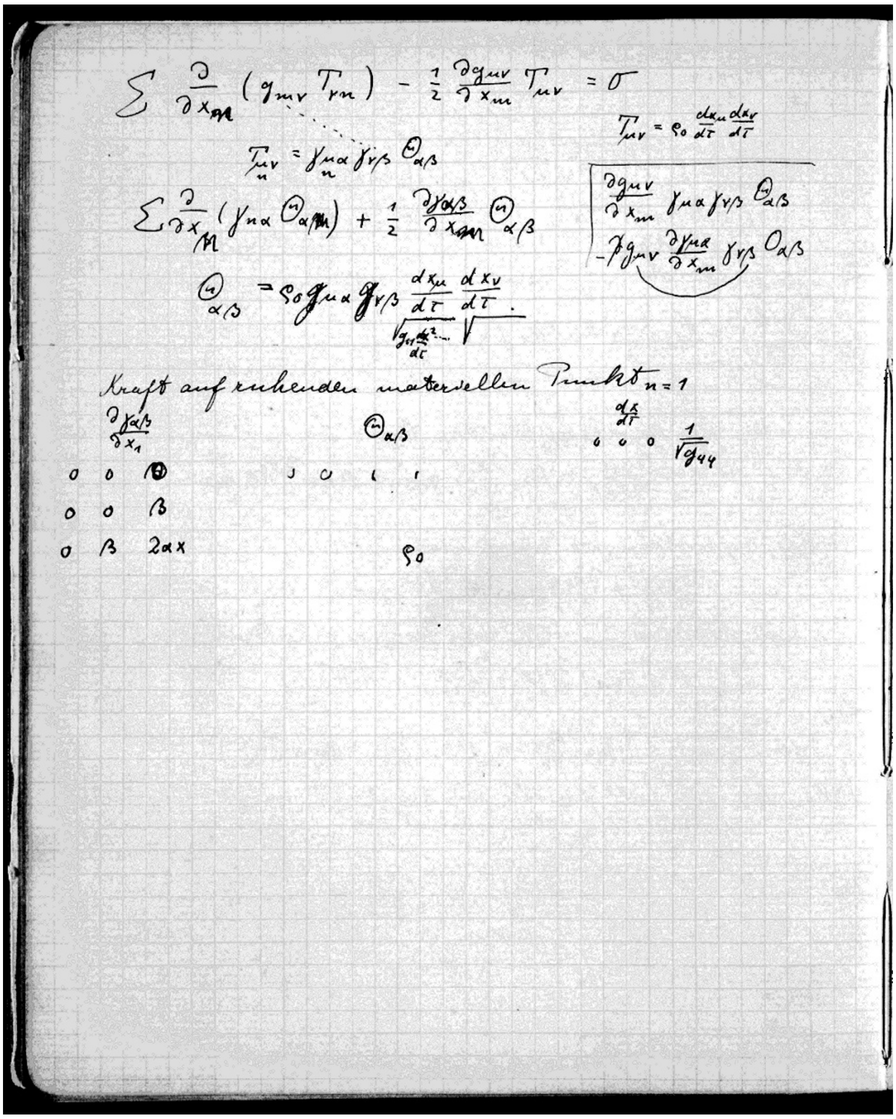
Durch Addition aus allen drei Termen erhält man

$$\vartheta_{\rho\sigma\tau} + \left[g_{\rho\alpha}\left(\frac{\partial \pi_{\tau\alpha}}{\partial x_\sigma} + \frac{\partial \pi_{\sigma\alpha}}{\partial x_\tau}\right) + \cdot + \cdot \right]$$

oder $\vartheta_{\rho\sigma\tau} + 2\left[g_{\rho\alpha}\dfrac{\partial \pi_{\sigma\alpha}}{\partial x_\tau} + g_{\sigma\alpha}\dfrac{\partial \pi_{\tau\alpha}}{\partial x_\rho} + g_{\tau\alpha}\dfrac{\partial \pi_{\rho\alpha}}{\partial x_\sigma} \right]$

Die Klammer soll für alle Kombinationen von $\rho\sigma\tau$ verschwinden.

[p. 24 L]



$$\sum \frac{\partial}{\partial x_m}(g_{m\nu} T_{\nu n}) - \frac{1}{2}\frac{\partial g_{\mu\nu}}{\partial x_m} T_{\mu\nu} = 0$$

$$T_{\mu\nu} = \rho_0 \frac{dx_\mu}{d\tau}\frac{dx_\nu}{d\tau}$$

$$\underset{n}{T_{\mu\nu}} = \underset{n}{\gamma_{\mu\alpha}}\gamma_{\nu\beta}\Theta_{\alpha\beta}$$

$$\sum \frac{\partial}{\partial x_m}(\gamma_{n\alpha}\Theta_{\alpha m}) + \frac{1}{2}\frac{\partial \gamma_{\alpha\beta}}{\partial x_m}\Theta_{\alpha\beta}$$

$$\frac{\partial g_{\mu\nu}}{\partial x_m}\gamma_{\mu\alpha}\gamma_{\nu\beta}\Theta_{\alpha\beta}$$

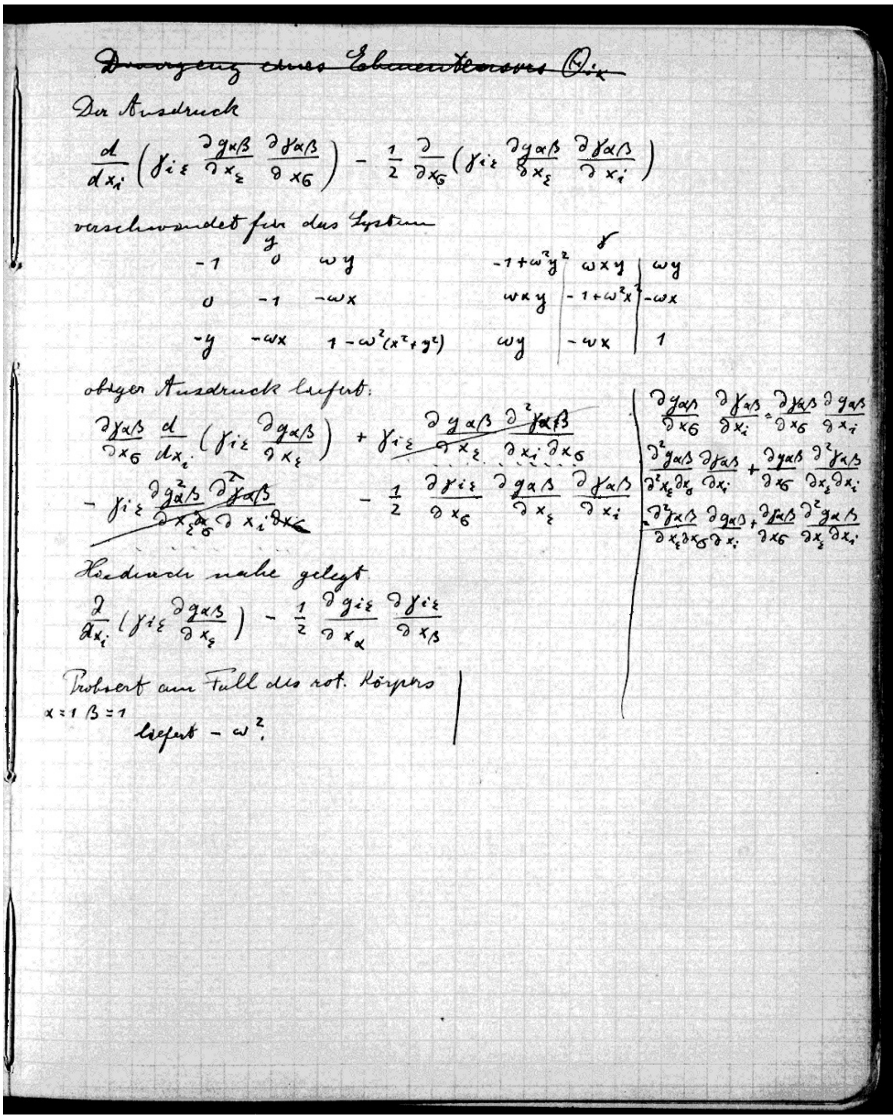
$$-g_{\mu\nu}\frac{\partial \gamma_{\mu\alpha}}{\partial x_m}\gamma_{\nu\beta}\Theta_{\alpha\beta}$$

$$\Theta_{\alpha\beta} = \rho_0 g_{\mu\alpha} g_{\nu\beta}\frac{dx_\mu}{d\tau}\frac{dx_\nu}{d\tau}.$$

Kraft auf ruhenden materiellen Punkt $n = 1$

$\frac{\partial \gamma_{\alpha\beta}}{\partial x_1}$			$\Theta_{\alpha\beta}$	$\frac{dx}{d\tau}$
0	0	Θ		0 0 0 $\frac{1}{\sqrt{g_{44}}}$
0	0	β		
0	β	$2\alpha x$	ρ_0	

[p. 24 R]



~~Drehung eines Elementartensors $Q_{i\kappa}$~~

Der Ausdruck

$$\frac{d}{dx_i}\left(\gamma_{i\varkappa}\frac{\partial g_{\alpha\beta}}{\partial x_\varkappa}\frac{\partial \gamma_{\alpha\beta}}{\partial x_\sigma}\right)-\frac{1}{2}\frac{\partial}{\partial x_\sigma}\left(\gamma_{i\varkappa}\frac{\partial g_{\alpha\beta}}{\partial x_\varkappa}\frac{\partial \gamma_{\alpha\beta}}{\partial x_i}\right)$$

verschwindet für das System

$$\begin{array}{ccc} & g & \\ -1 & 0 & \omega y \\ 0 & -1 & -\omega x \\ -y & -\omega x & 1-\omega^2(x^2+y^2) \end{array} \qquad \begin{array}{c|c|c} & \gamma & \\ -1+\omega^2y^2 & \omega xy & \omega y \\ \omega xy & -1+\omega^2x^2 & -\omega x \\ \omega y & -\omega x & 1 \end{array}$$

obiger Ausdruck liefert:

$$\frac{\partial \gamma_{\alpha\beta}}{\partial x_\sigma}\frac{d}{dx_i}\left(\gamma_{i\varkappa}\frac{\partial g_{\alpha\beta}}{\partial x_\varkappa}\right)+\cancel{\gamma_{i\varkappa}\frac{\partial g_{\alpha\beta}}{\partial x_\varkappa}\frac{\partial^2 \gamma_{\alpha\beta}}{\partial x_i\partial x_\sigma}}$$

$$-\cancel{\gamma_{i\varkappa}\frac{\partial g^2_{\alpha\beta}}{\partial x_\varkappa\partial x_\sigma}\frac{\partial^2 \gamma_{\alpha\beta}}{\partial x_i\partial x_\sigma}}\qquad-\frac{1}{2}\frac{\partial \gamma_{i\varkappa}}{\partial x_\sigma}\frac{\partial g_{\alpha\beta}}{\partial x_\varkappa}\frac{\partial \gamma_{\alpha\beta}}{\partial x_i}$$

$$\frac{\partial g_{\alpha\beta}}{\partial x_\sigma}\frac{\partial \gamma_{\alpha\beta}}{\partial x_i}=\frac{\partial \gamma_{\alpha\beta}}{\partial x_\sigma}\frac{\partial g_{\alpha\beta}}{\partial x_i}$$

$$\frac{\partial^2 g_{\alpha\beta}}{\partial^2 x_\varkappa\partial x_\sigma}\frac{\partial \gamma_{\alpha\beta}}{\partial x_i}+\frac{\partial g_{\alpha\beta}}{\partial x_\sigma}\frac{\partial^2 \gamma_{\alpha\beta}}{\partial x_\varkappa\partial x_i}$$

$$\frac{\partial^2 \gamma_{\alpha\beta}}{\partial x_\varkappa\partial x_\sigma}\frac{\partial g_{\alpha\beta}}{\partial x_i}+\frac{\partial \gamma_{\alpha\beta}}{\partial x_\sigma}\frac{\partial^2 g_{\alpha\beta}}{\partial x_\varkappa\partial x_i}$$

Hierdurch nahe gelegt

$$\frac{\partial}{\partial x_i}\left(\gamma_{i\varkappa}\frac{\partial g_{\alpha\beta}}{\partial x_\varkappa}\right)-\frac{1}{2}\frac{\partial g_{i\varkappa}}{\partial x_\alpha}\frac{\partial \gamma_{i\varkappa}}{\partial x_\beta}$$

Probiert am Fall des rot. Körpers

$\alpha=1\ \beta=1$

liefert $-\omega^2$.

[p. 24 L]

$$\sum \frac{\partial}{\partial x_n}(g_{m\nu} T_{\nu n}) - \frac{1}{2}\frac{\partial g_{\mu\nu}}{\partial x_m} T_{\mu\nu} = 0$$

$$T_{\underset{n}{\mu}\nu} = \gamma_{\underset{n}{\mu}\alpha}\gamma_{\nu\beta}\Theta_{\alpha\beta}$$

$$\sum \frac{\partial}{\partial x_n}(\gamma_{n\alpha}\Theta_{\alpha m}) + \frac{1}{2}\frac{\partial \gamma_{\alpha\beta}}{\partial x_m}\Theta_{\alpha\beta}$$

$$\Theta_{\alpha\beta} = \rho_0 g_{\mu\alpha} g_{\nu\beta} \frac{dx_\mu}{\underset{\sqrt{g_{11}\frac{dx^2}{d\tau}\cdots}}{d\tau}} \frac{dx_\nu}{\underset{\sqrt{}}{d\tau}} .$$

$$T_{\mu\nu} = \rho_0 \frac{dx_\mu}{d\tau}\frac{dx_\nu}{d\tau}$$

$$\frac{\partial g_{\mu\nu}}{\partial x_m}\gamma_{\mu\alpha}\gamma_{\nu\beta}\Theta_{\alpha\beta}$$

$$-\langle\partial\rangle g_{\mu\nu}\frac{\partial \gamma_{\mu\alpha}}{\partial x_m}\gamma_{\nu\beta}\Theta_{\alpha\beta}$$

Kraft auf ruhenden materiellen Punkt $n = 1$

$$\frac{\partial \gamma_{\alpha\beta}}{\partial x_1} \qquad \Theta_{\alpha\beta} \qquad \frac{dx}{d\tau}$$

$$\begin{matrix} 0 & 0 & 0 \\ 0 & 0 & \beta \\ 0 & \beta & 2\alpha x \end{matrix} \qquad \begin{matrix} 0 & 0 & 0 & 0 \\ & & & \\ & & & \rho_0 \end{matrix} \qquad \begin{matrix} 0 & 0 & 0 & \frac{1}{\sqrt{g_{44}}} \end{matrix}$$

[p. 24 R]

⟨Divergenz eines Ebenentensors $\Theta_{i\kappa}$⟩

Der Ausdruck

$$\frac{d}{dx_i}\left(\gamma_{i\varepsilon}\frac{\partial g_{\alpha\beta}}{\partial x_\varepsilon}\frac{\partial \gamma_{\alpha\beta}}{\partial x_\sigma}\right)-\frac{1}{2}\frac{\partial}{\partial x_\sigma}\left(\gamma_{i\varepsilon}\frac{\partial g_{\alpha\beta}}{\partial x_\varepsilon}\frac{\partial \gamma_{\alpha\beta}}{\partial x_i}\right)$$

verschwindet für das System

$$\begin{array}{ccc} & g & \\ -1 & 0 & \omega y \\ 0 & -1 & -\omega x \\ \cancel{y} & -\omega x & 1-\omega^2(x^2+y^2) \end{array} \qquad \begin{array}{c|c|c} & \gamma & \\ -1+\omega^2 y^2 & \omega xy & \omega y \\ \omega xy & -1+\omega^2 x^2 & -\omega x \\ \omega y & -\omega x & 1 \end{array}$$

obiger Ausdruck liefert:

$$\frac{\partial \gamma_{\alpha\beta}}{\partial x_\sigma}\frac{d}{dx_i}\left(\gamma_{i\varepsilon}\frac{\partial g_{\alpha\beta}}{\partial x_\varepsilon}\right)+\cancel{\gamma_{i\varepsilon}\frac{\partial g_{\alpha\beta}}{\partial x_\varepsilon}\frac{\partial^2 \gamma_{\alpha\beta}}{\partial x_i \partial x_\sigma}}$$

$$-\cancel{\gamma_{i\varepsilon}\frac{\partial g^2_{\alpha\beta}}{\partial x_\varepsilon \partial x_\sigma}\frac{\partial^2 \gamma_{\alpha\beta}}{\partial x_i \partial x_\sigma}}-\frac{1}{2}\frac{\partial \gamma_{i\varepsilon}}{\partial x_\sigma}\frac{\partial g_{\alpha\beta}}{\partial x_\varepsilon}\frac{\partial \gamma_{\alpha\beta}}{\partial x_i}$$

Hiedurch nahe gelegt

$$\frac{\partial}{\partial x_i}\left(\gamma_{i\varepsilon}\frac{\partial g_{\alpha\beta}}{\partial x_\varepsilon}\right)-\frac{1}{2}\frac{\partial g_{i\varepsilon}}{\partial x_\alpha}\frac{\partial \gamma_{i\varepsilon}}{\partial x_\beta}$$

Probiert am Fall des rot. Körpers

$\alpha = 1 \quad \beta = 1$

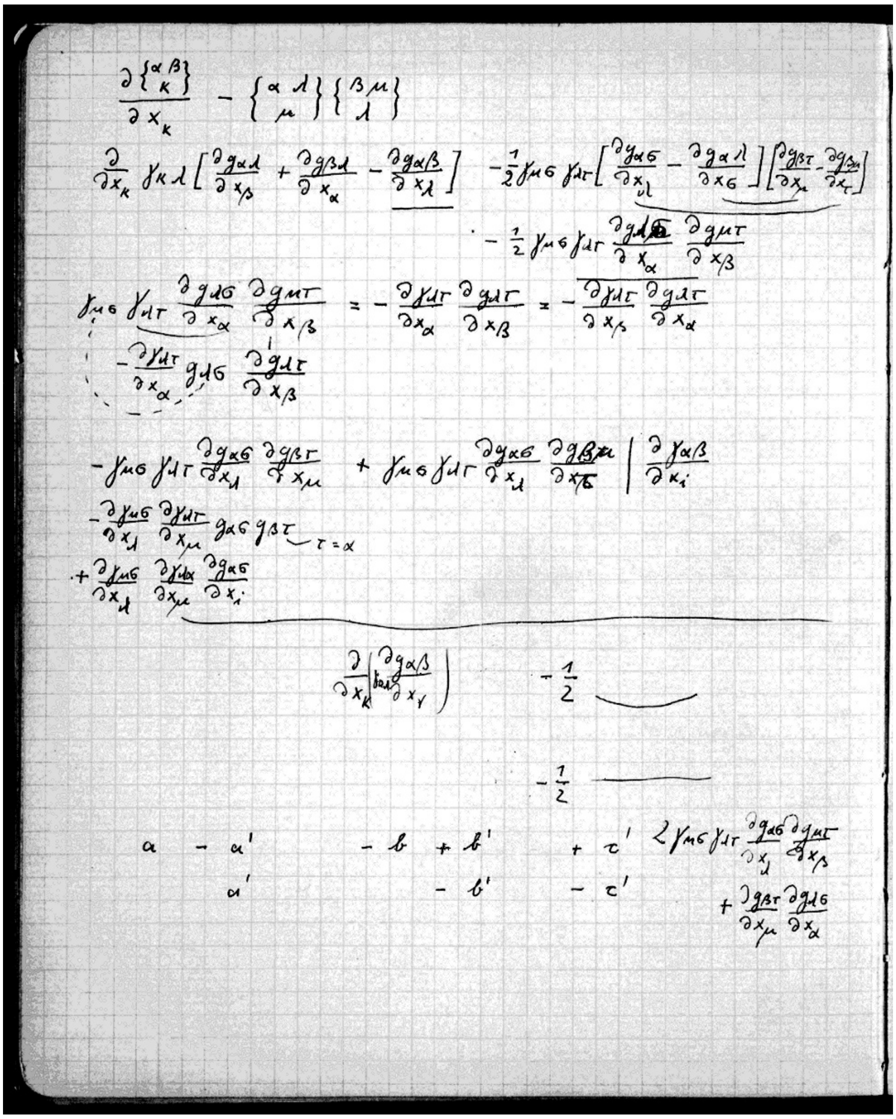
liefert $-\omega^2$.

$$\frac{\partial g_{\alpha\beta}}{\partial x_\sigma}\frac{\partial \gamma_{\alpha\beta}}{\partial x_i} = \frac{\partial \gamma_{\alpha\beta}}{\partial x_\sigma}\frac{\partial g_{\alpha\beta}}{\partial x_i}$$

$$\frac{\partial^2 g_{\alpha\beta}}{\partial x_\varepsilon \partial x_\sigma}\frac{\partial \gamma_{\alpha\beta}}{\partial x_i}+\frac{\partial g_{\alpha\beta}}{\partial x_\sigma}\frac{\partial^2 \gamma_{\alpha\beta}}{\partial x_\varepsilon \partial x_i}$$

$$=\frac{\partial^2 \gamma_{\alpha\beta}}{\partial x_\varepsilon \partial x_\sigma}\frac{\partial g_{\alpha\beta}}{\partial x_i}+\frac{\partial \gamma_{\alpha\beta}}{\partial x_\sigma}\frac{\partial^2 g_{\alpha\beta}}{\partial x_\varepsilon \partial x_i}$$

[p. 25 L]



$$\frac{\partial \left\{ {\alpha\beta \atop \kappa} \right\}}{\partial x_\kappa} - \left\{ {\alpha\lambda \atop \mu} \right\} \left\{ {\beta\mu \atop \lambda} \right\}$$

$$\frac{\partial}{\partial x_\kappa} \gamma_{\kappa\lambda} \left[\frac{\partial g_{\alpha\lambda}}{\partial x_\beta} + \frac{\partial g_{\beta\lambda}}{\partial x_\alpha} - \frac{\partial g_{\alpha\beta}}{\partial x_\lambda} \right] - \frac{1}{2} \gamma_{\mu\sigma} \gamma_{\lambda\tau} \left[\frac{\partial g_{\alpha\sigma}}{\partial x_\lambda} - \frac{\partial g_{\alpha\lambda}}{\partial x_\sigma} \right] \left[\frac{\partial g_{\beta\tau}}{\partial x_\mu} - \frac{\partial g_{\beta\mu}}{\partial x_\tau} \right]$$

$$- \frac{1}{2} \gamma_{\mu\sigma} \gamma_{\lambda\tau} \frac{\partial g_{\lambda\sigma}}{\partial x_\alpha} \frac{\partial g_{\mu\tau}}{\partial x_\beta}$$

$$\gamma_{\mu\sigma} \gamma_{\lambda\tau} \frac{\partial g_{\lambda\sigma}}{\partial x_\alpha} \frac{\partial g_{\mu\tau}}{\partial x_\beta} = - \frac{\partial \gamma_{\lambda\tau}}{\partial x_\alpha} \frac{\partial g_{\lambda\tau}}{\partial x_\beta} = - \frac{\partial \gamma_{\lambda\tau}}{\partial x_\beta} \frac{\partial g_{\lambda\tau}}{\partial x_\alpha}$$

$$- \frac{\partial \gamma_{\mu\tau}}{\partial x_\alpha} g_{\lambda\sigma} \frac{\partial g_{\lambda\tau}}{\partial x_\beta}$$

$$- \gamma_{\mu\sigma} \gamma_{\lambda\tau} \frac{\partial g_{\alpha\sigma}}{\partial x_\lambda} \frac{\partial g_{\beta\tau}}{\partial x_\mu} + \gamma_{\mu\sigma} \gamma_{\lambda\tau} \frac{\partial g_{\alpha\sigma}}{\partial x_\lambda} \frac{\partial g_{\beta\mu}}{\partial x_\tau} \quad \Big| \quad \frac{\partial \gamma_{\alpha\beta}}{\partial x_i}$$

$$- \frac{\partial \gamma_{\mu\sigma}}{\partial x_\lambda} \frac{\partial \gamma_{\lambda\tau}}{\partial x_\mu} g_{\alpha\sigma} g_{\beta\tau} \quad \tau = \alpha$$

$$+ \frac{\partial \gamma_{\mu\sigma}}{\partial x_\lambda} \frac{\partial \gamma_{\lambda\alpha}}{\partial x_\mu} \frac{\partial g_{\alpha\sigma}}{\partial x_i}$$

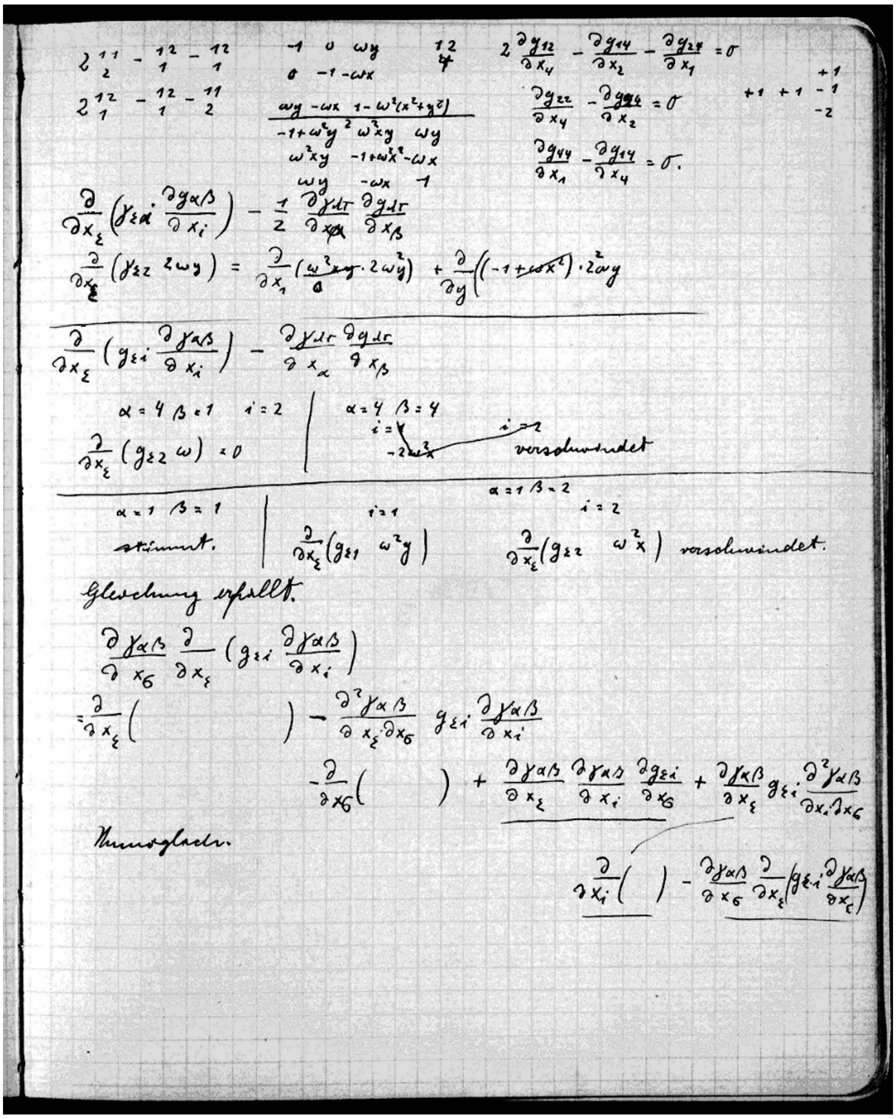
$$\frac{\partial}{\partial x_\kappa} \left(\gamma_{\kappa\alpha} \frac{\partial g_{\alpha\beta}}{\partial x_\gamma} \right) \qquad - \frac{1}{2}$$

$$- \frac{1}{2}$$

$$a \quad - a' \qquad - b \quad + b' \qquad + c' \quad 2 \gamma_{\mu\sigma} \gamma_{\lambda\tau} \frac{\partial g_{\alpha\sigma}}{\partial x_\lambda} \frac{\partial g_{\mu\tau}}{\partial x_\beta}$$

$$a' \qquad - b' \qquad - c' \qquad + \frac{\partial g_{\beta\tau}}{\partial x_\mu} \frac{\partial g_{\lambda\sigma}}{\partial x_\alpha}$$

[p. 25 R]



$2\begin{bmatrix}11\\2\end{bmatrix} - \begin{bmatrix}12\\1\end{bmatrix} - \begin{bmatrix}12\\1\end{bmatrix}$ $\quad$ $\begin{matrix}-1 & 0 & \omega y\\ 0 & -1 & -\omega x\end{matrix}$ $\quad$ $\begin{matrix}12\\4\end{matrix}$ $\quad$ $2\frac{\partial g_{12}}{\partial x_4} - \frac{\partial g_{14}}{\partial x_2} - \frac{\partial g_{24}}{\partial x_1} = 0$

$2\begin{bmatrix}12\\1\end{bmatrix} - \begin{bmatrix}12\\1\end{bmatrix} - \begin{bmatrix}11\\2\end{bmatrix}$ $\quad$ $\begin{matrix}\omega y & -\omega x & 1-\omega^2(x^2+y^2)\end{matrix}$ $\quad$ $\frac{\partial g_{22}}{\partial x_4} - \frac{\partial g_{24}}{\partial x_2} = 0$ $\quad$ $\begin{matrix} & & +1\\ +1 & +1 & -1\\ & & -2\end{matrix}$

$$\begin{matrix}-1+\omega^2 y^2 & \omega^2 xy & \omega y\\ \omega^2 xy & -1+\omega^2 x^2 & -\omega x\\ \omega y & -\omega x & 1\end{matrix}$$

$\frac{\partial g_{44}}{\partial x_1} - \frac{\partial g_{14}}{\partial x_4} = 0.$

$$\frac{\partial}{\partial x_\kappa}\left(\gamma_{\kappa\alpha}\frac{\partial g_{\alpha\beta}}{\partial x_i}\right) - \frac{1}{2}\frac{\partial \gamma_{\lambda\tau}}{\partial x_\alpha}\frac{\partial g_{\lambda\tau}}{\partial x_\beta}$$

$$\frac{\partial}{\partial x_\kappa}(\gamma_{\kappa 2}\, 2\omega y) = \frac{\partial}{\partial x_1}(\omega^2 \cancel{xy}\cdot 2\omega y) + \frac{\partial}{\partial y}((-1+\cancel{\omega x^2})\cdot 2\omega y$$

$$\frac{\partial}{\partial x_\kappa}\left(g_{\kappa i}\frac{\partial \gamma_{\alpha\beta}}{\partial x_i}\right) - \frac{\partial \gamma_{\lambda\tau}}{\partial x_\alpha}\frac{\partial g_{\lambda\tau}}{\partial x_\beta}$$

$\alpha = 4\ \beta = 1\quad i = 2$ | $\alpha = 4\ \beta = 4$

$\frac{\partial}{\partial x_\kappa}(g_{\kappa 2}\,\omega) = 0$ | $i = 1 \quad -2\omega^2 x \quad i = 2$ verschwindet

$\alpha = 1\ \beta = 2$

$\alpha = 1\ \beta = 1$ stimmt. | $i = 1 \quad \frac{\partial}{\partial x_\kappa}(g_{\kappa 1}\ \omega^2 y)$ $\quad i = 2 \quad \frac{\partial}{\partial x_\kappa}(g_{\kappa 2}\ \omega^2 x)$ verschwindet.

Gleichung erfüllt.

$$\frac{\partial \gamma_{\alpha\beta}}{\partial x_\sigma}\frac{\partial}{\partial x_\kappa}\left(g_{\kappa i}\frac{\partial \gamma_{\alpha\beta}}{\partial x_i}\right)$$

$$= \frac{\partial}{\partial x_\kappa}(\qquad) - \frac{\partial^2 \gamma_{\alpha\beta}}{\partial x_\kappa \partial x_\sigma} g_{\kappa i}\frac{\partial \gamma_{\alpha\beta}}{\partial x_i}$$

$$-\frac{\partial}{\partial x_\sigma}(\qquad) + \underline{\frac{\partial \gamma_{\alpha\beta}}{\partial x_\kappa}\frac{\partial \gamma_{\alpha\beta}}{\partial x_i}\frac{\partial g_{\kappa i}}{\partial x_\sigma}} + \frac{\partial \gamma_{\alpha\beta}}{\partial x_\kappa} g_{\kappa i}\frac{\partial^2 \gamma_{\alpha\beta}}{\partial x_i \partial x_\sigma}$$

Unmöglich.

$$\underline{\frac{\partial}{\partial x_i}(\quad)} - \underline{\frac{\partial \gamma_{\alpha\beta}}{\partial x_\sigma}\frac{\partial}{\partial x_\kappa}\left(g_{\kappa i}\frac{\partial \gamma_{\alpha\beta}}{\partial x_i}\right)}$$

[p. 25 L]

$$\frac{\partial\left\{\begin{matrix}\alpha\ \beta\\ \kappa\end{matrix}\right\}}{\partial x_\kappa} - \left\{\begin{matrix}\alpha\ \lambda\\ \mu\end{matrix}\right\}\left\{\begin{matrix}\beta\ \mu\\ \lambda\end{matrix}\right\}$$

$$\frac{\partial}{\partial x_\kappa}\gamma_{\kappa\lambda}\left[\frac{\partial g_{\alpha\lambda}}{\partial x_\beta} + \frac{\partial g_{\beta\lambda}}{\partial x_\alpha} - \frac{\partial g_{\alpha\beta}}{\underline{\partial x_\lambda}}\right] - \frac{1}{2}\gamma_{\mu\sigma}\gamma_{\lambda\tau}\left[\frac{\partial g_{\alpha\sigma}}{\partial x_\lambda} - \frac{\partial g_{\alpha\lambda}}{\partial x_\sigma}\right]\left[\frac{\partial g_{\beta\tau}}{\partial x_\mu} - \frac{\partial g_{\beta\mu}}{\partial x_\tau}\right]$$

$$-\frac{1}{2}\gamma_{\mu\sigma}\gamma_{\lambda\tau}\frac{\partial g_{\lambda\sigma}}{\partial x_\alpha}\frac{\partial g_{\mu\tau}}{\partial x_\beta}$$

$$\gamma_{\mu\sigma}\gamma_{\lambda\tau}\frac{\partial g_{\lambda\sigma}}{\partial x_\alpha}\frac{\partial g_{\mu\tau}}{\partial x_\beta} = -\frac{\partial \gamma_{\lambda\tau}}{\partial x_\alpha}\frac{\partial g_{\lambda\tau}}{\partial x_\beta} = -\overline{\frac{\partial \gamma_{\lambda\tau}}{\partial x_\beta}\frac{\partial g_{\lambda\tau}}{\partial x_\alpha}}$$

$$-\frac{\partial \gamma_{\lambda\tau}}{\partial x_\alpha}g_{\lambda\sigma}\frac{\partial g_{\lambda\tau}}{\partial x_\beta}$$

$$-\gamma_{\mu\sigma}\gamma_{\lambda\tau}\frac{\partial g_{\alpha\sigma}}{\partial x_\lambda}\frac{\partial g_{\beta\tau}}{\partial x_\mu} + \gamma_{\mu\sigma}\gamma_{\lambda\tau}\frac{\partial g_{\alpha\sigma}}{\partial x_\lambda}\frac{\partial g_{\beta\mu}}{\partial x_\tau} \quad \Bigg| \quad \frac{\partial y_{\alpha\beta}}{\partial x_i}$$

$$-\frac{\partial \gamma_{\mu\sigma}}{\partial x_\lambda}\frac{\partial \gamma_{\lambda\tau}}{\partial x_\mu}g_{\alpha\sigma}g_{\beta\tau} \quad \tau = \alpha$$

$$+\frac{\partial \gamma_{\mu\sigma}}{\partial x_\lambda}\frac{\partial \gamma_{\lambda\alpha}}{\partial x_\mu}\frac{\partial g_{\alpha\sigma}}{\partial x_i}$$

$$\frac{\partial}{\partial x_\kappa}\left(\gamma_{\kappa\lambda}\frac{\partial g_{\alpha\beta}}{\partial x_\nu}\right) \qquad -\frac{1}{2}$$

$$-\frac{1}{2} \quad \underline{\qquad\qquad}$$

$$\begin{array}{llllll} a & -\ a' & -b & +\ b' & +\ c' & 2\gamma_{\mu\sigma}\gamma_{\lambda\tau}\frac{\partial g_{\alpha\sigma}}{\partial x_\lambda}\frac{\partial g_{\mu\tau}}{\partial x_\beta} \\ & a' & & -\ b' & -\ c' & \\ & & & & & +\frac{\partial g_{\beta\tau}}{\partial x_\mu}\frac{\partial g_{\lambda\sigma}}{\partial x_\alpha} \end{array}$$

[p. 25 R]

$$2\,{}^{11}_{2} - {}^{12}_{1} - {}^{12}_{1} \qquad 2\,{}^{12}_{1} - {}^{12}_{1} - {}^{11}_{2}$$

$$\begin{array}{ccc|c} -1 & 0 & \omega y & 12 \\ 0 & -1 & -\omega x & 4 \\ \omega y & -\omega x & 1-\omega^2(x^2+y^2) & \\ \hline -1+\omega^2 y^2 & \omega^2 xy & \omega y & \\ \omega^2 xy & -1+\omega^2 x^2 & -\omega x & \\ \omega y & -\omega x & -1 & \end{array}$$

$$2\frac{\partial g_{12}}{\partial x_4} - \frac{\partial g_{14}}{\partial x_2} - \frac{\partial g_{24}}{\partial x_1} = 0 \qquad \begin{array}{c} +1 \\ +1+1\ \ -1 \\ -2 \end{array}$$

$$\frac{\partial g_{22}}{\partial x_4} - \frac{\partial g_{24}}{\partial x_2} = 0$$

$$\frac{\partial g_{44}}{\partial x_1} - \frac{\partial g_{14}}{\partial x_4} = 0.$$

$$\frac{\partial}{\partial x_\varepsilon}\left(\gamma_{\varepsilon i}\frac{\partial g_{\alpha\beta}}{\partial x_i}\right) - \frac{1}{2}\frac{\partial \gamma_{\lambda\tau}}{\partial x_\alpha}\frac{\partial g_{\lambda\tau}}{\partial x_\beta}$$

$$\frac{\partial}{\partial x_\varepsilon}(\gamma_{\varepsilon 2}2\omega y) = \underset{0[?]}{\frac{\partial}{\partial x_1}(<\omega^2 xy>\cdot\ 2\,\omega^2 y)} + \frac{\partial}{\partial y}\left((-1<+\omega x^2>)\cdot 2\omega^2 y\right.$$

$$\frac{\partial}{\partial x_\varepsilon}\left(g_{\varepsilon i}\frac{\partial \gamma_{\alpha\beta}}{\partial x_i}\right) - \frac{\partial \gamma_{\lambda\tau}}{\partial x_\alpha}\frac{\partial g_{\lambda\tau}}{\partial x_\beta}$$

$\alpha = 4 \quad \beta = 1 \quad i = 2$

$$\frac{\partial}{\partial x_\varepsilon}(g_{\varepsilon 2}\omega) = 0$$

$\alpha = 4 \quad \beta = 4$

$i = 1 \qquad\qquad i = 2$

$-2\omega^2 x$ verschwindet

$\alpha = 1 \quad \beta = 1$

stimmt.

$\alpha = 1 \quad \beta = 2$

$i = 1 \qquad\qquad i = 2$

$$\frac{\partial}{\partial x_\varepsilon}(g_{\varepsilon 1}\omega^2 y) \qquad \frac{\partial}{\partial x_\varepsilon}(g_{\varepsilon 2}\omega^2 x)$$ verschwindet.

Gleichung erfüllt.

$$\frac{\partial \gamma_{\alpha\beta}}{\partial x_\sigma}\frac{\partial}{\partial x_\varepsilon}\left(g_{\varepsilon i}\frac{\partial \gamma_{\alpha\beta}}{\partial x_i}\right)$$

$$= \frac{\partial}{\partial x_\varepsilon}\left(\qquad\right) - \frac{\partial^2 \gamma_{\alpha\beta}}{\partial x_\varepsilon \cdot \partial x_\sigma} g_{\varepsilon i}\frac{\partial \gamma_{\alpha\beta}}{\partial x_i}$$

$$- \frac{\partial}{\partial x_\sigma}\left(\qquad\right) + \underline{\frac{\partial \gamma_{\alpha\beta}}{\partial x_\varepsilon}\frac{\partial \gamma_{\alpha\beta}}{\partial x_i}\frac{\partial g_{\varepsilon i}}{\partial x_\sigma}} + \frac{\partial \gamma_{\alpha\beta}}{\partial x_\varepsilon} g_{\varepsilon i}\frac{\partial^2 \gamma_{\alpha\beta}}{\partial x_i \partial x_\sigma}$$

Unmöglich.

$$\underline{\frac{\partial}{\partial x_i}\left(\qquad\right)} - \underline{\frac{\partial \gamma_{\alpha\beta}}{\partial x_\sigma}\frac{\partial}{\partial x_\varepsilon}\left(g_{\varepsilon i}\frac{\partial \gamma_{\alpha\beta}}{\partial x_\varepsilon}\right)}$$

[p. 26 L]

System der Gleichungen für Materie

$$\frac{\partial}{\partial x_n}\left(\sqrt{G}\, g_{\mu\nu} T_{\nu n}\right) - \frac{1}{2}\sqrt{G}\,\frac{\partial \gamma_{\mu\nu}}{\partial x_m} T_{\mu\nu} = 0$$

$$T_{\mu\nu} = \varrho \frac{dx_\mu}{d\tau}\frac{dx_\nu}{d\tau}.$$

Ableitung der Gravitationsgleichungen

$$\underbrace{\frac{\partial g_{\mu\nu}}{\partial x_m}\frac{\partial}{\partial x_\alpha}\left(\gamma_{\alpha\beta}\sqrt{G}\frac{\partial \gamma_{\mu\nu}}{\partial x_\beta}\right)}_{0} = \underbrace{\frac{\partial}{\partial x_\alpha}\left(\gamma_{\alpha\beta}\sqrt{G}\frac{\partial \gamma_{\mu\nu}}{\partial x_\beta}\frac{\partial g_{\mu\nu}}{\partial x_m}\right)}_{+} - \sqrt{G}\,\gamma_{\alpha\beta}\frac{\partial \gamma_{\mu\nu}}{\partial x_\beta}\frac{\partial^2 g_{\mu\nu}}{\partial x_m \partial x_\alpha}$$

$$\underbrace{-\frac{\partial}{\partial x_m}\left(\gamma_{\alpha\beta}\sqrt{G}\frac{\partial \gamma_{\mu\nu}}{\partial x_\beta}\frac{\partial g_{\mu\nu}}{\partial x_\alpha}\right)}_{\times} + \frac{\partial g_{\mu\nu}}{\partial x_\alpha}\frac{\partial}{\partial x_m}\left(\sqrt{G}\,\gamma_{\alpha\beta}\frac{\partial \gamma_{\mu\nu}}{\partial x_\beta}\right)$$

$$\frac{1}{2}\frac{\partial g_{\sigma\tau}}{\partial x_m}\gamma_{\sigma\tau}\sqrt{G}\frac{\partial g_{\mu\nu}}{\partial x_\alpha}\frac{\partial \gamma_{\mu\nu}}{\partial x_\beta}\gamma_{\alpha\beta} + \frac{\partial \gamma_{\alpha\beta}}{\partial x_m}\sqrt{G}\frac{\partial g_{\mu\nu}}{\partial x_\alpha}\frac{\partial \gamma_{\mu\nu}}{\partial x_\beta} + \frac{\partial g_{\mu\nu}}{\partial x_\alpha}$$

$$\underline{\frac{1}{2}\frac{\partial g_{\mu\nu}}{\partial x_m}\gamma_{\mu\nu}\sqrt{G}\frac{\partial g_{\sigma\tau}}{\partial x_\alpha}\frac{\partial \gamma_{\sigma\tau}}{\partial x_\beta}\gamma_{\alpha\beta}} \qquad \frac{\partial \gamma_{\alpha\beta}}{\partial x_m}\underbrace{g_{\alpha\sigma}\, g_{\beta\tau}\,\Big|\,\gamma_{\alpha'\sigma}\,\gamma_{\beta'\tau}}_{1}\sqrt{G}\frac{\partial g_{\mu\nu}}{\partial x_{\alpha'}}\frac{\partial \gamma_{\mu\nu}}{\partial x_{\beta'}}$$

$$\underbrace{-\frac{\partial g_{\sigma\tau}}{\partial x_m}\sqrt{G}\,\gamma_{\alpha\sigma}\,\gamma_{\beta\tau}\frac{\partial g_{\mu\nu}}{\partial x_\alpha}\frac{\partial \gamma_{\mu\nu}}{\partial x_\beta}}_{0}$$

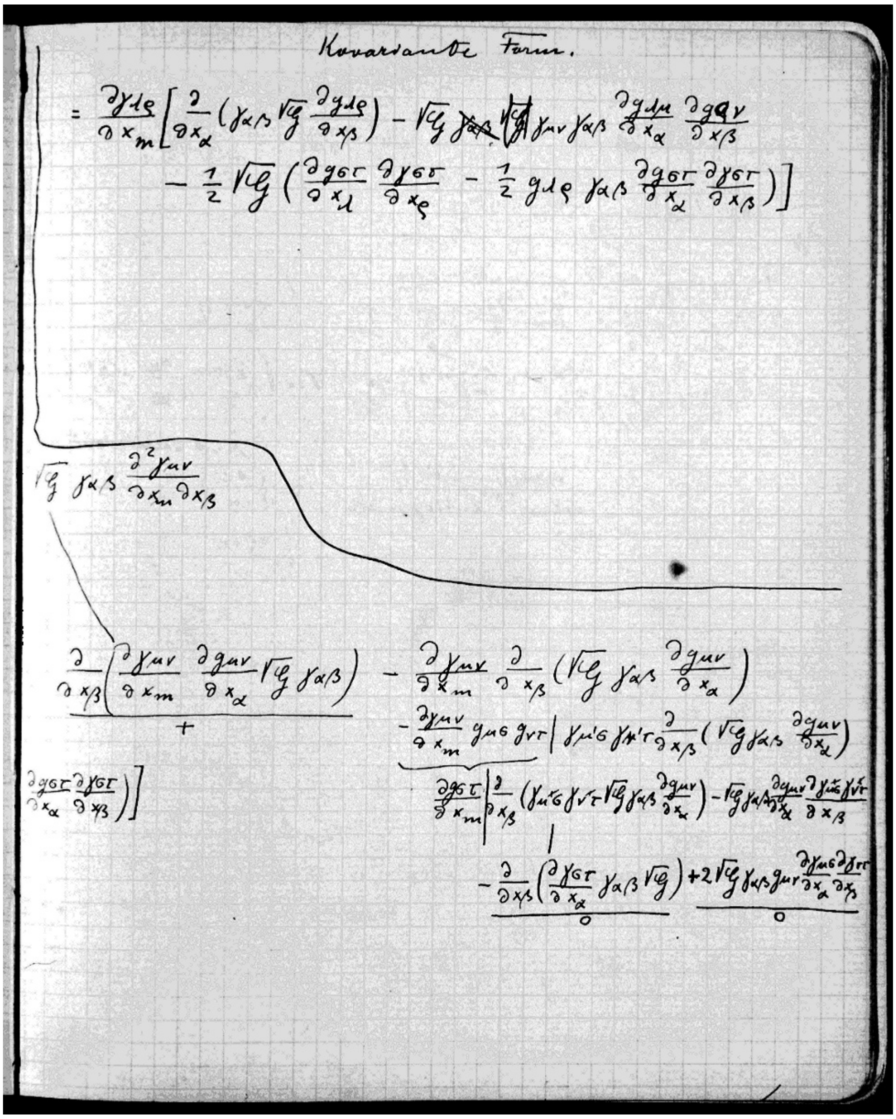
Zusammenfassung

$$\frac{\partial g_{\mu\nu}}{\partial x_m}\left[\frac{\partial}{\partial x_\alpha}\left(\gamma_{\alpha\beta}\sqrt{G}\frac{\partial \gamma_{\mu\nu}}{\partial x_\beta}\right) - \sqrt{G}\,\gamma_{\alpha\beta}\, g_{\sigma\tau}\frac{\partial \gamma_{\mu\sigma}}{\partial x_\alpha}\frac{\partial \gamma_{\nu\tau}}{\partial x_\beta}\right) + \frac{1}{2}\sqrt{G}\left(\gamma_{\alpha\mu}\gamma_{\beta\nu}\frac{\partial g_{\sigma\tau}}{\partial x_\alpha}\frac{\partial \gamma_{\sigma\tau}}{\partial x_\beta} - \frac{1}{2}\gamma_{\mu\nu}\gamma_{\alpha\beta}\right.$$

$$= \frac{\partial}{\partial x_\alpha}\left(\sqrt{G}\,\gamma_{\alpha\beta}\frac{\partial \gamma_{\sigma\tau}}{\partial x_\beta}\frac{\partial g_{\sigma\tau}}{\partial x_m}\right) - \frac{1}{2}\frac{\partial}{\partial x_m}\left(\sqrt{G}\,\gamma_{\alpha\beta}\frac{\partial g_{\sigma\tau}}{\partial x_\alpha}\frac{\partial \gamma_{\sigma\tau}}{\partial x_\beta}\right)$$

Dies ist die Kontra-Form.

[p. 26 R]



Kovariante Form.

$$= \frac{\partial \gamma_{\lambda\varrho}}{\partial x_m}\Big[\frac{\partial}{\partial x_\alpha}\Big(\gamma_{\alpha\beta}\sqrt{G}\frac{\partial g_{\lambda\varrho}}{\partial x_\beta}\Big) - \sqrt{G}\,\cancel{\gamma_{\alpha\beta}}\,\gamma_{\mu\nu}\gamma_{\alpha\beta}\frac{\partial g_{\lambda\mu}}{\partial x_\alpha}\frac{\partial g_{\varrho\nu}}{\partial x_\beta}$$

$$- \frac{1}{2}\sqrt{G}\Big(\frac{\partial g_{\sigma\tau}}{\partial x_\lambda}\frac{\partial \gamma_{\sigma\tau}}{\partial x_\varrho} - \frac{1}{2}g_{\lambda\varrho}\gamma_{\alpha\beta}\frac{\partial g_{\sigma\tau}}{\partial x_\alpha}\frac{\partial \gamma_{\sigma\tau}}{\partial x_\beta}\Big)\Big]$$

$$\sqrt{G}\,\gamma_{\alpha\beta}\frac{\partial^2 \gamma_{\mu\nu}}{\partial x_m \partial x_\beta}$$

$$\frac{\partial}{\partial x_\beta}\Big(\frac{\partial \gamma_{\mu\nu}}{\partial x_m}\frac{\partial g_{\mu\nu}}{\partial x_\alpha}\sqrt{G}\,\gamma_{\alpha\beta}\Big) - \frac{\partial \gamma_{\mu\nu}}{\partial x_m}\frac{\partial}{\partial x_\beta}\Big(\sqrt{G}\,\gamma_{\alpha\beta}\frac{\partial g_{\mu\nu}}{\partial x_\alpha}\Big)$$

$$+$$

$$\frac{\partial g_{\sigma\tau}}{\partial x_\alpha}\frac{\partial \gamma_{\sigma\tau}}{\partial x_\beta}\Big)\Big]$$

$$\underbrace{-\frac{\partial \gamma_{\mu\nu}}{\partial x_m}g_{\mu\sigma}g_{\nu\tau}}\Big|\,\gamma_{\mu'\sigma}\gamma_{\nu'\tau}\frac{\partial}{\partial x_\beta}\Big(\sqrt{G}\,\gamma_{\alpha\beta}\frac{\partial g_{\mu\nu}}{\partial x_\alpha}\Big)$$

$$\frac{\partial g_{\sigma\tau}}{\partial x_m}\Big|\frac{\partial}{\partial x_\beta}\Big(\gamma_{\mu'\sigma}\gamma_{\nu'\tau}\sqrt{G}\,\gamma_{\alpha\beta}\frac{\partial g_{\mu\nu}}{\partial x_\alpha}\Big) - \sqrt{G}\,\gamma_{\alpha\beta}\frac{\partial g_{\mu\nu}}{\partial x_\alpha}\frac{\partial \gamma_{\mu\sigma}\gamma_{\nu\tau}}{\partial x_\beta}$$

$$\underbrace{-\frac{\partial}{\partial x_\beta}\Big(\frac{\partial \gamma_{\sigma\tau}}{\partial x_\alpha}\gamma_{\alpha\beta}\sqrt{G}\Big)}_{0} \underbrace{+2\sqrt{G}\,\gamma_{\alpha\beta}g_{\mu\tau}\frac{\partial \gamma_{\mu\sigma}}{\partial x_\alpha}\frac{\partial \gamma_{\tau\tau}}{\partial x_\beta}}_{0}$$

[p. 26 L]

System der Gleichungen für Materie

$$\frac{\partial}{\partial x_n}(\sqrt{G} g_{\mu\nu} T_{\nu n}) - \frac{1}{2}\sqrt{G}\frac{\partial g_{\mu\nu}}{\partial x_m} T_{\mu\nu} = 0$$

$$T_{\mu\nu} = \rho \frac{dx_\mu}{d\tau}\frac{dx_\nu}{d\tau}.$$

Ableitung der Gravitationsgleichungen

$$\underbrace{\frac{\partial g_{\mu\nu}}{\partial x_m} \,\Big|\, \frac{\partial}{\partial x_\alpha}\left(\gamma_{\alpha\beta}\sqrt{G}\frac{\partial \gamma_{\mu\nu}}{\partial x_\beta}\right)}_{\mathrm{o}} = \underbrace{\frac{\partial}{\partial x_\alpha}\left(\gamma_{\alpha\beta}\sqrt{G}\frac{\partial \gamma_{\mu\nu}}{\partial x_\beta}\frac{\partial g_{\mu\nu}}{\partial x_m}\right)}_{+} - \sqrt{G}\gamma_{\alpha\beta}\frac{\partial \gamma_{\mu\nu}}{\partial x_\beta}\frac{\partial g^2_{\mu\nu}}{\partial x_m \partial x_\alpha}$$

$$\underbrace{-\frac{\partial}{\partial x_m}\left(\gamma_{\alpha\beta}\sqrt{G}\frac{\partial \gamma_{\mu\nu}}{\partial x_\beta}\frac{\partial g_{\mu\nu}}{\partial x_\alpha}\right)}_{\times} + \frac{\partial g_{\mu\nu}}{\partial x_\alpha}\frac{\partial}{\partial x_m}\left(\sqrt{G}\gamma_{\alpha\beta}\frac{\partial \gamma_{\mu\nu}}{\partial x_\beta}\right)$$

$$\frac{1}{2}\frac{\partial g_{\sigma\tau}}{\partial x_m}\gamma_{\sigma\tau}\sqrt{G}\frac{\partial g_{\mu\nu}}{\partial x_\alpha}\frac{\partial \gamma_{\mu\nu}}{\partial x_\beta}\gamma_{\alpha\beta} + \frac{\partial \gamma_{\alpha\beta}}{\partial x_m}\sqrt{G}\frac{\partial g_{\mu\nu}}{\partial x_\alpha}\frac{\partial \gamma_{\mu\nu}}{\partial x_\beta} + \frac{\partial g_{\mu\nu}}{\partial x_\alpha}$$

$$\underline{\frac{1}{2}\frac{\partial g_{\mu\nu}}{\partial x_m}\gamma_{\mu\nu}\sqrt{G}\frac{\partial g_{\sigma\tau}}{\partial x_\alpha}\frac{\partial \gamma_{\sigma\tau}}{\partial x_\beta}\gamma_{\alpha\beta}} \qquad \frac{\partial \gamma_{\alpha\beta}}{\partial x_m}\underbrace{g_{\alpha\sigma}g_{\beta\tau} \,\big|\, \gamma_{\alpha'\sigma}\gamma_{\beta'\tau}} \sqrt{G}\frac{\partial g_{\mu\nu}}{\partial x_{\alpha'}}\frac{\partial \gamma_{\mu\nu}}{\partial x_{\beta'}}$$

$$\underbrace{-\frac{\partial g_{\sigma\tau}}{\partial x_m}\sqrt{G}\gamma_{\alpha\sigma}\gamma_{\beta\tau}\frac{\partial g_{\mu\nu}}{\partial x_\alpha}\frac{\partial \gamma_{\mu\nu}}{\partial x_\beta}}_{\mathrm{o}}$$

Zusammenfassung

$$\frac{\partial g_{\mu\nu}}{\partial x_m}\left[\left(\frac{\partial}{\partial x_\alpha}\left(\gamma_{\alpha\beta}\sqrt{G}\frac{\partial \gamma_{\mu\nu}}{\partial x_\beta}\right) - \sqrt{G}\gamma_{\alpha\beta}g_{\sigma\tau}\frac{\partial \gamma_{\mu\sigma}}{\partial x_\alpha}\frac{\partial \gamma_{\nu\tau}}{\partial x_\beta}\right)\right.$$
$$\left. + \frac{1}{2}\sqrt{G}\left(\gamma_{\alpha\mu}\gamma_{\beta\nu}\frac{\partial g_{\sigma\tau}}{\partial x_\alpha}\frac{\partial \gamma_{\sigma\tau}}{\partial x_\beta} - \frac{1}{2}\gamma_{\mu\nu}\gamma_{\alpha\beta}\frac{\partial g_{\sigma\tau}}{\partial x_\alpha}\frac{\partial \gamma_{\sigma\tau}}{\partial x_\beta}\right)\right]$$

$$= \frac{\partial}{\partial x_\alpha}\left(\sqrt{G}\gamma_{\alpha\beta}\frac{\partial \gamma_{\sigma\tau}}{\partial x_\beta}\frac{\partial g_{\sigma\tau}}{\partial x_m}\right) - \frac{1}{2}\frac{\partial}{\partial x_m}\left(\sqrt{G}\gamma_{\alpha\beta}\frac{\partial g_{\sigma\tau}}{\partial x_\alpha}\frac{\partial \gamma_{\sigma\tau}}{\partial x_\beta}\right)$$

Dies ist die Kontra-Form.

[p. 26 R]

Kovariante Form.

$$= \frac{\partial\gamma_{\lambda\rho}}{\partial x_m}\left[\frac{\partial}{\partial x_\alpha}\left(\gamma_{\alpha\beta}\sqrt{G}\frac{\partial g_{\lambda\rho}}{\partial x_\beta}\right) - \sqrt{G}\langle\gamma_{\alpha\beta}\sqrt{G}\rangle\gamma_{\mu\nu}\gamma_{\alpha\beta}\frac{\partial g_{\lambda\mu}}{\partial x_\alpha}\frac{\partial g_{\rho\nu}}{\partial x_\beta}\right.$$

$$\left.-\frac{1}{2}\sqrt{G}\left(\frac{\partial g_{\sigma\tau}}{\partial x_\lambda}\frac{\partial\gamma_{\sigma\tau}}{\partial x_\rho}-\frac{1}{2}g_{\lambda\rho}\gamma_{\alpha\beta}\frac{\partial g_{\sigma\tau}}{\partial x_\alpha}\frac{\partial\gamma_{\sigma\tau}}{\partial x_\beta}\right)\right]$$

$$\sqrt{G}\,\gamma_{\alpha\beta}\frac{\partial^2\gamma_{\mu\nu}}{\partial x_m\partial x_\beta}$$

$$\underline{\frac{\partial}{\partial x_\beta}\left(\frac{\partial\gamma_{\mu\nu}}{\partial x_m}\frac{\partial g_{\mu\nu}}{\partial x_\alpha}\sqrt{G}\gamma_{\alpha\beta}\right)}_{+} - \frac{\partial\gamma_{\mu\nu}}{\partial x_m}\frac{\partial}{\partial x_\beta}\left(\sqrt{G}\gamma_{\alpha\beta}\frac{\partial g_{\mu\nu}}{\partial x_\alpha}\right)$$

$$\underbrace{-\frac{\partial\gamma_{\mu\nu}}{\partial x_m}g_{\mu\sigma}g_{\nu\tau}}\Big|\gamma_{\mu'\sigma}\gamma_{\nu'\tau}\frac{\partial}{\partial x_\beta}\left(\sqrt{G}\gamma_{\alpha\beta}\frac{\partial g_{\mu\nu}}{\partial x_\alpha}\right)$$

$$\frac{\partial g_{\sigma\tau}}{\partial x_m}\Bigg|\ \frac{\partial}{\partial x_\beta}\left(\gamma_{\mu\sigma}\gamma_{\nu\tau}\sqrt{G}\gamma_{\alpha\beta}\frac{\partial g_{\mu\nu}}{\partial x_\alpha}\right) - \sqrt{G}\gamma_{\alpha\beta}\frac{\partial g_{\mu\nu}}{\partial x_\alpha}\frac{\partial\gamma_{\mu\sigma}\gamma_{\nu\tau}}{\partial x_\beta}$$

$$\underline{-\frac{\partial}{\partial x_\beta}\left(\frac{\partial\gamma_{\sigma\tau}}{\partial x_\alpha}\gamma_{\alpha\beta}\sqrt{G}\right)}_{\text{o}} + \underline{2\sqrt{G}\gamma_{\alpha\beta}g_{\mu\nu}\frac{\partial\gamma_{\mu\sigma}}{\partial x_\alpha}\frac{\partial\gamma_{\nu\tau}}{\partial x_\beta}}_{\text{o}}$$

[p. 27 L]

$$g_{\sigma\alpha}\,\Theta_{\alpha\beta} \qquad -\frac{\partial g_{\mu\nu}}{\partial x_\sigma}\gamma_{\alpha\mu}\gamma_{\beta\nu}$$

$$g_{\sigma\alpha}\gamma_{\alpha\mu}\gamma_{\beta\nu}T_{\mu\nu} \qquad +\frac{\partial\gamma_{\alpha\mu}^{\beta}}{\partial x_\sigma}T_{\alpha\beta}$$

$$\frac{\partial}{\partial x_\nu}(\gamma_{\beta\mu}T_{\mu\sigma})$$

Divergenz antisymmetrischer Tensoren

$$T_\nu = \sum_{s\mu}\frac{\partial T_{\nu s}}{\partial x_s} + \cancel{\gamma_{\nu\alpha}\left(\frac{\partial g_{s\alpha}}{\partial}\right.}\begin{Bmatrix} s\mu \\ \nu \end{Bmatrix}T_{\mu s} + \begin{Bmatrix} s\mu \\ s \end{Bmatrix}T_{\nu\mu}$$

$$\frac{1}{2}\gamma_{\nu\alpha}\left(\frac{\partial g_{s\alpha}}{\partial x_\mu}+\frac{\partial g_{\mu\alpha}}{\partial x_s}-\frac{\partial g_{\mu s}}{\partial x_\alpha}\right)T_{\mu s} \qquad \cancel{-\begin{Bmatrix} \mu s \\ \nu \end{Bmatrix}T_{s\mu}}\;\frac{1}{2}\gamma_{s\alpha}\left(\frac{\partial g_{s\alpha}}{\partial x_\mu}+\frac{\partial g_{\alpha\mu}}{\partial x_s}-\frac{\partial g_{s\mu}}{\partial x_\alpha}\right)T_{\nu\mu}$$

μ, s vertauscht

verschwindet, wenn T antisymmetr.

$$\begin{Bmatrix} s\mu \\ \mu \end{Bmatrix}T_{\nu s}$$

$$\gamma_{\mu\alpha}\left(\frac{\partial g_{s\alpha}}{\partial x_\mu}+\frac{\partial g_{\mu\alpha}}{\partial x_s}-\frac{\partial g_{s\mu}}{\partial x_\alpha}\right)T_{\nu s}$$

$$\frac{\partial G}{\partial x_s}T_{\nu s} \qquad 0$$

$$\sum\frac{\partial T_{\nu s}}{\partial x_s}+\frac{1}{2}\frac{\partial G}{\partial x_s}T_{\nu s}$$

$$= \sqrt{G}\frac{\partial T_{\nu s}}{\partial x_s}+\frac{G}{2\sqrt{G}}\frac{\partial G}{\partial x_s}T_{\nu s} = \frac{1}{\sqrt{G}}\sum_s\frac{\partial(\sqrt{G}\,T_{\nu s})}{\partial x_s}$$

Elektr. Menge ist Skalar ebenso wahre el. Dichte

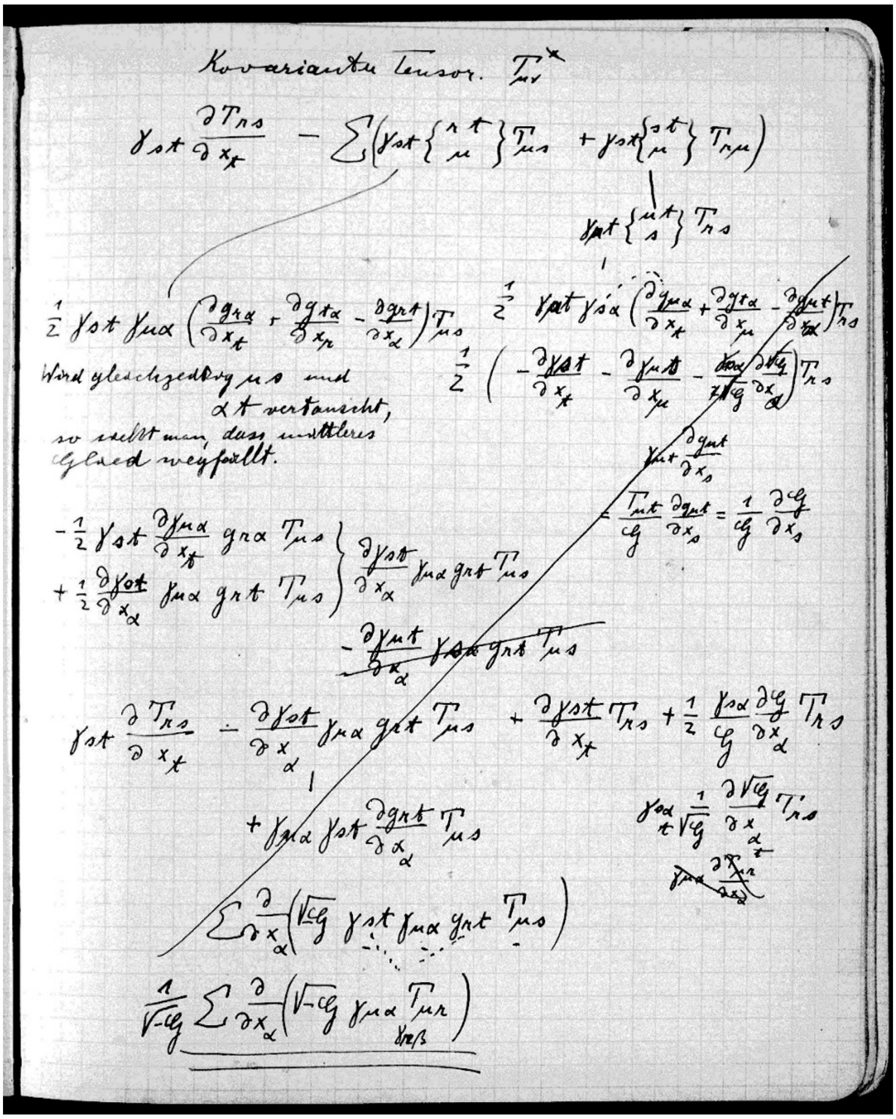
$$\rho_0\frac{dx_\nu}{ds} = \text{kontravarianter Vektor} = \frac{\rho}{\sqrt{-g}}\frac{ds}{dt}\frac{dx_\nu}{ds}$$

$$\rho = \frac{e_0}{V} = \frac{e_0}{V_0}\frac{dt}{ds}\sqrt{-g} = \rho_0\sqrt{-g}\frac{dt}{ds} \qquad \frac{\rho}{\sqrt{-g}}\frac{dx_\nu}{dt} \quad \frac{\rho}{\sqrt{-g}}$$

ist kontravarianter Vektor.

Hieraus Feldgleichungen (1. System.

[p. 27 R]


Kovarianter Tensor. $T_{\mu\nu}^{*}$

$$\gamma_{st}\frac{\partial T_{rs}}{\partial x_t} - \sum\left(\gamma_{st}\left\{{rt \atop \mu}\right\}T_{\mu s} + \gamma_{st}\left\{{st \atop \mu}\right\}T_{r\mu}\right)$$

$$\gamma_{\mu t}\left\{{\mu t \atop s}\right\}T_{rs}$$

$$\frac{1}{2}\gamma_{st}\gamma_{\mu\alpha}\left(\frac{\partial g_{r\alpha}}{\partial x_t}+\frac{\partial g_{t\alpha}}{\partial x_r}-\frac{\partial g_{rt}}{\partial x_\alpha}\right)T_{\mu s}$$

$$\frac{1}{2}\gamma_{\mu t}\gamma_{s\alpha}\left(\frac{\partial g_{\mu\alpha}}{\partial x_t}+\frac{\partial g_{t\alpha}}{\partial x_\mu}-\frac{\partial g_{\mu t}}{\partial x_\alpha}\right)T_{rs}$$

Wird gleichzeitig μs und αt vertauscht, so sieht man, dass mittleres Glied wegfällt.

$$\frac{1}{2}\left(-\frac{\partial \gamma_{st}}{\partial x_t}-\frac{\partial \gamma_{\mu t}}{\partial x_\mu}-\frac{\gamma_{s\alpha}}{2g}\frac{\partial g}{\partial x_\alpha}\right)T_{rs}$$

$$\gamma_{\mu t}\frac{\partial g_{\mu t}}{\partial x_s}$$

$$=\frac{T_{\mu t}}{g}\frac{\partial g_{\mu t}}{\partial x_s}=\frac{1}{g}\frac{\partial g}{\partial x_s}$$

$$\left.\begin{array}{l}-\frac{1}{2}\gamma_{st}\frac{\partial\gamma_{\mu\alpha}}{\partial x_t}g_{r\alpha}T_{\mu s}\\ +\frac{1}{2}\frac{\partial\gamma_{st}}{\partial x_\alpha}\gamma_{\mu\alpha}g_{rt}T_{\mu s}\end{array}\right\}\frac{\partial\gamma_{st}}{\partial x_\alpha}\gamma_{\mu\alpha}g_{rt}T_{\mu s}$$

$$\cancel{-\frac{\partial\gamma_{\mu t}}{\partial x_\alpha}\gamma_{s\alpha}g_{rt}T_{\mu s}}$$

$$\gamma_{st}\frac{\partial T_{rs}}{\partial x_t}-\frac{\partial\gamma_{st}}{\partial x_\alpha}\gamma_{\mu\alpha}g_{rt}T_{\mu s}+\frac{\partial\gamma_{st}}{\partial x_t}T_{rs}+\frac{1}{2}\frac{\gamma_{s\alpha}}{g}\frac{\partial g}{\partial x_\alpha}T_{rs}$$

$$+\gamma_{\mu\alpha}\gamma_{st}\frac{\partial g_{rt}}{\partial x_\alpha}T_{\mu s}$$

$$\gamma_{s\alpha}\frac{1}{\sqrt{g}}\frac{\partial\sqrt{g}}{\partial x_\alpha}T_{rs}$$

$$\cancel{\gamma_{\mu\alpha}\frac{\partial T_{\mu r}}{\partial x_\alpha}}$$

$$\sum\frac{\partial}{\partial x_\alpha}\left(\sqrt{-g}\,\gamma_{st}\gamma_{\mu\alpha}g_{rt}T_{\mu s}\right)$$

$$\frac{1}{\sqrt{-g}}\sum\frac{\partial}{\partial x_\alpha}\left(\sqrt{-g}\,\gamma_{\mu\alpha}\underset{\mathfrak{T}_{\alpha\beta}}{T_{\mu r}}\right)$$

[p. 27 L]

$$g_{\sigma\alpha}\Theta_{\alpha\beta} \qquad\qquad -\frac{\partial g_{\mu\nu}}{\partial x_\sigma}\gamma_{\alpha\mu}\gamma_{\beta\nu}$$

$$g_{\sigma\alpha}\gamma_{\alpha\mu}\gamma_{\beta\nu}T_{\mu\nu} \qquad\qquad +\frac{\partial \gamma_{\alpha\mu}^{\ \beta}}{\partial x_\sigma}T_{\alpha\beta}$$

$$\frac{\partial}{\partial x_\beta}(\gamma_{\beta\mu}T_{\mu\sigma})$$

Divergenz antisymmetrischer Tensoren

$$T_r = \sum_{s\mu}\frac{\partial T_{rs}}{\partial x_s} + \langle\gamma_{r\alpha}\Big(\frac{\partial g_{s\alpha}}{\partial}\rangle\begin{Bmatrix} s & \mu \\ r \end{Bmatrix}T_{\mu s} + \begin{Bmatrix} s & \mu \\ s \end{Bmatrix}T_{r\mu}$$

$$\frac{1}{2}\gamma_{r\alpha}\Big(\frac{\partial g_{s\alpha}}{\partial x_\mu}+\frac{\partial g_{\mu\alpha}}{\partial x_s}-\frac{\partial g_{\mu s}}{\partial x_\alpha}\Big)T_{\mu s} \qquad \cancel{-\begin{Bmatrix} \mu & s \\ r \end{Bmatrix}}T_{s\mu} \quad \frac{1}{2}\gamma_{s\alpha}\Big(\frac{\partial g_{s\alpha}}{\partial x_\mu}+\frac{\partial g_{\alpha\mu}}{\partial x_s}-\frac{\partial g_{s\mu}}{\partial x_\alpha}\Big)T_{r\mu}$$

verschwin~~det~~,
wenn $\cancel{T}$ antisymmetr.

μ u s vertauscht

$$\begin{Bmatrix} s & \mu \\ \mu \end{Bmatrix}T_{rs}$$

$$\gamma_{\mu\alpha}\Big(\frac{\partial g_{s\alpha}}{\partial x_\mu}+\frac{\partial g_{\mu\alpha}}{\partial x_s}-\frac{\partial g_{s\mu}}{\partial x_\alpha}\Big)T_{rs}$$

$$\frac{\partial G}{\partial x_s}T_{rs} \qquad\qquad 0$$

$$\sum\frac{\partial T_{rs}}{\partial x_s}+\frac{1}{2}\frac{\partial G}{\partial x_s}T_{rs}$$

$$= \sqrt{G}\frac{\partial T_{rs}}{\partial x_s}+\frac{G}{2\sqrt{G}}\frac{\partial G}{\partial x_s}T_{rs} = \frac{1}{\sqrt{-G}}\sum_s\frac{\partial(\sqrt{-G}T_{rs})}{\partial x_s}$$

Elektr. Menge ist Skalar ebenso wahre el. Dichte

$$\rho_0\frac{dx_\nu}{ds} = \text{kontravarianter Vektor} = \frac{\rho}{\sqrt{-G}}\ \frac{ds}{dt}\ \frac{dx_\nu}{ds}$$

$$\rho_0 = \frac{e_0}{V} = \frac{e_0}{V_0}\frac{dt}{ds}\sqrt{-G} = \rho_0\sqrt{-G}\frac{dt}{ds} \qquad \frac{\rho}{\sqrt{-G}}\frac{dx_\nu}{dt}\ \cdot\ \cdot\ \frac{\rho}{\sqrt{-G}}$$

ist kontravarianter Vektor.

Hieraus Feldgleichungen (1. System.

[p. 27 R]

Kovarianter Tensor. $T_{\mu\nu}^{\langle x\rangle}$

$$\gamma_{st}\frac{\partial T_{rs}}{\partial x_t}-\sum\left(\gamma_{st}\begin{Bmatrix} r & t \\ & \mu \end{Bmatrix}T_{\mu s}+\gamma_{st}\begin{Bmatrix} s & t \\ & \mu \end{Bmatrix}T_{r\mu}\right)$$

$$\gamma_{\mu t}\begin{Bmatrix} \mu & t \\ & s \end{Bmatrix}T_{rs}$$

$$\frac{1}{2}\gamma_{st}\gamma_{\mu\alpha}\left(\frac{\partial g_{r\alpha}}{\partial x_t}+\frac{\partial g_{t\alpha}}{\partial x_r}-\frac{\partial g_{rt}}{\partial x_\alpha}\right)T_{\mu s}$$

$$\frac{1}{2}\gamma_{\mu t}\gamma_{s\alpha}\left(\frac{\partial g_{\mu\alpha}}{\partial x_t}+\frac{\partial g_{t\alpha}}{\partial x_\mu}-\frac{\partial g_{\mu t}}{\partial x_\alpha}\right)T_{rs}$$

Wird gleichzeitig μs und
αt vertauscht,
so sieht man, dass mittleres
Glied wegfällt.

$$\frac{1}{2}\left(-\frac{\partial\gamma_{st}}{\partial x_t}-\frac{\partial\gamma_{\mu t}}{\partial x_\mu}-\frac{\gamma_{s\alpha}}{\langle 2\sqrt{}\rangle G}\frac{\partial\langle\sqrt{}\rangle G}{\partial x_\alpha}\right)T_{rs}$$

$$\gamma_{\mu t}\frac{\partial g_{\mu t}}{\partial x_s}$$

$$=\frac{\Gamma_{\mu t}}{G}\frac{\partial g_{\mu t}}{\partial x_s}=\frac{1}{G}\frac{\partial G}{\partial x_s}$$

$$\left.\begin{aligned}&-\frac{1}{2}\gamma_{st}\frac{\partial\gamma_{\mu\alpha}}{\partial x_t}g_{r\alpha}T_{\mu s}\\&+\frac{1}{2}\frac{\partial\gamma_{st}}{\partial x_\alpha}\gamma_{\mu\alpha}g_{rt}T_{\mu s}\end{aligned}\right\}\ \frac{\partial\gamma_{st}}{\partial x_\alpha}\gamma_{\mu\alpha}g_{rt}T_{\mu s}$$

$$\cancel{-\frac{\partial\gamma_{\mu t}}{\partial x_\alpha}\gamma_{s\alpha}g_{rt}T_{\mu s}}$$

$$\gamma_{st}\frac{\partial T_{rs}}{\partial x_t}\ -\ \frac{\partial\gamma_{st}}{\partial x_\alpha}\gamma_{\mu\alpha}g_{rt}T_{\mu s}\ +\ \frac{\partial\gamma_{st}}{\partial x_t}T_{rs}\ +\ \frac{1}{2}\frac{\gamma_{s\alpha}}{G}\frac{\partial G}{\partial x_\alpha}T_{rs}$$

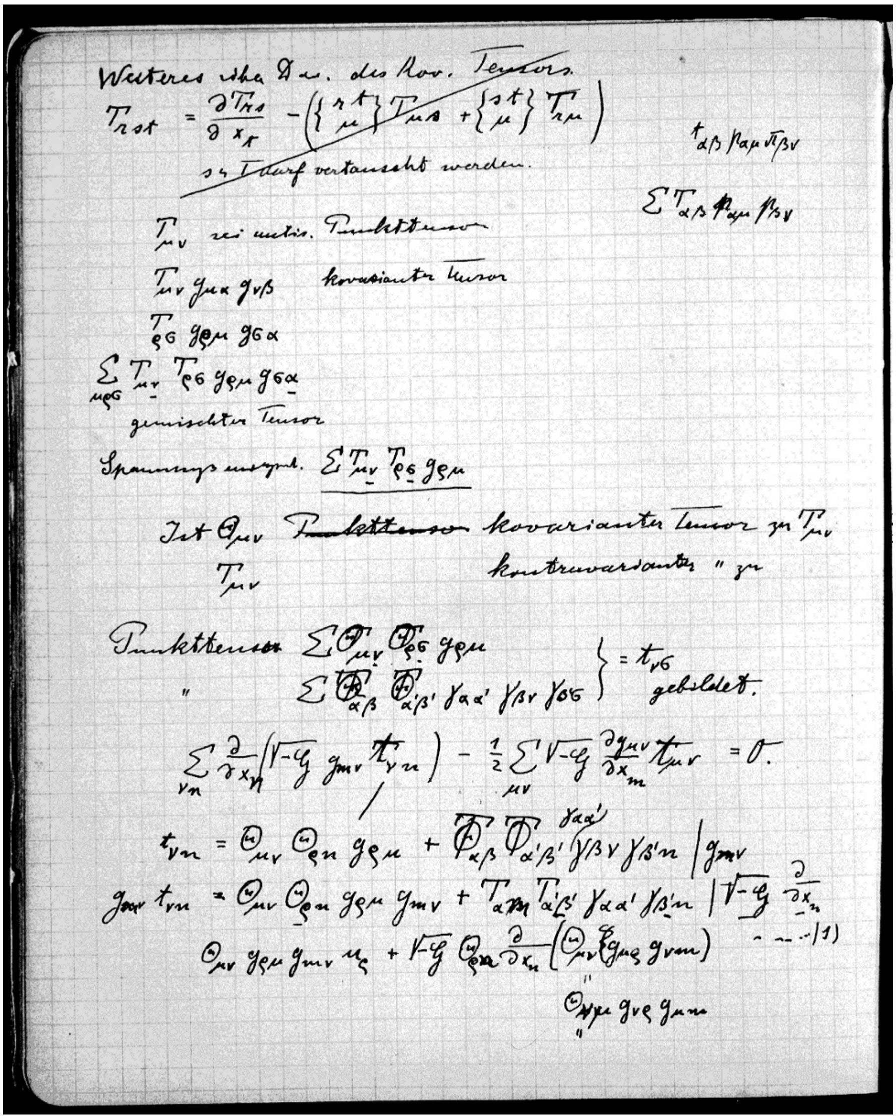
$$+\gamma_{\mu\alpha}\gamma_{st}\frac{\partial g_{rt}}{\partial x_\alpha}T_{\mu s}\qquad\qquad \gamma_{s\alpha}\frac{1}{\sqrt{G}}\frac{\partial\sqrt{G}}{\partial x_\alpha}T_{rs}$$

$$\cancel{\gamma_{\mu\alpha}\frac{\partial T_{\mu r}}{\partial x_\alpha}}$$

$$\sum\frac{\partial}{\partial x_\alpha}(\sqrt{-G}\gamma_{st}\gamma_{\mu\alpha}g_{rt}T_{\mu s})$$

$$\frac{1}{\sqrt{-G}}\sum\frac{\partial}{\partial x_\alpha}\left(\sqrt{-G}\gamma_{\mu\alpha}T_{\mu r}\right)\quad \gamma_{r\beta}$$

[p. 28 L]



Weiteres über Div. des kov. ~~Tensors.~~

$$\cancel{T_{rst} = \frac{\partial T_{rs}}{\partial x_t} - \left(\left\{ {rt \atop \mu} \right\} T_{\mu s} + \left\{ {st \atop \mu} \right\} T_{r\mu} \right)}$$

~~r, s darf vertauscht werden.~~

$t_{\alpha\beta} \gamma_{\alpha\mu} \gamma_{\beta\nu}$

$\sum T_{\alpha\beta} \gamma_{\alpha\mu} \gamma_{\beta\nu}$

$T_{\mu\nu}$ sei antis. Punkttensor

$T_{\mu\nu} g_{\mu\alpha} g_{\nu\beta}$ kovarianter Tensor

$T_{\varrho\sigma} g_{\varrho\mu} g_{\sigma\alpha}$

$\sum_{\mu\varrho\sigma} T_{\mu\underline{\nu}} T_{\varrho\sigma} g_{\varrho\mu} g_{\sigma\underline{\alpha}}$

gemischter Tensor

Spannungs unsymmet. $\underline{\sum T_{\mu\underline{\nu}} T_{\varrho\underline{\sigma}} g_{\varrho\mu}}$

Ist $\Theta_{\mu\nu}$ ~~Punkttensor~~ kovarianter Tensor zu $T_{\mu\nu}$

$T_{\mu\nu}$ kontravarianter " zu

~~Punkttensor~~ $\sum \Theta_{\mu\underline{\nu}} \Theta_{\varrho\underline{\sigma}} g_{\varrho\mu}$

" $\sum \Phi_{\alpha\beta} \Phi_{\alpha'\beta'} \gamma_{\alpha\alpha'} \gamma_{\beta\nu} \gamma_{\beta'\sigma}$ $\Big\} = t_{\nu\sigma}$ gebildet.

$$\sum_{\nu n} \frac{\partial}{\partial x_n} \left(\sqrt{-g}\, g_{m\nu} t_{\nu n} \right) - \frac{1}{2} \sum_{\mu\nu} \sqrt{-g} \frac{\partial g_{\mu\nu}}{\partial x_m} t_{\mu\nu} = 0.$$

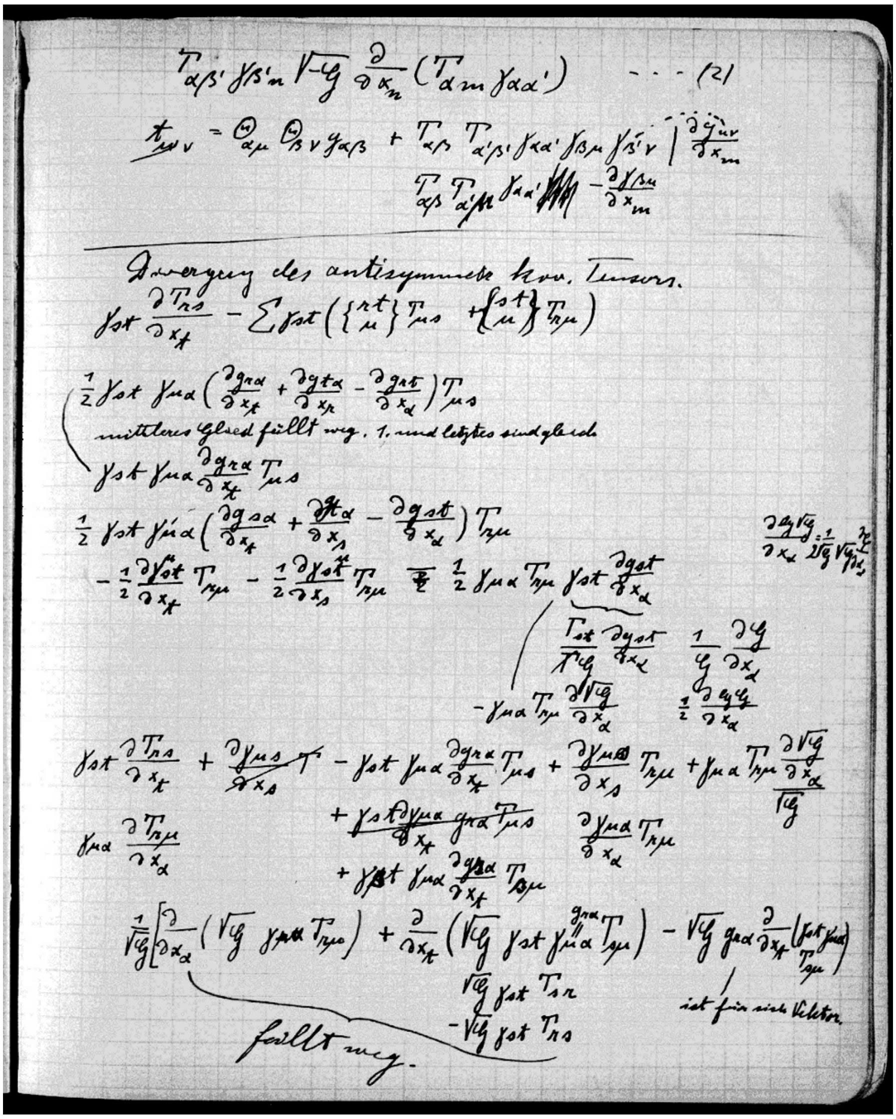
$$t_{\nu n} = \Theta_{\mu\nu} \Theta_{\varrho n} g_{\varrho\mu} + \Phi_{\alpha\beta} \Phi_{\alpha'\beta'} \gamma_{\alpha\alpha'} \gamma_{\beta\nu} \gamma_{\beta' n} \ \Big|\ g_{m\nu}$$

$$g_{m\nu} t_{\nu n} = \Theta_{\mu\nu} \Theta_{\underline{\varrho} n} g_{\varrho\mu} g_{m\nu} + T_{\alpha m} T_{\alpha'\beta'} \gamma_{\alpha\alpha'} \gamma_{\beta' \underline{n}} \ \Big|\ \sqrt{-g} \frac{\partial}{\partial x_{\underline{n}}}$$

$$\Theta_{\mu\nu} g_{\varrho\mu} g_{m\nu} u_\varrho + \sqrt{-g}\, \Theta_{\varrho n} \frac{\partial}{\partial x_n} \left(\Theta_{\mu\nu} g_{\mu\varrho} g_{\nu m} \right) \quad \cdots (1)$$

$$\Theta_{\nu\mu} g_{\nu\varrho} g_{\mu m}$$

[p. 28 R]


$$T_{\alpha\beta'}\gamma_{\beta' n}\sqrt{-g}\,\frac{\partial}{\partial x_n}(T_{\alpha m}\gamma_{\alpha\alpha'}) \quad \ldots (2)$$

$$t_{\mu\nu} = \Theta_{\alpha\mu}\Theta_{\beta\nu}\gamma_{\alpha\beta} + T_{\alpha\beta}T_{\alpha'\beta'}\gamma_{\alpha\alpha'}\gamma_{\beta\mu}\gamma_{\beta'\nu} \;\Big|\; \frac{\partial g_{\mu\nu}}{\partial x_m}$$

$$T_{\alpha\beta}T_{\alpha'\mu}\gamma_{\alpha\alpha'} \quad -\frac{\partial\gamma_{\beta\mu}}{\partial x_m}$$

Divergenz des antisymmetr. kov. Tensors.

$$\gamma_{st}\frac{\partial T_{rs}}{\partial x_t} - \sum\gamma_{st}\left(\left\{{rt \atop \mu}\right\}T_{\mu s} + \left\{{st \atop \mu}\right\}T_{r\mu}\right)$$

$$\frac{1}{2}\gamma_{st}\gamma_{\mu\alpha}\left(\frac{\partial g_{r\alpha}}{\partial x_t} + \frac{\partial g_{t\alpha}}{\partial x_r} - \frac{\partial g_{rt}}{\partial x_\alpha}\right)T_{\mu s}$$

mittleres Glied fällt weg. 1. und letztes sind gleich

$$\gamma_{st}\gamma_{\mu\alpha}\frac{\partial g_{r\alpha}}{\partial x_t}T_{\mu s}$$

$$\frac{1}{2}\gamma_{st}\gamma_{\mu\alpha}\left(\frac{\partial g_{s\alpha}}{\partial x_t} + \frac{\partial g_{t\alpha}}{\partial x_s} - \frac{\partial g_{st}}{\partial x_\alpha}\right)T_{r\mu}$$

$$\frac{\partial \lg\sqrt{g}}{\partial x_\alpha} = \frac{1}{2\sqrt{g}}\,\frac{1}{\sqrt{g}}\frac{\partial g}{\partial x_\alpha}$$

$$-\frac{1}{2}\frac{\partial\gamma_{\mu t}}{\partial x_t}T_{r\mu} - \frac{1}{2}\frac{\partial\gamma_{s\mu}}{\partial x_s}T_{r\mu} \mp \frac{1}{2}\gamma_{\mu\alpha}T_{r\mu}\gamma_{st}\frac{\partial g_{st}}{\partial x_\alpha}$$

$$\frac{T_{st}}{\sqrt{g}}\frac{\partial g_{st}}{\partial x_\alpha} \qquad \frac{1}{g}\frac{\partial g}{\partial x_\alpha}$$

$$-\gamma_{\mu\alpha}T_{r\mu}\frac{\partial\sqrt{g}}{\partial x_\alpha} \qquad \frac{1}{2}\frac{\partial \lg g}{\partial x_\alpha}$$

$$\gamma_{st}\frac{\partial T_{rs}}{\partial x_t} + \cancel{\frac{\partial\gamma_{\mu s}}{\partial x_s}T} - \gamma_{st}\gamma_{\mu\alpha}\frac{\partial g_{r\alpha}}{\partial x_t}T_{\mu s} + \frac{\partial\gamma_{\mu s}}{\partial x_s}T_{r\mu} + \gamma_{\mu\alpha}T_{r\mu}\frac{\frac{\partial\sqrt{g}}{\partial x_\alpha}}{\sqrt{g}}$$

$$\gamma_{\mu\alpha}\frac{\partial T_{r\mu}}{\partial x_\alpha} \qquad + \cancel{\gamma_{st}\frac{\partial\gamma_{\mu\alpha}}{\partial x_t}g_{r\alpha}T_{\mu s}} \qquad \frac{\partial\gamma_{\mu\alpha}}{\partial x_\alpha}T_{r\mu}$$

$$+\gamma_{st}\gamma_{\mu\alpha}\frac{\partial g_{s\alpha}}{\partial x_t}T_{s\mu}$$

$$\frac{1}{\sqrt{g}}\left[\frac{\partial}{\partial x_\alpha}(\sqrt{g}\,\gamma_{\mu\alpha}T_{r\mu}) + \frac{\partial}{\partial x_t}\left(\sqrt{g}\,\gamma_{st}\gamma_{\mu\alpha}^{g_{r\alpha}}T_{s\mu}\right) - \sqrt{g}\,g_{r\alpha}\frac{\partial}{\partial x_t}(\gamma_{st}\gamma_{\mu\alpha}T_{s\mu})\right.$$

$$\sqrt{g}\,\gamma_{st}T_{sr}$$

$$-\sqrt{g}\,\gamma_{st}T_{rs}$$

fällt weg.

ist für sich Vektor.

[p. 28 L]

Weiteres über Div. des Kov. Tensors.

$$T_{rst} = \frac{\partial T_{rs}}{\partial x_t} - \left(\begin{Bmatrix} r & t \\ & \mu \end{Bmatrix} T_{\mu s} + \begin{Bmatrix} s & t \\ & \mu \end{Bmatrix} T_{r\mu} \right)$$

s u T darf vertauscht werden.

$t_{\alpha\beta} p_{\alpha\mu} \pi_{\beta\nu}$

$\sum T_{\alpha\beta} p_{\alpha\mu} p_{\beta\nu}$

$T_{\mu\nu}$ sei antis. Punkttensor

$T_{\mu\nu} g_{\mu\alpha} g_{\nu\beta}$ kovarianter Tensor

$T_{\rho\sigma} g_{\rho\mu} g_{\sigma\alpha}$

$$\sum_{\mu\rho\sigma} T_{\mu\underline{\nu}} T_{\rho\sigma} g_{\rho\mu} g_{\sigma\underline{\alpha}}$$

gemischter Tensor

Spannungs energiet. $\underline{\sum T_{\mu\underline{\nu}} T_{\rho\underline{\sigma}} g_{\rho\mu}}$

Ist $\Theta_{\mu\nu}$ <Punkttensor> kovarianter Tensor zu $T_{\mu\nu}$

$T_{\mu\nu}$ kontravarianter " zu

$$\left.\begin{array}{ll} \text{Punkttensor} & \sum \Theta_{\mu\underline{\nu}} \Theta_{\rho\underline{\sigma}} g_{\rho\mu} \\ \quad " & \sum T_{\alpha\beta} T_{\alpha'\beta'} \gamma_{\alpha\alpha'} \gamma_{\beta\nu} \gamma_{\beta'\sigma} \end{array}\right\} = t_{\nu\sigma} \text{ gebildet.}$$

$$\sum_{\nu n} \frac{\partial}{\partial x_n} (\sqrt{-G} g_{m\nu} t_{\nu n}) - \frac{1}{2} \sum_{\mu\nu} \sqrt{-G} \frac{\partial g_{\mu\nu}}{\partial x_m} t_{\mu\nu} = 0.$$

$$t_{\nu n} = \Theta_{\mu\nu} \Theta_{\rho n} g_{\rho\mu} + T_{\alpha\beta} T_{\alpha'\beta'} \overset{\gamma_{\alpha\alpha'}}{\gamma_{\beta\nu}} \gamma_{\beta' n} \Big| g_{m\nu}$$

$$g_{m\nu} t_{\nu n} = \Theta_{\mu\nu} \Theta_{\underline{\rho} n} g_{\rho\mu} g_{m\nu} + T_{\alpha m} T_{\alpha'\underline{\beta}'} \gamma_{\alpha\alpha'} \gamma_{\underline{\beta}' n} \Big| \sqrt{-\underline{G}} \frac{\partial}{\underline{\partial x}_n}$$

$$\Theta_{\mu\nu} g_{\rho\mu} g_{m\nu} u_\rho + \sqrt{-G} \Theta_{\rho n} \frac{\partial}{\partial x_n} (\Theta_{\mu\nu} g_{\mu\rho} g_{\nu m}) \cdots (1)$$

$$"$$

$$\Theta_{\nu\mu} g_{\nu\rho} g_{\mu m}$$

$$"$$

[p. 28 R]

$$T_{\alpha\beta'}\gamma_{\beta'n}\sqrt{-G}\frac{\partial}{\partial x_n}(T_{\alpha m}\gamma_{\alpha\alpha'}) \quad \cdots \quad (2)$$

$$t_{\mu\nu} = \Theta_{\alpha\mu}\Theta_{\beta\nu}g_{\alpha\beta} + T_{\alpha\beta}T_{\alpha'\beta'}\gamma_{\alpha\alpha'}\gamma_{\beta\mu}\gamma_{\beta'\nu} \mid \frac{\partial g_{\mu\nu}}{\partial x_m}$$

$$T_{\alpha\beta}T_{\alpha'\mu}\gamma_{\alpha\alpha'}\langle\gamma_{\beta\beta'}\rangle - \frac{\partial\gamma_{\beta\mu}}{\partial x_m}$$

Divergenz des antisymmetr kov. Tensors.

$$\gamma_{st}\frac{\partial T_{rs}}{\partial x_t} - \sum\gamma_{st}\left(\begin{Bmatrix} r & t \\ & \mu \end{Bmatrix}T_{\mu s} + \begin{Bmatrix} s & t \\ & \mu \end{Bmatrix}T_{r\mu}\right)$$

$$\frac{1}{2}\gamma_{st}\gamma_{\mu\alpha}\left(\frac{\partial g_{r\alpha}}{\partial x_t} + \frac{\partial g_{t\alpha}}{\partial x_r} - \frac{\partial g_{rt}}{\partial x_\alpha}\right)T_{\mu s}$$

mittleres Glied fällt weg. 1. und letztes sind gleich

$$\gamma_{st}\gamma_{\mu\alpha}\frac{\partial g_{r\alpha}}{\partial x_t}T_{\mu s}$$

$$\frac{1}{2}\gamma_{st}\gamma_{\mu\alpha}\left(\frac{\partial g_{s\alpha}}{\partial x_t} + \frac{\partial g_{t\alpha}}{\partial x_s} - \frac{\partial g_{st}}{\partial x_\alpha}\right)T_{r\mu} \qquad \frac{\partial \lg\sqrt{G}}{\partial x_\alpha} = \frac{1}{2\sqrt{G}}\ \sqrt{G}\frac{\partial G}{\partial x_\alpha}$$

$$-\frac{1}{2}\frac{\partial\gamma^{\mu}_{st}}{\partial x_t}T_{r\mu} - \frac{1}{2}\frac{\partial\gamma^{\mu}_{st}}{\partial x_s}T_{r\mu} - \frac{1}{2}\gamma_{\mu\alpha}T_{r\mu}\underbrace{\gamma_{st}\frac{\partial g_{st}}{\partial x_\alpha}}$$

$$\frac{\Gamma_{st}}{\langle\Gamma\rangle G}\ \frac{\partial g_{st}}{\partial x_\alpha} \qquad \frac{1}{G}\frac{\partial G}{\partial x_\alpha}$$

$$-\gamma_{\mu\alpha}T_{r\mu}\frac{\partial\sqrt{G}}{\partial x_\alpha} \qquad \frac{1}{2}\frac{\partial \lg G}{\partial x_\alpha}$$

$$\gamma_{st}\frac{\partial T_{rs}}{\partial x_t} + \cancel{\frac{\partial\gamma_{\mu s}}{\partial x_s}T} - \gamma_{st}\gamma_{\mu\alpha}\frac{\partial g_{r\alpha}}{\partial x_t}T_{\mu s} + \frac{\partial\gamma_{\mu s}}{\partial x_s}T_{r\mu} + \gamma_{\mu\alpha}T_{r\mu}\frac{\frac{\partial\sqrt{G}}{\partial x_\alpha}}{\sqrt{G}}$$

$$\gamma_{\mu\alpha}\frac{\partial T_{r\mu}}{\partial x_\alpha} \qquad + \cancel{\gamma_{st}\frac{\partial\gamma_{\mu\alpha}}{\partial x_t}g_{r\alpha}T_{\mu s}} \quad \frac{\partial\gamma_{\mu\alpha}}{\partial x_\alpha}T_{r\mu}$$

$$+ \gamma_{st}\gamma_{\mu\alpha}\frac{\partial g_{s\alpha}}{\partial x_t}T_{s\mu}$$

$$\frac{1}{\sqrt{G}}\Big[\underbrace{\frac{\partial}{\partial x_\alpha}(\sqrt{G}\gamma_{\mu\alpha}T_{r\mu}) + \frac{\partial}{\partial x_t}(\sqrt{G}\gamma_{st}\overset{g_{r\alpha}}{\cancel{\gamma_{\mu\alpha}}}T_{s\mu})}_{\text{fällt weg.}} - \sqrt{G}g_{r\alpha}\frac{\partial}{\partial x_t}(\gamma_{st}\gamma_{\mu\alpha}T_{s\mu})$$

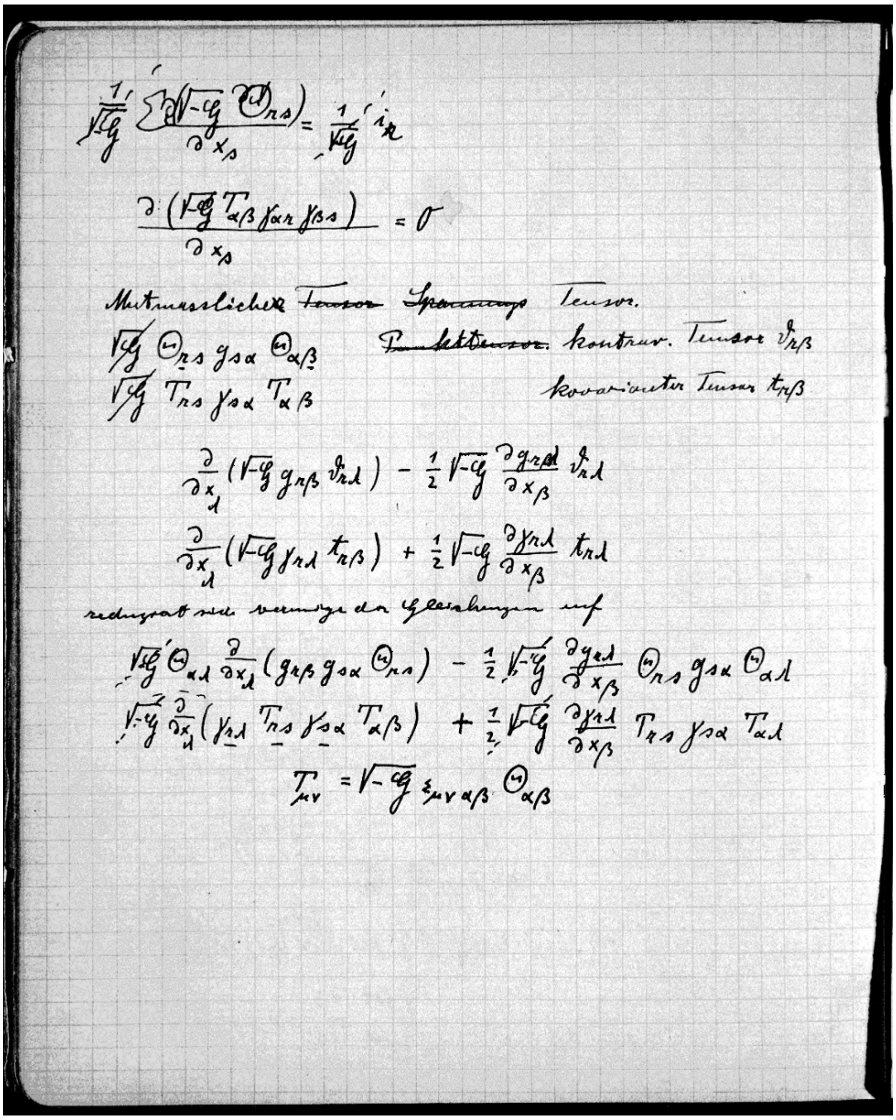
$$\sqrt{G}\gamma_{st}T_{sr}$$

$$-\sqrt{G}\gamma_{st}T_{rs}$$

ist für sich Vektor.

fällt weg.

[p. 29 L]



$$\frac{1}{\sqrt{G}}\,\frac{\sum \partial(\sqrt{-G}\,\Theta_{\kappa\sigma})}{\partial x_\sigma} = \frac{1}{\sqrt{G}}\, i_\kappa$$

$$\frac{\partial(\sqrt{G}\,T_{\alpha\beta}\gamma_{\alpha\kappa}\gamma_{\beta\sigma})}{\partial x_\sigma} = 0$$

Mutmassliche ~~Tensor~~ ~~Spannungs~~ Tensor.

$\sqrt{G}\,\Theta_{\kappa\sigma}\, g_{\sigma\alpha}\,\Theta_{\alpha\beta}$ ~~Punkttensor.~~ kontrav. Tensor $\vartheta_{\kappa\beta}$

$\sqrt{G}\,T_{\kappa\sigma}\,\gamma_{\sigma\alpha}\,T_{\alpha\beta}$ kovarianter Tensor $t_{\kappa\beta}$

$$\frac{\partial}{\partial x_\lambda}(\sqrt{-G}\,g_{\kappa\beta}\,\vartheta_{\kappa\lambda}) - \frac{1}{2}\sqrt{-G}\,\frac{\partial g_{\kappa\lambda}}{\partial x_\beta}\,\vartheta_{\kappa\lambda}$$

$$\frac{\partial}{\partial x_\lambda}(\sqrt{-G}\,\gamma_{\kappa\lambda}\,t_{\kappa\beta}) + \frac{1}{2}\sqrt{-G}\,\frac{\partial \gamma_{\kappa\lambda}}{\partial x_\beta}\,t_{\kappa\lambda}$$

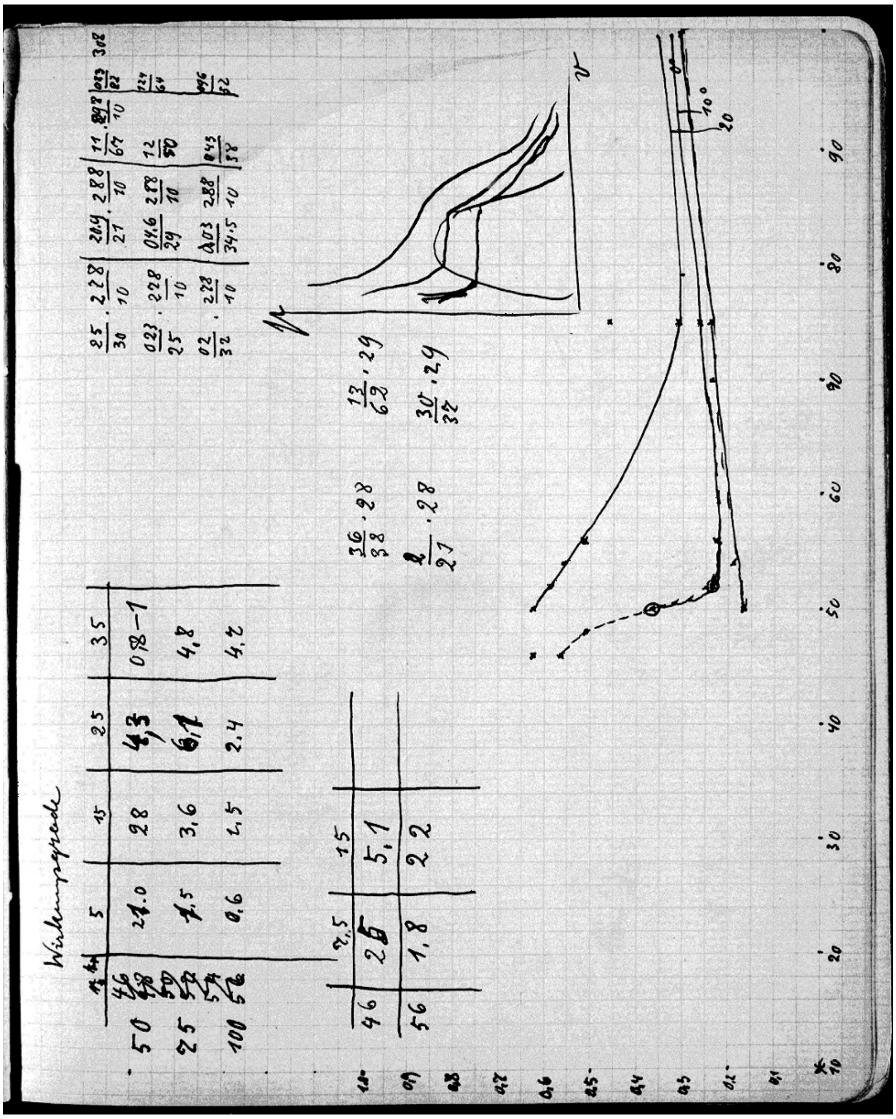
reduziert sich vermöge der Gleichungen auf

$$\sqrt{G}\,\Theta_{\alpha\lambda}\,\frac{\partial}{\partial x_\lambda}(g_{\kappa\beta}\,g_{\sigma\alpha}\,\Theta_{\kappa\sigma}) - \frac{1}{2}\sqrt{-G}\,\frac{\partial g_{\kappa\lambda}}{\partial x_\beta}\,\Theta_{\kappa\sigma}\,g_{\sigma\alpha}\,\Theta_{\alpha\lambda}$$

$$\sqrt{-G}\,\frac{\partial}{\partial x_\lambda}(\gamma_{\kappa\lambda}\,T_{\kappa\sigma}\,\gamma_{\sigma\alpha}\,T_{\alpha\beta}) + \frac{1}{2}\sqrt{-G}\,\frac{\partial \gamma_{\kappa\lambda}}{\partial x_\beta}\,T_{\kappa\sigma}\,\gamma_{\sigma\alpha}\,T_{\alpha\lambda}$$

$$T_{\mu\nu} = \sqrt{-G}\,\varepsilon_{\mu\nu\alpha\beta}\,\Theta_{\alpha\beta}$$

[p. 29 R]



[p. 29 L]

$$\frac{1}{\sqrt{-G}}\sum\frac{\partial(\sqrt{-G}\Theta_{rs})}{\partial x_s} = \frac{1}{\sqrt{G}}i_r$$

$$\frac{\partial(\sqrt{-G}T_{\alpha\beta}\gamma_{\alpha r}\gamma_{\beta s})}{\partial x_s} = 0$$

Mutmassliche ⟨r⟩ ⟨Tensor Spannungs⟩ Tensor.

$\cancel{\sqrt{G}}\ \Theta_{\underline{r}s}g_{s\alpha}\Theta_{\alpha\underline{\beta}}$ ⟨Punkttensor.⟩ kontrav. Tensor $\vartheta_{r\beta}$

$\cancel{\sqrt{G}}\ T_{rs}\gamma_{s\alpha}T_{\alpha\beta}$ kovarianter Tensor $t_{r\beta}$

$$\frac{\partial}{\partial x_\lambda}(\sqrt{-G}g_{r\beta}\vartheta_{r\lambda}) - \frac{1}{2}\sqrt{-G}\frac{\partial g_{r\lambda}}{\partial x_\beta}\vartheta_{r\lambda}$$

$$\frac{\partial}{\partial x_\lambda}(\sqrt{-G}\gamma_{r\lambda}t_{r\beta}) + \frac{1}{2}\sqrt{-G}\frac{\partial \gamma_{r\lambda}}{\partial x_\beta}t_{r\lambda}$$

reduziert sich vermöge der Gleichungen auf

$$\sqrt{-G}\Theta_{\alpha\lambda}\frac{\partial}{\partial x_\lambda}(g_{r\beta}g_{s\alpha}\Theta_{rs}) - \frac{1}{2}\sqrt{-G}\frac{\partial g_{r\lambda}}{\partial x_\beta}\Theta_{rs}g_{s\alpha}\Theta_{\alpha\lambda}$$

$$\sqrt{-G}\frac{\partial}{\partial x_\lambda}(\gamma_{\underline{r}\lambda}T_{\underline{rs}}\gamma_{\underline{s}\alpha}T_{\alpha\beta}) + \frac{1}{2}\sqrt{-G}\frac{\partial \gamma_{r\lambda}}{\partial x_\beta}T_{rs}\gamma_{s\alpha}T_{\alpha\lambda}$$

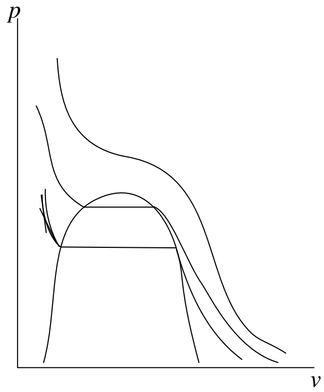
$$T_{\mu\nu} = \sqrt{-G}\varepsilon_{\mu\nu\alpha\beta}\Theta_{\alpha\beta}$$

[p. 29 R]

Wirkungsgrade

	1	5	15	25	35
	46				
50	48	21,0	28	4, 3	0, 8 – 1
	50				
75	52	1, 5	3, 6	6, 1	4, 8
	54				
100	56	0, 6	2, 5	2, 4	4, 7

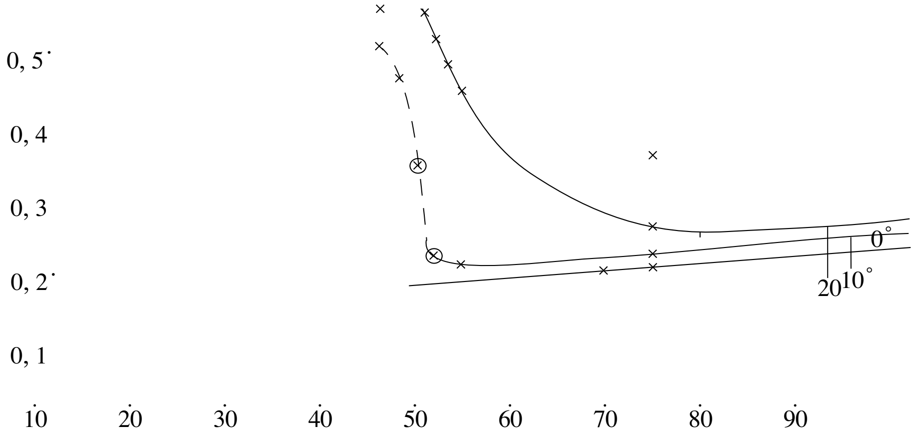
$\frac{25}{30}\cdot\frac{278}{10}$	$\frac{20\cdot 9}{21}\cdot\frac{288}{10}$	$\frac{11}{67}\cdot\frac{298}{10}$	$\frac{083}{82}$ 308
$\frac{023}{25}\cdot\frac{278}{10}$	$\frac{04\cdot 6}{29}\ \frac{288}{10}$	$\frac{12}{50}$	$\frac{124}{64}$
$\frac{02}{32}\cdot\frac{278}{10}$	$\frac{\cancel{0}03}{34\cdot 5}\ \frac{288}{10}$	$\frac{043}{38}$	$\frac{096}{32}$



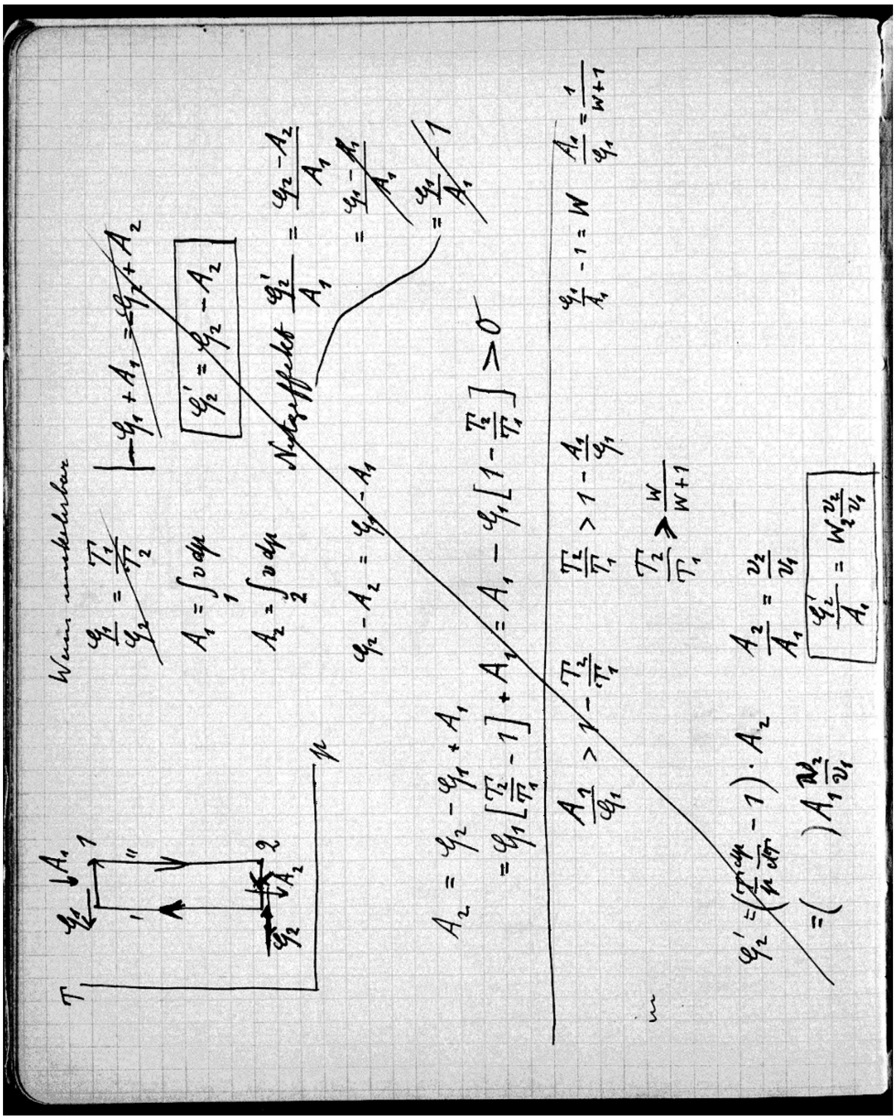
	7, 5	15	
46	25	5, 1	
56	1, 8	22	

$\frac{36}{38}\cdot 28$ $\frac{13}{62}\cdot 29$

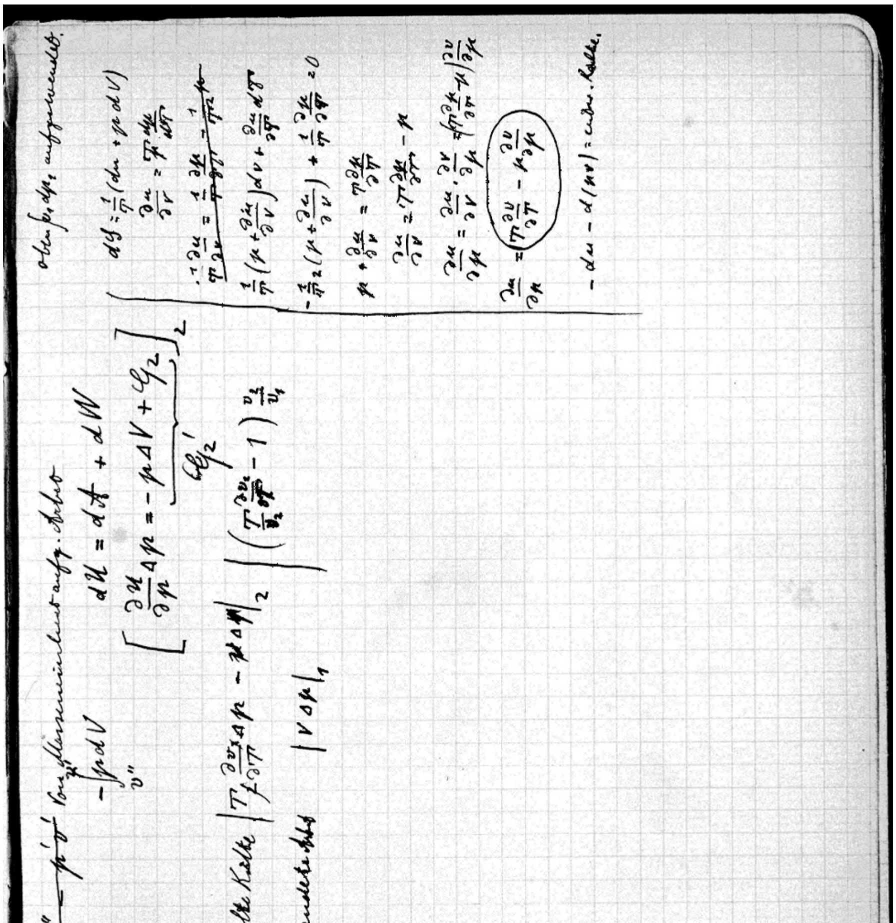
$\frac{2}{21}\cdot 28$ $\frac{30}{37}\cdot 29$



[p. 30 L]



[p. 30 R]



~~$p''v'' - p'v'$~~ von Massenelement aufg. Arbeit

$-\int_{v'}^{v''} p\,dV$ $\quad dU = dA + dW$

$\left[\frac{\partial U}{\partial p}\Delta p = \underbrace{-p\Delta V + G_2}_{G_2'}\right]_2$

Entwickelte Kälte $\left|T\frac{\partial v}{\partial T}\Delta p - \cancel{p\Delta v}\right|_2$ $\left|\left(\frac{T}{v_2}\frac{\partial v_2}{\partial T} - 1\right)\frac{v_2}{v_1}\right.$

aufgewendete Arb. $|v\Delta p|_1$

oben $p_1 dp_1$ aufgewendet.

$dS = \frac{1}{T}(du + p\,dV)$

$\frac{\partial u}{\partial v} = T\frac{dp}{dT}$

$\cancel{\frac{1}{T}\frac{\partial u}{\partial v} = \frac{1}{T}\frac{\partial p}{\partial T} - \frac{1}{T^2}p}$

$\frac{1}{T}\left(p + \frac{\partial u}{\partial v}\right)dv + \frac{\partial u}{\partial T}dT$

$-\frac{1}{T^2}\left(p + \frac{\partial u}{\partial v}\right) + \frac{1}{T}\frac{\partial p}{\partial T} = 0$

$p + \frac{\partial u}{\partial v} = T\frac{\partial p}{\partial T}$

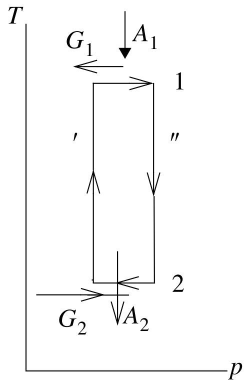
$\frac{\partial u}{\partial v} = T\frac{\partial p}{\partial T} - p$

$\frac{\partial u}{\partial p} = \frac{\partial u}{\partial v}\cdot\frac{\partial v}{\partial p} = \left(T\frac{\partial p}{\partial T} - p\right)\frac{\partial v}{\partial p}$

$\frac{\partial u}{\partial p} = T\frac{\partial v}{\partial T} - p\frac{\partial v}{\partial p}$

$-du - d(pv) =$ entw. Kälte.

[p. 30 L]



Wenn umkehrbar

$$\cancel{\frac{G_1}{G_2} = \frac{T_1}{T_2}} \qquad \cancel{-G_1 + A_1 = -G_2 + A_2}$$

$$A_1 = \int_1 v\,dp \qquad \boxed{G_2' = G_2 - A_2}$$

$$A_2 = \int_2 v\,dp$$

Nutzeffekt $\frac{G_2'}{A_1} = \frac{G_2 - A_2}{A_1}$

$$= \cancel{\frac{G_1 - A_1}{A_1}}$$

$$G_2 - A_2 = G_1 - A_1$$

$$= \cancel{\frac{G_1}{A_1}} - 1$$

$$A_2 = G_2 - G_1 + A_1$$

$$= G_1\left[\frac{T_2}{T_1} - 1\right] + A_1 = A_1 - G_1\left[1 - \frac{T_2}{T_1}\right] > 0$$

$$\frac{A_1}{G_1} > 1 - \frac{T_2}{T_1} \qquad \frac{T_2}{T_1} > 1 - \frac{A_1}{G_1} \qquad \frac{G_1}{A_1} - 1 = W \qquad \frac{A_1}{G_1} = \frac{1}{W+1}$$

$$\frac{T_2}{T_1} > \frac{W}{W+1}$$

$$G_2' = \left(\frac{T}{p}\frac{dp}{dT} - 1\right) \cdot A_2 \qquad \frac{A_2}{A_1} = \frac{v_2}{v_1}$$

$$= \left(\qquad\right) A_1 \frac{v_2}{v_1} \qquad \boxed{\frac{G_2'}{A_1} = W_2 \frac{v_2}{v_1}}$$

[p. 30 R]

$\cancel{p''v'' - p'v'}$ Von Masseneinheit aufg. Arbeit

$$-\int_{v''}^{v'} p dV \qquad dU = dA + dW$$

$$\left[\frac{\partial U}{\partial p}\Delta p = \underbrace{-p\Delta V + G_2}_{G_2'}\right]_2$$

Entwickelte Kälte $\left| T_{\langle 2\rangle} \frac{\partial v_{\langle 2\rangle}}{\partial T}\Delta p - v\Delta p \right|_2 \quad \left(\frac{T}{v_2}\frac{\partial v_2}{T} - 1\right)\frac{v_2}{v_1}$

aufgewendete Arbeit $|v\Delta p|_1$

oben $\int v_1 dp_2$ aufgewendet.

$$dS = \frac{1}{T}(du + pdV)$$

$$\frac{\partial u}{\partial v} = T\frac{dp}{dT}$$

$$\cancel{\cdot \frac{1}{T}\frac{\partial u}{\partial v} = \frac{1}{T}\frac{\partial p}{\partial T} - \frac{1}{T^2}p}$$

$$\frac{1}{T}\left(p + \frac{\partial u}{\partial v}\right)dv + \frac{\partial u}{\partial T}dT$$

$$-\frac{1}{T^2}\left(p + \frac{\partial u}{\partial v}\right) + \frac{1}{T}\frac{\partial p}{\partial T} = 0$$

$$p + \frac{\partial u}{\partial v} = T\frac{\partial p}{\partial T}$$

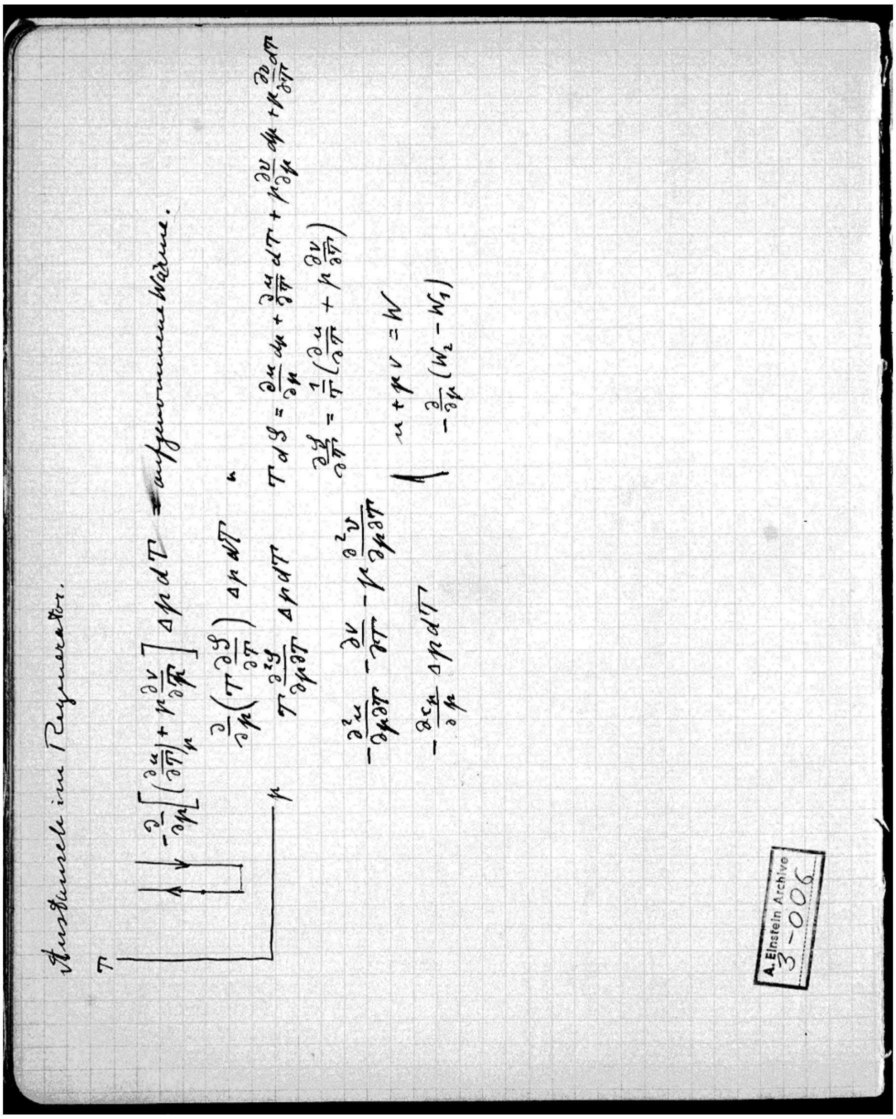
$$\frac{\partial u}{\partial v} = T\frac{\partial p}{\partial T} - p$$

$$\frac{\partial u}{\partial p} = \frac{\partial u}{\partial v}\cdot\frac{\partial v}{\partial p} = \left(T\frac{\partial p}{\partial T} - p\right)\frac{\partial v}{\partial p}$$

$$\frac{\partial u}{\partial p} = T\frac{\partial v}{\partial T} - p\frac{\partial v}{\partial p}$$

$-du - d(pv) =$ entw. Kälte.

[p. 31 L]



Austausch im Regenerator.

$$-\frac{\partial}{\partial p}\left[\left(\frac{\partial u}{\partial T}\right)_p + p\frac{\partial v}{\partial T}\right]\Delta p\, dT = \text{aufgenommene Wärme.}$$

$$\frac{\partial}{\partial p}\left(T\frac{\partial S}{\partial T}\right)\Delta p\, dT \quad "$$

$$T\frac{\partial^2 S}{\partial p\,\partial T}\Delta p\, dT \qquad T\,dS = \frac{\partial u}{\partial p}dp + \frac{\partial u}{\partial T}dT + p\frac{\partial v}{\partial p}dp + p\frac{\partial v}{\partial T}dT$$

$$\frac{\partial S}{\partial T} = \frac{1}{T}\left(\frac{\partial u}{\partial T} + p\frac{\partial v}{\partial T}\right)$$

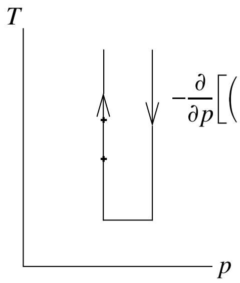
$$-\frac{\partial^2 u}{\partial p\,\partial T} - \frac{\partial v}{\partial T} - p\frac{\partial^2 v}{\partial p\,\partial T}$$

$$-\frac{\partial c_p}{\partial p}\Delta p\, dT \qquad u + pv = W$$

$$-\frac{\partial}{\partial p}(W_2 - W_1)$$

[p. 31 L]

Austausch im Regenerator.



$$-\frac{\partial}{\partial p}\left[\left(\frac{\partial u}{\partial T}\right)_p + p\frac{\partial v}{\partial T}\right]\Delta p dT = \text{aufgenommene Wärme.}$$

$$\frac{\partial}{\partial p}\left(T\frac{\partial S}{\partial T}\right)\Delta p dT \qquad \text{"}$$

$$T\frac{\partial^2 S}{\partial p \partial T}\Delta p dT \qquad TdS = \frac{\partial u}{\partial p}dp + \frac{\partial u}{\partial T}dT + p\frac{\partial v}{\partial p}\partial p + p\frac{\partial v}{\partial T}dT$$

$$\frac{\partial S}{\partial T} = \frac{1}{T}\left(\frac{\partial u}{\partial T} + p\frac{\partial v}{\partial T}\right)$$

$$-\frac{\partial^2 u}{\partial p \partial T} - \frac{\partial v}{\partial T} - p\frac{\partial^2 v}{\partial p \partial T}$$

$$u + pv = W$$

$$-\frac{\partial c_p}{\partial p}\Delta p dT \qquad -\frac{\partial}{\partial p}(W_2 - W_1)$$

MICHEL JANSSEN, JÜRGEN RENN, TILMAN SAUER,
JOHN D. NORTON, AND JOHN STACHEL

A COMMENTARY ON THE NOTES ON GRAVITY IN THE ZURICH NOTEBOOK

TABLE OF CONTENTS

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 2
Einstein's Zurich Notebook: Commentary and Essays.

1. INTRODUCTION

The Zurich Notebook provides what appears to be a virtually complete record of Einstein's search for gravitational field equations in the winter of 1912–1913. He had just started to explore a new theory of gravitation in which the ten components of the metric tensor take over the role of the gravitational potential in Newton's theory.[1] The notes documenting Einstein's search for field equations for this theory take up the better part of the notebook. They start on pp. 39L–41R and continue on pp. 5R–29L and pp. 42L–43L. Our text is a detailed running commentary on these notes.[2] It provides line-by-line reconstructions of all calculations and discusses the purpose behind them.[3] The commentary is self-contained and can in principle be read independently of the notebook. It is designed, however, to be used in conjunction with the facsimile

1 For discussion of how Einstein arrived at this theory, see "The First Two Acts" (in vol. 1 of this series). In terms of this metaphor, the search for field equations comprises the third act. It starts with the Zurich Notebook and ends with the four communications to Berlin Academy of November 1915 in which Einstein completed general relativity as we know it today with the formulation of the generally-covariant Einstein field equations (Einstein 1915a, b, c, d).

2 The commentary does not cover pp. 27L–29L, which deal with the behavior of matter in a given metric field.

3 We follow the notation used in the notebook and clearly indicate whenever we use elements of modern notation (such as the Kronecker delta or a more compact notation for derivatives) in our reconstructions of Einstein's calculations. Einstein's notation should be easy to follow for readers familiar with the basics of the standard modern notation of general relativity. The one exception is Einstein's idiosyncratic convention before (Einstein 1914b) of writing all indices downstairs and distinguishing between covariant and contravariant components (e.g., the components $g_{\mu\nu}$ and $g^{\mu\nu}$ of the metric) by using Latin letters for one ($g_{\mu\nu}$) and Greek letters for the other ($\gamma_{\mu\nu}$). In the notebook, to make matters worse, Einstein sometimes used Latin for contravariant and Greek for covariant components. The contravariant stress-energy tensor, e.g., is denoted by $T_{\mu\nu}$ in the notebook while its covariant counterpart is denoted by $\Theta_{\mu\nu}$ (see, e.g., p. 5R and equation (64)), which is just the reverse of the notation used in (Einstein and Grossmann 1913) and subsequent publications. The summation convention (repeated indices are summed over) was only introduced in (Einstein 1916a, 788). In the notebook, however, Einstein occasionally omitted summation signs, thus implicitly using it. When providing intermediate steps for calculations in the notebook we shall frequently and tacitly use the summation convention.

and the transcription of the notebook presented in vol. 1 of this series. The reader seeking guidance in reading a particular passage of the notebook can go directly to the section of the commentary in which that passage is analyzed. To help the reader find the relevant section, both the table of contents and the running heads of the text match (sub) (sub-)sections of the commentary to pages of the notebook. When looking up the annotation for a specific passage of the notebook, the reader is advised to consult the introduction to the (sub-)(sub-)section of the commentary dealing with that passage first.

After covering the earliest research on gravity documented by the notebook (pp. 39L–41R) in sec. 2 and the derivation of the law for the energy-momentum balance between matter and gravitational field (p. 5R) in sec. 3, the commentary continues with its two longest sections, sec. 4 covering pp. 6L–13R and 41L–R and sec. 5 covering pp. 14L–25R and 42L–43L. These two sections have extensive introductions, providing an overview of the material cross-referenced both with the page numbers of the notebook and the numbers of the (sub-)subsections of the commentary where the material is covered in detail. Sec. 4 covers a stage in Einstein's research during which, through trial and error, he found a body of results, strategies, and techniques that he drew on for the more systematic search for field equations during the next stage. Sec. 5 follows the fate of a series of candidate field equations as Einstein checked them against a list of criteria they would have to satisfy. All these candidates are extracted from the Riemann tensor, which makes its first appearance on p. 14L, the first of the pages covered by sec. 5.

In the course of his investigations, Einstein already came across some of the field equations published in November 1915. One of the central tasks of the commentary is to analyze why these candidate field equations were rejected three years earlier. In the notebook Einstein eventually gave up trying to extract field equations from the Riemann tensor. Drawing on results and techniques found during the earlier stage instead, he developed a way of generating field equations guaranteed to meet what he deemed to be the most important of the requirements to be satisfied by such equations. In this way he found the field equations of severely limited covariance published in the spring of 1913 in a paper co-authored with Marcel Grossmann (Einstein and Grossmann 1913). The theory and the field equations presented in this paper are known, after the title of the paper, as the "*Entwurf*" ("outline") theory and the "*Entwurf*" (field) equations. The derivation of the *Entwurf* field equations on pp. 26L–R of the notebook is covered in sec. 6, the short section that concludes the commentary.

To get an overview of the contents of the gravitational part of the Zurich Notebook, the reader is advised to read the balance of this section as well as secs. 4.1 and 5.1–5.2, the introductions to the two main sections of the commentary.

1.1 *The Four Heuristic Requirements*

For our analysis of the notebook it is important to distinguish two strategies employed by Einstein in his search for suitable gravitational field equations and four requirements serving as guideposts and touchstones in this search. In sec. 1.2 we introduce the two strategies; in this section we introduce the four heuristic requirements, which, for ease of reference, we have labeled relativity, equivalence, correspondence, and conservation.[4]

1.1.1 *The Relativity Principle*

Throughout the period in which Einstein formulated general relativity as we know it today (i.e., the years 1912–1915), he was under the impression that the principle of relativity for uniform motion of special relativity can be generalized to arbitrary motion by extending the manifest Lorentz invariance of special relativity in the formulation of Minkowski (1908), Sommerfeld (1910a, b), Laue (1911b) and others to general covariance.[5] Only in the fall of 1916 did Einstein come to realize that general covariance does not automatically lead to relativity of arbitrary motion.[6] Kretschmann (1917) was the first to give a precise formulation of the difference between general covariance and general relativity.[7] Given his conflation of the two at the time, Einstein tried to implement the relativity principle in the notebook by constructing field equations of the broadest possible covariance. He was thus drawn to generally-covariant objects such as the Beltrami invariants (p. 6L) and the Riemann tensor (p. 14L). He found these in the mathematical literature with the help of Marcel Grossmann. Einstein would sacrifice some covariance to meet the other requirements the field equations had to satisfy. Contrary to what we know today, he assumed that both energy-momentum conservation and the recovery of Newtonian theory for weak static fields put constraints on the class of admissible coordinate transformations. Initially, his hope was that this class would still include transformations to frames of reference in arbitrary states of motion. In the course of the research documented in the notebook it became clear to him that this is not the case. By the end of the notebook (pp. 26–R), Einstein had settled for invariance under general linear transformations, hoping to extend the principle of relativity to non-uniform motion in a different way, suggested by the equivalence principle.

1.1.2 *The Equivalence Principle*

Einstein's fundamental insight in developing general relativity was that there is an intimate connection between acceleration and gravity. As Einstein put it in 1918, the two are "of the exact same nature" ("wesensgleich," Einstein 1918, 176). They are two

---

4 For further discussion, see "Pathways out of Classical Physics …" (in vol. 1 of this series) and (Renn and Sauer 1999).

5 For an analysis of Einstein's conflation of the status of Lorentz invariance in special relativity and the status of general covariance in general relativity, see (Norton 1992, 1999).

6 This became clear to Einstein in discussions with Willem de Sitter. For further discussion, see the headnote, "The Einstein-De Sitter-Weyl-Klein Debate," in (CPAE 8, 351–357).

7 For discussion of Kretschmann's paper and Einstein's brief response to it in (Einstein 1918), see, e.g., (Norton 1993, sec. 5).

sides of the same coin, which explains Galileo's principle that all bodies fall alike in a given gravitational field, or, in Newtonian terms, that inertial mass is equal to gravitational mass. Einstein first explored the connection between acceleration and gravity in (Einstein 1907) and started calling it the equivalence principle in (Einstein 1912a). Einstein wanted to use this principle to extend the relativity of uniform motion to non-uniform motion and develop a new theory of gravity at the same time.[8]

The general-relativity principle that Einstein formulated on the basis of the equivalence principle is of a somewhat peculiar nature. It is best illustrated with a couple of paradigmatic examples. Two observers in the vicinity of some massive body, one in free fall, one resisting the pull of gravity, can both claim to be at rest as long as they agree to disagree about whether or not there is a gravitational field present in their region of spacetime. The observer in free fall can legitimately claim that there is *no* gravitational field, that he is at rest, and that the other observer is accelerating upward. The observer resisting the pull of gravity is equally justified in claiming that there *is* a gravitational field, that she is at rest, and that the other observer is accelerating downward. They can make similar claims when they are both in some flat region of spacetime, one hovering freely, the other firing up the engines of her spacecraft. The first observer can claim that there is no gravitational field, that he is at rest, and that the other observer is accelerating upwards. The second observer can claim that a gravitational field came into existence the moment she turned on her engines, that she is at rest in this field, and that the other observer is accelerating downward in it. In neither of these two cases—close to some massive body or in some flat region of spacetime—are the situations of these two observers physically equivalent to one another. The equivalence, in fact, is between the observer in free fall in the first case and the one hovering in outer space in the second and between the observer resisting the pull of gravity in the first case and the one in the accelerating rocket in the second. In both cases it is not the motion of the two observers with respect to one another that is relative—in the sense of being determined only with reference to the observer making the call—but the presence or absence of a gravitational field.[9]

Einstein implemented the equivalence principle by letting the metric field $g_{\mu\nu}$ represent both the gravitational field and the inertial structure of spacetime. For the equivalence principle to go through it is crucial that in all cases such as the ones considered above the metric field is a solution of the same field equations in the coordinate systems of both observers. This is automatically true if the field equations are generally covariant. But general covariance, while sufficient, is not necessary to meet this requirement. Whenever general covariance proved unattainable in the notebook (and similarly during the subsequent reign of the *Entwurf* theory with its field equations of severely restricted covariance), Einstein tried to meet the requirement in a different way, involving what he first called "non-autonomous" ("unselbständige") transformations[10] and later "justified" ("berechtigte") transformations between "adapted"

8 For discussion of Einstein's equivalence principle, see (Norton 1985).
9 For further discussion, see (Janssen 2005, 63–66).

("angepaßte") coordinates.[11] In the case of ordinary, autonomous, transformations, the new coordinates are simply functions of the old ones. If the field equations are invariant under some autonomous transformation, any solution in the old coordinates is guaranteed to turn into a solution in the new ones under that transformation. In the case of non-autonomous transformations, the new coordinates are functions of the old coordinates and the metric field (expressed in terms of the old coordinates).[12] If the field equations are invariant under a non-autonomous transformation for some metric field that is a solution of the field equations in the old coordinates, only that particular solution is guaranteed to turn into a solution in the new coordinates. This suffices for the implementation of the equivalence principle. Given the difficulty of finding field equations invariant under a broad enough class of autonomous transformations, it need not surprise us that non-autonomous transformations play a prominent role in the notebook. Einstein was especially interested in non-autonomous transformations to uniformly rotating and uniformly accelerating frames of reference in flat spacetime, in which case the metric in the old coordinates is simply the standard diagonal Minkowski metric.

1.1.3 *The Correspondence Principle*

An obvious constraint on Einstein's new gravitational theory was that Newton's theory be recovered under the appropriate circumstances. We call this the correspondence principle. In the case of weak static fields, the 44-component of the metric (where x_4 is the time coordinate multiplied by the velocity of light) is proportional to the gravitational potential in Newtonian theory. The correspondence principle thus requires that, in the case of weak static fields, the component of the field equations that determines g_{44} reduce to the Poisson equation for the Newtonian potential. Einstein expected g_{44} to be the only variable component of the metric in this case so that the spatial part of the metric would remain flat. This would allow him to connect his new theory both to Newtonian theory and to his own earlier theory for static fields in which the gravitational potential is represented by a variable speed of light (Einstein 1912a, b).

It turns out that the correspondence principle does not require the metric for weak static fields to be spatially flat. Other components of the metric besides g_{44} can be variable without losing compatibility with Newtonian theory. This is because, regardless of the values of the other components, g_{44} is the only component to enter into the equations of motion for matter moving in the gravitational field in the relevant weak-field slow-motion approximation. Einstein only came to realize this in the course of his calculation of the perihelion motion of Mercury in November 1915 (Einstein 1915c).[13] Throughout the notebook he assumed that only g_{44} can be variable for weak

10 Einstein to H. A. Lorentz, 14 August 1913 (CPAE 5, Doc. 467). The relevant passage is quoted in footnote 94.

11 (Einstein and Grossmann 1914, 221; Einstein 1914b, 1070)

12 For more detailed discussion of non-autonomous transformations, see the introduction to sec. 4.3.

static fields. At one point he contemplated relaxing this requirement but quickly convinced himself that this is not allowed.[14]

Despite this additional constraint on the form of the metric for weak static fields, Einstein found various field equations that satisfy the correspondence principle. On the face of it, it looks as if Einstein applied coordinate conditions to various field equations of broad covariance to establish that they reduce to the Poisson equation in the appropriate limit.[15] On closer examination, it turns out that Einstein actually used what we shall call coordinate *restrictions*. He took expressions of broad covariance and truncated them by imposing additional conditions on the metric to obtain candidate field equations that reduce to the Poisson equation in the case of weak static fields. Einstein did not see such truncated equations as representing candidate field equations of broader covariance in a limited class of coordinate systems for the purpose of comparing them with Newtonian theory, as we would nowadays, but as candidates for the fundamental field equations of the theory. The status of the conditions on the metric with which Einstein did the truncating is therefore very different from that of modern coordinate conditions. This is why we introduced the special term coordinate restrictions.[16]

Coordinate restrictions played an important role in Einstein's attempts to find non-autonomous coordinate transformations under which candidate field equations would be invariant. If such candidates had been extracted from expressions of broad covariance with the help of some coordinate restriction, their covariance was determined by the covariance of that coordinate restriction. The expressions involved in coordinate restrictions are much simpler than the field equations themselves and the covariance properties of the former are therefore much more tractable than those of the latter. Many pages of the notebook are thus given over to the investigation of the covariance properties of various coordinate restrictions.[17] Einstein routinely checked whether the coordinate restrictions he imposed allow non-autonomous transformations to uniformly rotating and uniformly accelerating frames of reference in Minkowski spacetime.

The correspondence principle led Einstein to expect the left-hand side of the field equations to have the form of a sum of what we shall call a core operator, a term with second-order derivatives of the metric that for weak static fields reduces to the Laplacian acting on the metric, and various correction terms quadratic in first-order derivatives of the metric that vanish in a weak-field approximation. He either extracted

---

13 See (Norton 1984, 146–147), (Earman and Janssen 1993, 144–145), and "Untying the Knot …" sec. 7 (in this volume).

14 See p. 21R discussed in secs. 5.4.4 and 5.4.6.

15 See in particular pp. 19L and 22R discussed in sec. 5.4.1 and sec. 5.5.2, respectively. When considered in isolation, these two pages strongly suggest that Einstein applied coordinate conditions in the modern sense in the notebook, a conclusion that was indeed drawn in (Norton 1984).

16 For a particularly illuminating example of a coordinate restriction, see pp. 23L–R discussed in sec. 5.5.4.

17 See, e.g., pp. 10L–11L (covered in secs. 4.5.1–4.5.3), p. 22L (covered in 5.5.3), and pp. 23R, 42L–R (covered in secs. 5.5.5–5.5.6).

equations of this form from equations of broad covariance with the help of coordinate restrictions or he determined what terms had to be added to the core operator on the basis of considerations of energy-momentum conservation.

1.1.4 *The Conservation Principle*

Einstein's 1912 theory for static gravitational fields had taught him to pay close attention to what we shall call the conservation principle, the compatibility between the field equations and energy-momentum conservation. The field equations proposed in (Einstein 1912a) had turned out to be incompatible with energy-momentum conservation and Einstein had been forced to add a term to them in (Einstein 1912b).[18] The extra term, it turned out, gave the energy density of the gravitational field and entered the field equations on the same footing as the energy density of the field's material sources. This, as Einstein realized, had to be the case given the equivalence of energy and inertial mass expressed in $E = mc^2$, the equality of inertial and gravitational mass asserted by the equivalence principle, and the equality of active and passive gravitational mass. The field equations of the new metric theory would thus have to satisfy a similar requirement. Einstein accordingly sought to interpret the correction terms to the core operator mentioned above as the energy-momentum density of the gravitational field, occurring on a par with the stress-energy tensor of matter.[19]

That the stress-energy tensor of matter should replace the mass of Newtonian theory as the source of the gravitational field Einstein had learned from the development of special-relativistic mechanics.[20] The early years of special relativity had brought a transition from Galilean-invariant particle mechanics based on Newton's second law to Lorentz-invariant continuum mechanics based on energy-momentum conservation, expressed by the vanishing of the four-divergence of the total stress-energy tensor of closed systems, $\partial_\nu T^{\mu\nu}_{\mathrm{tot}} = 0$ (∂_ν shorthand for $\partial / \partial x_\nu$). The conservation principle requires that the gravitational field equations be compatible with this law in a weak-field approximation in which the energy-momentum of the gravitational field itself can be neglected. The weak-field field equations have the form $\Box g^{\mu\nu} = -\kappa T^{\mu\nu}$

18 See footnote 119 for more details.

19 Following the terminology in (Einstein and Grossmann 1913, e.g., p. 11 and p. 16), we shall use the term stress-energy tensor' throughout the commentary rather than the more common 'energy-momentum tensor' or the more cumbersome 'stress-energy-momentum (SEM) tensor'. In the notebook Einstein referred both to the "tensor of momentum and energy" ("Tensor der Bewegungsgröße u. Energie," p. 5R) and to the "stress-energy tensor" [*Sp[annungs]-Energie-Tensor*] p. 20R).

20 In his 1912 manuscript on special relativity, Einstein called the extension of the central role of the stress-energy tensor in electrodynamics to all of physics "the most important recent advance in relativity theory" ("den wichtigsten neueren Fortschritt der Relativitätstheorie." CPAE 4, Doc. 1, p. [63]). He gave credit to Minkowski, Abraham, Planck, and Laue for this development. The symmetry of the stress-energy tensor in its two indices encodes such physical knowledge as the inertia of energy and the conservation of angular momentum (Laue 1911a). This means that the differential operator acting on the metric set equal to the stress-energy tensor in the field equations must have that same symmetry. This requirement is indeed imposed, though not emphasized, in the notebook (cf. footnotes 219 and 220).

($\Box$ is the d'Alembertian, $\eta^{\mu\nu}\partial_\mu\partial_\nu$, with $\eta^{\mu\nu} \equiv \mathrm{diag}(-1, -1, -1, 1)$ the standard diagonal Minkowski metric). Einstein imposed the coordinate restriction $\partial_\nu g^{\mu\nu} = 0$ in which case these field equations imply energy-momentum conservation in the weak-field approximation.[21] This coordinate restriction, which Einstein used both to satisfy the conservation principle and to satisfy the correspondence principle, occurs so frequently in the notebook that we have given it a special name. We call it the Hertz restriction.[22]

When the energy-momentum coming from the gravitational field cannot be neglected, the situation gets more complicated. On p. 5R Einstein derived an equation giving the exact energy-momentum balance between matter and gravitational field. In modern terms, this equation states that the covariant divergence of the stress-energy tensor of matter vanishes. It is the sum of two terms, the ordinary divergence of the stress-energy tensor of matter and an expression, once again containing this tensor, that can be interpreted as the gravitational force. The field equations set some second-order differential operator acting on the metric equal to the stress-energy tensor of matter. One of the most important tests to which Einstein submitted candidate field equations was to use them to eliminate the stress-energy tensor from the expression for the gravitational force in the energy-momentum balance and check whether the resulting expression can be written as the divergence of an expression that could be interpreted as the energy-momentum density of the gravitational field itself.[23] Given the form of the stress-energy tensor of the electromagnetic field and of the stress tensor of the gravitational field in his static theory, Einstein expected and required this quantity to be quadratic in first-order derivatives of the metric. That way gravitational energy-momentum could indeed be neglected in the weak-field approximation. With the help of this quantity, Einstein could now rewrite the energy-momentum balance between matter and gravitational field as an ordinary divergence of the total energy-momentum density—of matter and of the gravitational field. The balance equation thus turns into a genuine conservation law. That a covariant divergence can be rewritten as an ordinary divergence in this manner immediately makes it clear that gravitational energy-momentum density cannot be represented by a generally-covariant tensor. It is what we now call a pseudo-tensor. Throughout the notebook, however, Einstein tacitly assumed that its transformation properties are the same as those of any other stress-energy tensor. He only recognized in early 1914 that this is not true.[24] Einstein eventually turned this test of the conservation principle into a powerful method for generating field equations. It was this method that gave him the *Entwurf* field equations.[25]

---

21 See p. 19R discussed in sec. 5.4.2.

22 Named after Paul Hertz, the recipient of an important letter in which Einstein discussed this restriction (Einstein to Paul Hertz, 22 August 1915 [CPAE 8, Doc. 111]).

23 See, e.g., p. 19R.

24 See (Einstein and Grossmann 1914). For discussion, see (Norton 1984, sec. 5), "Pathways out of Classical Physics …" (in vol. 1 of this series), and "What Did Einstein Know …" sec. 2 (in this volume).

In summary, the conservation principle resulted in (at least) four related but distinct requirements that candidate field equations have to satisfy. First, the energy-momentum density of the gravitational field has to enter into the field equations in the same way as the energy-momentum density of matter. Secondly, the field equations should guarantee that the four-divergence of the stress-energy tensor of matter vanishes in the weak-field approximation. Thirdly, the field equations should allow the gravitational force density in the energy-momentum balance between matter and gravitational field to be written as the divergence of some gravitational stress-energy pseudo-tensor. Finally, this pseudo-tensor should be an expression quadratic in first-order derivatives of the metric. Given this rich harvest of requirements, the conservation principle was probably the most fruitful of the four heuristic principles that guided Einstein in his search for suitable field equations.

1.2 *The Two Strategies*

Einstein attacked the problem of finding suitable field equations for the metric field from two directions, clearing the hurdles he had himself erected with his four heuristic requirements— relativity, equivalence, correspondence, and conservation—in a different order. In what we call the 'mathematical strategy,' Einstein tackled relativity and equivalence first and then moved on to correspondence and conservation. In what we call the 'physical strategy' it is just the other way around. There he started with correspondence and conservation and then turned to relativity and equivalence.[26] The identification of these two complementary strategies not only turned out to be key to our reconstruction of many of Einstein's arguments and calculations in the notebook, it also greatly enhanced our understanding of his work on general relativity during the subsequent period of 1913–1915.[27] In this section we briefly characterize these two strategies.

The mathematical strategy was to use one of the generally-covariant quantities that can be found in the mathematical literature, such as the Beltrami invariants or the Riemann tensor, to construct a second-order differential operator acting on the metric (or its determinant) that is then set equal to the stress-energy tensor of matter (or its trace). If this can be done without compromising the general covariance of the initial quantity too much, such field equations will automatically meet the relativity and equivalence requirements. The problem that Einstein ran into was that the correspondence and conservation requirements, if they could be met at all, called for severe coordinate restrictions. Still, as explained in sec. 1.1.3, knowing that candidate field equations can be extracted from equations of broad covariance with the help of a coordinate restriction makes their covariance properties much more tractable. Their covariance is fully

---

25 See p. 13R, p. 24R, and p. 26L-R, discussed in secs. 4.6.2, 5.6.1, and 6, respectively.

26 For more detailed discussion of these two strategies, see (Renn and Sauer 1999) and "Pathways out of Classical Physics ..." (in vol. 1 of this series).

27 See "Untying the Knot ..." (in this volume).

determined by the covariance of the coordinate restriction. Unfortunately, Einstein found again and again that the coordinate restrictions he needed to satisfy the correspondence and conservation requirements ruled out the kind of transformations to accelerating frames of reference needed to meet the relativity and equivalence requirements.

The physical strategy was to model the field equations for the metric field on the Poisson equation of Newtonian gravitational theory and Maxwell's equations for the electromagnetic field.[28] As explained in secs. 1.1.3 and 1.1.4, the correspondence and conservation requirements suggest that the field equations have a core operator, which for weak static fields reduces to the Laplacian acting on the metric, and a term representing gravitational energy-momentum density on the left-hand side and the stress-energy tensor of matter on the right-hand side. The conservation principle can be used to determine the exact form of the gravitational stress-energy pseudo-tensor. The physical strategy thus amounts to constructing candidate field equations guaranteed to meet the correspondence and conservation requirements. The problem is that their construction sheds little light on their covariance properties. Only their covariance under general linear transformations is assured. It thus remains completely unclear whether they satisfy the relativity and equivalence requirements. The best Einstein could do on this score was to check whether they allowed the Minkowski metric in various accelerated frames of reference so that they would at least be invariant under some non-autonomous non-linear transformations corresponding to acceleration.

In the first half of the notes (pp. 39L–41R, 5R–13R), we see Einstein vacillate between the mathematical and the physical strategy. On p. 6L, for instance, the Beltrami invariants are introduced and used as input for the mathematical strategy. Two pages later, Einstein switched to the core operator and the physical strategy. On the following pages Einstein combined his two strategies trying in vain to connect field equations based on the core operator to the Beltrami invariants in an attempt to clarify their transformation properties. On p. 13R he made further progress along the lines of the physical strategy by introducing considerations of energy-momentum conservation. Then, on p. 14L, the Riemann tensor makes its first appearance in the notebook, and Einstein abruptly switched from the physical to the mathematical strategy. What follows is a concerted effort, taking up most of the second half of the notes (pp. 14L–23R, 42L–43L), to extract field equations from the Riemann tensor. Eventually (p. 24R–26R), Einstein went back to the physical strategy, continuing the line of reasoning begun on p. 13R. He briefly combined the two strategies again, trying to find (on p. 25L) a coordinate restriction with which to extract field equations found along the lines of the physical strategy (on p. 24R) from the Riemann tensor. He then decided to go exclusively with the physical strategy, which led him to the *Entwurf* field equations on pp. 26L–R.

---

28 For detailed analysis of Einstein's use of this analogy, see "Pathways out of Classical Physics ..." (in vol. 1 of this series).

39L–41L

2. FIRST EXPLORATION OF A METRIC THEORY OF GRAVITATION (39L–41L)

2.1 *Introduction*

Einstein's attempt to construct a new dynamical theory of gravitation starts from three basic elements: (1) the representation of the gravitational field by the ten components of the metric tensor; (2) the four-dimensional spacetime formalism of special relativity; and (3) his scalar theory of the static gravitational field. For Einstein, all three components were recent additions to his stock of knowledge.

In Prague, in the spring of 1912, he brought his attempts to formulate a theory of gravitation for the special case of a static field to a satisfactory conclusion. Only after returning from Prague to Zurich in the summer of 1912 did he recognize the relevance of Gauss' theory of surfaces to the gravitation problem. Gauss' theory represents the metrical geometry of surfaces of variable curvature by a line element, the square root of a quadratic differential form invariant under arbitrary coordinate transformations. Einstein had become familiar with the four-dimensional formalism for special relativity developed by Minkowski, Sommerfeld, and Laue, and he realized that a four-dimensional extension of Gauss' theory could provide a mathematical framework suitable for a new, dynamical theory of gravitation.

The three components, out of which he hoped to build the new theory, each posed distinct but interrelated problems. Gauss' theory for two-dimensional surfaces had to be extended to a four-dimensional space with indefinite signature. The flat Minkowski spacetime formalism had to be extended to a vector and tensor analysis valid for arbitrary coordinate systems in a non-flat spacetime. Einstein's static gravitational theory was formulated in terms of a single (three)-scalar gravitational potential. The single partial differential equation governing this potential had to be generalized to a system of partial differential equations for the ten-component tensorial gravitational potential.

On pp. 39L–41L of the notebook, Einstein explored a few simple ways of combining these three components to find the gravitational field equations for a static field in special coordinates. No clear candidates emerged from Einstein's first foray into the problem. This may have signaled to him that a higher level of mathematical sophistication was called for. Even at the elementary level of these early calculations, however, one can see an alternation between physically and mathematically motivated approaches foreshadowing the two basic strategies that we distinguished in sec. 1.2.

39L

2.2 *The Three Building-Blocks: Gauss, Minkowski, Einstein (39L)*

This page includes three groups of formulas, separated from each other by two horizontal lines. Each of these groups can be associated with one of the elements mentioned above. Einstein started by writing down the square of the four-dimensional line element

$$ds^2 = \Sigma G_{\lambda\mu} dx_\lambda dx_\mu \,. \tag{1}$$

His use of a capital letter "G" indicates that this is the earliest occurrence of the metric tensor in the notebook. After p. 40R, Einstein switched to the common lower-case g.

He then derived the transformational behavior of the metric tensor under four-dimensional coordinate transformations from the condition that the line element be invariant under such transformations. The transformations between the unprimed and primed coordinates are expressed by a matrix of coefficients and its inverse[29]

$$\begin{array}{c|c|c|c|c} & x'_1 & x'_2 & x'_3 & x'_4 \\ \hline x_1 & \alpha_{11} & \alpha_{21} & \alpha_{31} & \alpha_{41} \\ \hline x_2 & \alpha_{12} & & & \\ \hline x_3 & \alpha_{13} & & & \\ \hline x_4 & \alpha_{14} & & & \end{array} \qquad \begin{array}{ccccc} & x_1 & x_2 & x_3 & x_4 \\ x'_1 & \beta_{11} & & & \\ x'_2 & \beta_{12} & & & \\ x'_3 & \beta_{13} & & & \\ x'_4 & \beta_{14} & & & \end{array} \tag{2}$$

From the invariance of the square of the line element under such a transformation,

$$\begin{aligned} \sum\sum G_{\lambda\mu} dx_\lambda dx_\mu &= \sum\sum G'_{\rho\sigma} dx'_\rho dx'_\sigma \\ &= \sum_\rho \sum_\sigma \sum_\eta \sum_\zeta G'_{\rho\sigma} \alpha_{\rho\eta} \alpha_{\sigma\zeta} dx_\eta dx_\zeta , \end{aligned} \tag{3}$$

Einstein read off the transformation laws for the components of the metric

$$G_{\lambda\mu} = \sum_\rho \sum_\sigma G'_{\rho\sigma} \alpha_{\rho\lambda} \alpha_{\sigma\mu} \,, \tag{4}$$

and "analogously" ("analog")

$$G'_{\lambda\mu} = \sum_\rho \sum_\sigma G_{\rho\sigma} \beta_{\rho\lambda} \beta_{\sigma\mu} \,. \tag{5}$$

Next to these results, Einstein noted explicitly the equations for the coordinate transformation

$$x'_r = \sum_s \alpha_{rs} x_s \,. \tag{6}$$

So apparently Einstein was considering linear transformations at this point.[30] For the corresponding first-order partial derivative operators, he wrote

29 As is indicated by a line through the first column, the transformation is given by $x'_1 = \alpha_{11}x_1 + \alpha_{12}x_2 + \alpha_{13}x_3 + \alpha_{14}x_4$, etc. A similar matrix is found in (Laue 1911b, 57) and in Einstein's 1912 manuscript on special relativity (CPAE 4, Doc. 1, secs. 15 and 16).

30 In the transformation matrix at the top of the page he may have had non-linear transformations in mind with non-constant α's: In the top row of that matrix Einstein seems to have written dx instead of simply x for the first three entries, and it may have been only later that he restricted himself to linear transformations.

$$\frac{\partial}{\partial x_s} = \sum_r \alpha_{rs} \frac{\partial}{\partial x'_r}. \tag{7}$$

In the middle of the page, below the first horizontal line, Einstein considered a "special case for the $G_{\lambda\mu}$" ("Spezialfall für die $G_{\lambda\mu}$"), namely the case of a coordinate system in which the metric is diagonal

$$\begin{array}{cccc} G_{11} & G_{12} & G_{13} & G_{14} \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & c^2 \end{array} \cdot \tag{8}$$

The spatial metric is flat and expressed in Cartesian coordinates, and the 44-component of the metric is identified with the square of the speed of light.[31] If c^2 is a constant, this metric represents the Minkowski spacetime of special relativity in quasi-Cartesian coordinates. If c^2 is a function of the spatial coordinates, $c^2 = c^2(x_1, x_2, x_3)$, it represents the metrical generalization of Einstein's static theory.

At the bottom of the page, under the second horizontal line, Einstein turned to this theory of the static gravitational field. The gravitational field equation of his static theory,[32]

$$c\Delta c - \frac{1}{2}\mathrm{grad}^2 c = kc^2\sigma, \tag{9}$$

is a partial differential equation for the velocity of light, which serves as the single gravitational potential of this theory. In a four-dimensional metric theory, the field equations will be partial differential equations for all ten components of the metric tensor, resulting from the action of some differential operator acting on it. Einstein assumed that, in the special case of a static field and with an appropriate choice of coordinates, the metric tensor would reduce to the form given in equation (8) with $c = c(x_1, x_2, x_3)$.[33] In this case, the gravitational field equations would be expected to reduce to equation (9). The problem was to reverse this reduction and find the differential operator that enters into field equations for the metric tensor.

The first step in his attempt to find this operator was to rewrite the left-hand side of equation (9) in terms of the components of the metric tensor in equation (8), i.e., to

31 A striking feature of this expression is the signature of the metric, which is in contrast to the appearance of the metric for this special case on pp. 6R and 21R, where Einstein wrote $\mathrm{diag}(-1, -1, -1, c^2)$. One possible explanation for this is that Einstein implicitly used an imaginary time coordinate at this point, which he had explicitly introduced on p. 32R in the context of his treatment of electrodynamics in moving media.

32 (Einstein 1912b, 456, equation (3b)).

33 Einstein later made this expectation explicit (Einstein and Grossmann 1913, Sec. 2).

rewrite the equation in terms of c^2 instead of c. Einstein began with the left-hand side of equation (9)

$$c\Delta c - \frac{1}{2}\mathrm{grad}^2 c \,. \tag{10}$$

He then wrote down the first- and second-order partial derivatives of c^2 with respect to x

$$c^2 \qquad 2c\frac{\partial c}{\partial x} \qquad 2\left(\frac{\partial c}{\partial x}\right)^2 + 2c\frac{\partial^2 c}{\partial x^2} \,. \tag{11}$$

Since the y- and z-derivatives behave similarly, Einstein could now read off the Laplacian of c^2,

$$\Delta(c^2) = 2\mathrm{grad}^2 c + 2c\Delta c \,, \tag{12}$$

as well as the gradient of c^2,

$$\mathrm{grad}(c^2) = 2c\,\mathrm{grad}\,c \,, \tag{13}$$

in terms of c and its derivatives.

At this point, Einstein probably realized that various factors of 2 on the right-hand sides of equations (11)–(13) cancel if one looks at $c^2/2$ instead of c^2. In any case, he defined a function

$$\gamma = \frac{c^2}{2} \tag{14}$$

such that

$$\Delta\gamma = \mathrm{grad}^2 c + c\Delta c \,, \tag{15}$$

and

$$\mathrm{grad}\,\gamma = c\,\mathrm{grad}\,c \,. \tag{16}$$

This equation allowed Einstein to express $\mathrm{grad}^2 c$ in terms of γ:

$$\frac{\mathrm{grad}^2\gamma}{2\gamma} = \mathrm{grad}^2 c \,. \tag{17}$$

Using equations (15) and (17), he finally wrote the expression (10) in terms of γ:

$$\Delta\gamma - \frac{3}{4}\frac{\mathrm{grad}^2\gamma}{\gamma} \,. \tag{18}$$

Einstein presumably expected that this expression could be recovered from the tensorial field equations. He may have hoped that it would emerge as the 44-component of the left-hand side of these field equation, but he could, of course, not be certain that tensorial equations would yield the familiar static theory in this way.[34]

34 Alternatives are conceivable but less likely. For example, expression (18) might be the trace of the left-hand side of the unknown gravitational field equations.

39R

2.3 *Finding a Metric Formulation of the Static Field Equation (39R)*

At the top of p. 39R, Einstein wrote down the vacuum field equations of his 1912 static theory in terms of the variable γ introduced on the facing page:

$$\gamma\Delta\gamma - \frac{3}{4}\mathrm{grad}^2\gamma = 0 . \tag{19}$$

Next to this equation, he started to write the word "Umfo[rmen]" (to re-express), then deleted it in favour of "Transformieren" (to transform). Presumably, the point of such transformations was to generalize this field equation for one component of the static field in a special coordinate system to field equations for all components of the field—static or non-static—in more general coordinates.

However, instead of dealing with the field equations, Einstein turned to a simpler but related problem. In the Cartesian coordinates used so far, the static character of the spatially flat metric in equation (8) can be expressed as[35]

$$\frac{\partial G_{44}}{\partial x_4} = 0 , \tag{20}$$

the vanishing of the time derivative of the 44-component. Einstein now tried to find a covariant formulation of the condition that this metric be static. To this end, he began to transform condition (20) to primed coordinates. Using equation (7), he first expressed the derivative with respect to x_4 in terms of primed coordinates:

$$\frac{\partial}{\partial x_4} = \alpha_{14}\frac{\partial}{\partial x'_1} + \alpha_{24}\frac{\partial}{\partial x'_2} + \cdot + \cdot \tag{21}$$

Using equation (4), he then did the same for the 44-component of the metric

$$G_{44} = \sum_{\rho}\sum_{\sigma} G'_{\rho\sigma}\alpha_{\rho 4}\alpha_{\sigma 4} . \tag{22}$$

He now used the inverse transformation for $G'_{\rho\sigma}$ on the right-hand side:

$$\begin{aligned} G'_{\lambda\mu} &= k + \sum\sum G_{\rho\sigma}\beta_{\rho\lambda}\beta_{\sigma\mu} \\ &= k + G_{44} \underbrace{\sum\sum \beta_{4\lambda}\beta_{4\mu}}_{B_{\lambda\mu}} . \end{aligned} \tag{23}$$

The inclusion of the term k in the first line appears to be an error, perhaps anticipating the need for such a term on the next line, which absorbs all terms in the summation except those containing G_{44}.[36] If k is a constant, as the notation suggests, then Einstein seems to be considering the special case of linear transformations (in which case only G_{44} is non-constant). It is evident from equation (23) that, under arbitrary linear transformations, all components of $G'_{\lambda\mu}$ can be non-constant. This expression, how-

35 Originally, this equation was written as $\partial G_{44}/\partial t = 0$.

36 In the second line, Einstein presumably forgot to take out the summation signs.

ever, does not indicate how the $G'_{\lambda\mu}$ depend on the primed coordinates, which would require expressing G_{44} as a function of the primed coordinates. Perhaps looking for such an expression, Einstein began rewriting an expression for G_{44} but quickly gave up, possibly because such a substitution would just result in the composition of a transformation and its inverse. Finally, he deleted the entire calculation and started afresh in the top right corner of the page. (This part is ruled off by a horizontal and a vertical line.)

So Einstein had failed to find a covariant reformulation of the physically-motivated condition (20) through direct transformation. He now turned his attention to the mathematically more promising relation

$$\mathrm{Div}\,\Gamma = 0\,, \tag{24}$$

which reduces to equation (20) in the case of the spatially flat static metric (8). That Einstein did indeed find this equation promising is indicated by the word "probable" ("wahrscheinlich") that he wrote above it. The differential operator "Div" is analogous to the quasi-Cartesian coordinate divergence of a tensor, an operation well known in Minkowski's four-dimensional spacetime formalism.[37] Einstein was familiar with the use of this operation in four-dimensional electrodynamics. The symbol "Γ" denotes the metric tensor ("Tensor der G").

Einstein now asked "Is this invariant?" ("Ist dies invariant?") and wrote out equation (24) explicitly:

$$\sum_{\mu}\frac{\partial G_{\lambda\mu}}{\partial x_{\mu}} = 0,\ \lambda = 1234\,. \tag{25}$$

Using the transformation equations (4) and (7), he transformed this equation to primed coordinates:

$$\sum_{\tau}\alpha_{\tau\mu}\frac{\partial}{\partial x'_{\tau}}\left\{\sum_{\rho}\sum_{\sigma}G'_{\rho\sigma}\alpha_{\rho\lambda}\alpha_{\sigma\mu}\right\} = 0\,. \tag{26}$$

Equation (26) shows that the condition does indeed transform as a vector (i.e., is "invariant") under linear transformations (i.e., for constant α's). This concludes the investigation of the condition that the metric be static.

At this point, Einstein presumably returned to the field equations. This is suggested by his derivation on the bottom half of p 39R of the transformation equations for second-order derivatives. To simplify matters, he suppressed two spatial coordinates: "Everything only dependent on x_1 and x_2 (time)." ("Alles nur von x_1 und x_2 (Zeit) abhängig"). From the actual calculations at the bottom of p. 39R, it is clear that Ein-

37 The notation "Div" for a four-dimensional generalization of the ordinary three-dimensional divergence was introduced by (Sommerfeld 1910b, 650). It is also used in (Laue 1911b, 70). The four-divergence is implicitly used in the derivation of Maxwell's equations on p. 33L. Note that the ordinary divergence is written with a lower case "*d*" on p. 5R.

stein focused on linear transformations with symmetric transformation matrices (i.e., $a_{12} = a_{21}$).[38] The end result of his calculation is

$$\left(\frac{\partial^2}{\partial x_1^2}\right) = \alpha_{11}^2 \frac{\partial^2}{\partial x'^2_1} + 2\alpha_{11}\alpha_{21}\frac{\partial^2}{\partial x'_1 \partial x'_2} + \alpha_{22}^2 \frac{\partial^2}{\partial x'^2_2} \tag{27}$$

(the last term on the right does not appear in the notebook). This equation shows that, under (symmetric) linear transformations, the set of second-order partial derivatives of a scalar function transforms like the components of the metric tensor, an insight that Einstein put to good use on the next page.

40L–R 2.4 *Searching for a Generalization of the Laplacian in the Metric Formalism (40L–R)*

At the bottom of p. 39R, Einstein had found that second-order partial derivatives of a scalar function have the same transformation behavior under linear transformations as the metric tensor. This insight probably prompted the calculations on pp. 40L–R, which are an attempt to find candidates for the differential operator acting on the metric tensor in the field equations on the basis of expressions involving well-known differential operators acting on a scalar field. Given some striking similarities between expressions found on p. 40L and a list of differential invariants in (Wright 1908), Einstein (or Grossmann) probably consulted this book while working on the calculations on this page.[39] Since the field equations have to be of second order and the components of the metric tensor are equivalent—as far as their transformation properties are concerned[40]—to second-order derivatives of a scalar, Einstein had to consider expressions containing fourth-order derivatives of the scalar function. In the course of his calculations, he further imposed the condition that all four coordinates enter on the same footing into the differential operator acting on the scalar function. The attempt was abandoned after the first few lines on p. 40R.

40L After a few false starts, Einstein listed a number of expressions with the Laplacian and the gradient operator acting on a scalar function φ. First, he wrote down three expressions with “second-order” (“2. Ordnung”) derivatives:

$$\Delta\varphi,\ \ \varphi\Delta\varphi,\ \ \mathrm{grad}^2\varphi\,. \tag{28}$$

---

38 In the 1+1 dimensional spacetime considered by Einstein, all Lorentz transformations are boosts. These are all represented by symmetric matrices. Presumably, this is the physical rationale behind Einstein’s focus on such matrices.

39 Einstein was familiar with this book, as can be inferred from Einstein to Felix Klein, 21 April 1917: “Grossmann (I believe) had the little book by Weight [sic] when we were working together on relativity four years ago” (“Das Büchlei[n] von Weight [sic] hatte Grossmann (glaube ich), als wir vor 4 Jahren zusammen über Relativität arbeiteten.” CPAE 8, Doc. 328). For historical discussion of (Wright 1908), see (Reich 1994, 105–107).

40 The equivalence can only refer to the behavior under coordinate transformations since a direct identification of the metric tensor with the second partial derivatives of a scalar function would result in severe, unacceptable restrictions on the metric.

Then he wrote down a number of expressions with "fourth-order" ("4. Ordnung") derivatives

$$\Delta\Delta\varphi,\ \Delta(\varphi\Delta\varphi),\ \Delta(\mathrm{grad}^2\varphi),\ \varphi(\Delta\Delta\varphi),\ \mathrm{grad}^2\Delta\varphi\,. \tag{29}$$

The list shows some resemblance to a list on pp. 56–57 of (Wright 1908).[41] On the next line, the first of these expressions, $\Delta\Delta\varphi$, is expanded, with the help of the definition of the Laplacian in Cartesian coordinates in two dimensions, $\Delta \equiv \partial^2/\partial x^2 + \partial^2/\partial y^2$. He then expanded the second term of the list, $\Delta(\varphi\Delta\varphi)$, as[42]

$$\underline{\Delta\varphi\cdot\Delta\varphi} + \varphi\Delta\Delta\varphi + \underbrace{\left(\frac{\partial\varphi}{\partial x}\frac{\partial}{\partial x}\Delta\varphi + \cdot + \cdot\right)}_{\mathrm{grad}\varphi\,\mathrm{grad}\Delta\varphi}. \tag{30}$$

The terms $\Delta\Delta\varphi$ and $\Delta\varphi\Delta\varphi$ only contain second-order derivatives of the scalar function φ and can therefore be translated into expressions with (differential operators acting on) components of the metric. The underlining of these two terms in the notebook thus confirms our interpretation of the rationale for these calculations.

Einstein then wrote "the first two steps" ("Die ersten 2 Schritte"), drew a horizontal line, wrote "2 dimensions" ("2 Dimensionen"), and then crossed out the calculations he had made so far and tried again.

Einstein's next comment makes his strategy explicit: "system of the G equivalent to system $\partial^2\varphi/\partial x_\mu\partial x_\nu$" ("System der G äquivalent dem System $\partial^2\varphi/\partial x_\mu\partial x_\nu$"). In other words, he now spelled out the equivalence of second-order derivatives of a scalar function and components of the metric tensor as far as their behavior under linear coordinate transformations is concerned. Einstein's next step was to impose an additional constraint on candidate field equations constructed from the kind of expressions he had been examining above: "The equation should be such that in every term one has the same number of differentiations with respect to every x" ("Gleichung soll so sein, dass in jedem Glied nach allen x gleich oft diff[erenziert] wird.").

Einstein now had three constraints that any acceptable expression had to satisfy. First, for every factor φ there should be two partial-derivative operators $\partial/\partial x_\mu$ to enable the translation to components of the metric tensor. Second, there should be two additional partial-derivative operators to yield second-order field equations. Third, for each of the four coordinates the expression should contain an equal number of operators $\partial/\partial x_\mu$.

He first considered an expression linear in φ:

41 In Wright's book the list is formed from combinations of the two Beltrami invariants (Wright 1908, 56–57). At this point in the notebook, the operator Δ does not seem to be related to the Beltrami invariants. Einstein nevertheless may have hoped that Wright's book would give him some guidance in finding a differential invariant that he could use to construct the left-hand side of the gravitational field equations. The Beltrami invariants explicitly appear on later pages of the notebook (for the first time on p. 7L; for discussion see sec. 4.2).

42 There is a factor of 2 missing in the third term.

$$\frac{\partial^8\varphi}{\partial x_1^2\partial x_2^2\partial x_3^2\partial x_4^2}. \tag{31}$$

Einstein rejected this possibility, writing: "linear impossible, of 8th order in φ" ("Linear unmöglich von 8. Ordnung in φ"). The problem with this expression is that it leads to sixth-order field equations. For instance, using the derivatives with respect to x_4 for the translation to components of the metric tensor, one arrives at the expression

$$\frac{\partial^6 g_{44}}{\partial x_1^2\partial x_2^2\partial x_3^2}. \tag{32}$$

The next possibility he considered was an expression quadratic in φ:

$$\frac{\partial^2\varphi}{\partial x_1^2}\frac{\partial^6\varphi}{\partial x_2^2\partial x_3^2\partial x_4^2}. \tag{33}$$

This term, as Einstein noted, would correspond to an expression with fourth-order derivatives of the metric: "will necessarily be of fourth order" ("wird notwendig 4. Ordnung").

To obtain an expression satisfying all three constraints, one needs an expression cubic in φ. Einstein thus wrote: "third degree in φ will be of second order, as it has to be" ("dritten Grades in φ wird 2. Ordnung, wie es sein muss") and then wrote down an example of such an expression:

$$\frac{\partial^2\varphi}{\partial x_1^2}\frac{\partial^2\varphi}{\partial x_2^2}\frac{\partial^4\varphi}{\partial x_3^2\partial x_4^2}. \tag{34}$$

40R The equivalence under linear coordinate transformations of the metric tensor and the partial derivatives of a scalar function is further explored at the top of p. 40R. There Einstein tried a different way of achieving symmetry between the four space-time coordinates. He changed what we identified above as the third constraint. Rather than looking at the expression corresponding to the field equations, he now exclusively concentrated on the second-order derivatives of φ representing components of the metric in this context. Taking as his starting point the product of two such second-order derivatives (corresponding to $g_{12}g_{34}$ [43]), he ensured that all four coordinates occur on equal footing by adding terms obtained through permutation of the indices. Einstein initially wrote down the cyclic permutations $g_{23}g_{41}$ and $g_{34}g_{12}$. However, $g_{34}g_{12} = g_{12}g_{34}$, so he crossed out this first attempt and tried again. He now wrote down the terms $g_{13}g_{24}$ and $g_{14}g_{23}$, the only non-redundant terms given the symmetries $g_{ik} = g_{ki}$ and $g_{mn}g_{ik} = g_{ik}g_{mn}$. He then translated these three terms into partial differential operators acting on φ:

43 Here, on top of p. 40R Einstein used the lower case g for the first time, a practice he continued throughout the rest of the notebook.

$$\frac{\partial^2\varphi}{\partial x_1\partial x_2}\frac{\partial^2\varphi}{\partial x_3\partial x_4}+\frac{\partial^2\varphi}{\partial x_1\partial x_3}\frac{\partial^2\varphi}{\partial x_2\partial x_4}+\frac{\partial^2\varphi}{\partial x_1\partial x_4}\frac{\partial^2\varphi}{\partial x_2\partial x_3}\,. \tag{35}$$

This expression is fully symmetric under permutation of the indices. By having a second-order derivative operator symmetric in all four coordinates (such as the d'Alembertian) acting on this expression, one can now construct field equations that meet all three constraints, the two original ones and the modified version of the third one. However, Einstein did not pursue this line of inquiry any further.

2.5 *Transforming the Ellipsoid Equation as a Model for Transforming the Line Element (40R–41L)* 40R–41L

Einstein's first exploration of a metric theory of gravitation ends at the horizontal line drawn on p. 40R with the calculations discussed above. The remaining pages of this part, pp. 40R–43L, do not seem to be a direct continuation of these investigations.

On the bottom half of p. 40R and the top half of p. 41L, Einstein considered transformations of equations describing three-dimensional ellipsoids. Only these calculations, which bear on Einstein's exploration of the transformation properties of the metric tensor, will be discussed here. On the bottom half of p. 41L and the top half of p. 41R, Einstein examined some properties of infinitesimal unimodular transformations. These calculations will be discussed in sec. 4.5.7 along with very similar calculations at the bottom of p. 12R. The bottom half of p. 41R, dealing with constrained motion along a two-dimensional surface, will be discussed in the sec. 4.5.8. Finally, most, if not all, of the material on pp. 42L–43L is related to Einstein's considerations on pp. 23L–R and will be discussed in secs. 5.5.6–5.5.10.

On the bottom half of p. 40R, Einstein transformed the equation for a three-dimensional ellipsoid to its principal-axis form. On the top half of p. 41L, he then tried to determine the class of linear transformations that would leave this description of the ellipsoid invariant. The calculation on p. 40R is analogous to finding coordinate transformations that take the line element in arbitrary coordinates to its standard Minkowski form. This geometrical analogy can also be found in (Einstein and Grossmann 1913, sec. 3): "the real cone $ds^2 = 0$ appears brought to its principal axes" ("der reelle Kegel $ds^2 = 0$ erscheint auf seine Hauptachsen bezogen"). The calculation on p. 41L is analogous to finding the class of linear transformations leaving the Minkowski line element in its standard diagonal form invariant.[44]

The calculation starts from the defining equation for an ellipsoid in Cartesian coordinates x, y, and z: 40R

$$\alpha_{11}x^2+2\alpha_{12}xy+\ldots..\alpha_{33}z^2 = 1\,. \tag{36}$$

44 These calculations may have been motivated by the following remark in (Wright 1908, 18): "The problem of the equivalence of two quadratic differential forms is reduced to that of the equivalence of two sets of algebraic forms, where one set is obtained from the other by a linear transformation."

For this equation to describe an ellipsoid, the left-hand side must be a positive definite quadratic form. Immediately below this expression, Einstein wrote the equation for the same ellipsoid in rotated primed coordinates such that the coordinate axes are aligned with the principal axes of the ellipsoid,

$$\frac{x'^2}{A^2} + \frac{y'^2}{B^2} + \frac{z'^2}{C^2} = 1 , \tag{37}$$

where A, B, and C are the lengths of the three semi-principal axes of the ellipsoid.

Next he wrote down the matrix of coefficients of the orthogonal coordinate transformation corresponding to this rotation:[45]

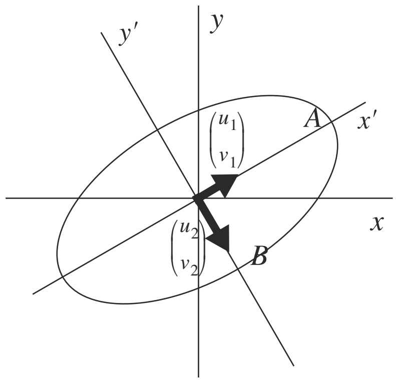
$$\begin{array}{c|c|c|c} & x' & y' & z' \\ \hline x & \alpha_1 & \alpha_2 & \alpha_3 \\ \hline y & \beta_1 & \beta_2 & \beta_3 \\ \hline z & \gamma_1 & \gamma_2 & \gamma_3 \end{array} . \tag{38}$$

Under the three columns of this matrix, Einstein wrote $1/A$, $1/B$, and $1/C$, respectively, indicating that he wanted to rescale the coefficients by these factors. He denoted the rescaled coefficients by $u_1 = \alpha_1/A$, $v_1 = \beta_1/A$, etc. The matrix of the rescaled coefficients is written below the transformation matrix,

$$\begin{array}{ccc} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{array} \tag{39}$$

For the remainder of these calculations, Einstein used the geometrical meaning of these quantities. To understand Einstein's reasoning, it is helpful to examine the two-dimensional case.

45 Similar transformation schemes appear on p. 39L and on the following p. 41L. More explicitly, the transformation reads $x' = \alpha_1 x + \beta_1 y + \gamma_1 z$, etc.



Consider the diagram on the left. In the primed coordinate system, the equation for the ellipse is:

$$\frac{x'^2}{A^2} + \frac{y'^2}{B^2} = 1\,, \tag{40}$$

which can be rewritten as

$$(\mathbf{u}_1 \cdot \mathbf{x})^2 + (\mathbf{u}_2 \cdot \mathbf{x})^2 = 1\,. \tag{41}$$

In the primed coordinate system, the vectors $\mathbf{x}$, $\mathbf{u}_1$ and $\mathbf{u}_2$ have components (x', y'), $(1/A, 0)$ and $(0, -1/B)$, respectively.

Einstein wrote down the three-dimensional analogue of equation (41) in the unprimed coordinates:[46]

$$(u_1 x + v_1 y + w_1 z)^2 + (u_2 x + v_2 y + w_2 z)^2 + (u_3 x + v_3 y + w_3 z)^2 = 1\,. \tag{42}$$

Expanding the binomials and comparing the resulting coefficients with those of equation (36), one finds[47]

$$\begin{aligned} u_1^2 + u_2^2 + u_3^2 &= \alpha_{11} & \quad v_1 w_1 + v_2 w_2 + v_3 w_3 &= \alpha_{23} \\ v_1^2 + v_2^2 + v_3^2 &= \alpha_{22} & \quad w_1 u_1 + w_2 u_2 + w_3 u_3 &= \alpha_{31} \\ w_1^2 + w_2^2 + w_3^2 &= \alpha_{33} & \quad u_1 v_1 + u_2 v_2 + u_3 v_3 &= \alpha_{12}. \end{aligned} \tag{43}$$

The vectors (u_1, v_1, w_1), (u_2, v_2, w_2), and (u_3, v_3, w_3) are in the direction of the semi-axes of the ellipsoid (cf. the figure above). It follows that these three vectors are orthogonal and that their norms are the reciprocal of the lengths of the ellipsoid's semi-axes. The orthogonality of the three vectors is expressed by:

$$\begin{aligned} u_1 u_2 + v_1 v_2 + w_1 w_2 &= 0, \\ u_1 u_3 + v_1 v_3 + w_1 w_3 &= 0, \\ u_2 u_3 + v_2 v_3 + w_2 w_3 &= 0. \end{aligned} \tag{44}$$

Their norms are given by:

46 The third term is only indicated by a dot in the notebook.

47 The equations for α_{33}, α_{31}, and α_{12} are indicated by dashes in the notebook, and instead of α_{23} Einstein erroneously wrote α_{12}.

$$
\begin{aligned}
u_1^2 + v_1^2 + w_1^2 &= \frac{1}{A^2}, \\
u_2^2 + v_2^2 + w_2^2 &= \frac{1}{B^2}, \\
u_3^2 + v_3^2 + w_3^2 &= \frac{1}{C^2}.
\end{aligned}
\tag{45}
$$

Equations (44) and (45) are the ones written down at the bottom of p. 40R.[48]

41L On the next page, Einstein recapitulated, remarking that “the u, v, w determine orientation and size of the ellipsoid” (“Die u, v, w bestimmen Lage und Grösse des Ellipsoids.”). Equation (42) for the ellipsoid can be written $\sum(u_i x + v_i y + w_i z)^2 = 1$, the left-hand side of which can be interpreted as the sum of the squared scalar products of (u_1, v_1, w_1), (u_2, v_2, w_2), and (u_3, v_3, w_3) with the radius vector x, y, z.

On p. 40R, Einstein had been concerned with the transformation to principal axes for the ellipsoid. On p. 41L, he investigated “arbitrary linear transformations of x, y, z to x', y', z' for an invariant ellipsoid function” (“Beliebige lineare Transformationen der x, y, z in x', y', z' bei invarianter Ellipsoidfunktion.”). In other words, he asked which linear transformation leave the principal-axes form of the equation for the ellipsoid invariant. He thus wrote down the equation for the ellipsoid in two different coordinate systems:[49]

$$\sum(u_i x + v_i y + w_i z)^2 = \sum(u'_i x' + v'_i y' + w'_i z')^2 \ . \tag{46}$$

Einstein next wrote down transformation matrices for the linear transformations between the three spatial coordinates that parallel those introduced earlier for the four spacetime coordinates:

$$
\begin{array}{c|ccc}
 & x' & y' & z' \\
x & \beta_{11} & \beta_{21} & \beta_{31} \\
y & \beta_{12} & \beta_{22} & \beta_{32} \\
z & \beta_{13} & \beta_{23} & \beta_{33},
\end{array}
\tag{47}
$$

and:[50]

48 The last two lines of equations (44) and (45) are only indicated by dashes in the notebook.

49 In writing this expression, Einstein seems to have added the summation signs as an afterthought: additional terms on the right-hand side were deleted in favor of the summation sign, and the summation index “i” in that equation is an “l” in the notebook.

50 Note that the notation here is the reverse of the notation introduced on p. 39L (cf. equation (2)).

$$
\begin{array}{c|ccc}
 & x & y & z \\
x' & \alpha_{11} & \alpha_{21} & \alpha_{31} \\
y' & \alpha_{12} & \alpha_{22} & \alpha_{32} \\
z' & \alpha_{13} & \alpha_{23} & \alpha_{33}.
\end{array}
\tag{48}
$$

From equation (46), it follows that the transformation matrix for the vectors (u_1, v_1, w_1), (u_2, v_2, w_2), and (u_3, v_3, w_3) is the transposed of the matrix in equation (47), i.e., in Einstein's notation:

$$
\begin{array}{c|ccc}
 & u'_1 & v'_1 & w'_1 \\
u_1 & \beta_{11} & \beta_{12} & \beta_{13} \\
v_1 & \beta_{21} & \beta_{22} & \beta_{23} \\
w_1 & \beta_{31} & \beta_{32} & \beta_{33}.
\end{array}
\tag{49}
$$

Using these transformation equations, one can write:

$$
\begin{aligned}
u'^2_1 &= (u_1\beta_{11} + v_1\beta_{21} + w_1\beta_{31})^2, \\
u'^2_2 &= (u_2\beta_{11} + v_2\beta_{21} + w_2\beta_{31})^2, \\
u'^2_3 &= (u_3\beta_{11} + v_3\beta_{21} + w_3\beta_{31})^2.
\end{aligned}
\tag{50}
$$

The sum of these three expressions may be more conveniently expressed as:

$$
u'^2_1 + u'^2_2 + u'^2_3 = \sum_{i=1}^{3} (u_i\beta_{11} + v_i\beta_{21} + w_i\beta_{31})^2 \tag{51}
$$

In the notebook this equation is written as[51]

$$
u'^2_1 + u'^2_2 + u'^2_3 = \sum (u_1\beta_{11} + v_1\beta_{21} + w_1\beta_{31})^2 . \tag{52}
$$

Similarly, he found:

$$
v'^2_1 + v'^2_2 + v'^2_3 = \sum (u_1\beta_{12} + v_1\beta_{22} + w_1\beta_{32})^2 . \tag{53}
$$

The third equation, for $w'^2_1 + w'^2_2 + w'^2_3$, is indicated by a dashed line. At this point, the calculation breaks off.

The remaining pages of the part starting from the back of the notebook contain material that was added later and will be covered later. The calculations on the bottom half of p. 41L and on p. 41R will be discussed in section 4; the calculations on p. 42L–43L in section 5.

---

51 In other words, Einstein neglected to change u_1, v_1, w_1 to u_i, v_i, w_i on the right-hand side.

3. ENERGY-MOMENTUM BALANCE BETWEEN MATTER AND GRAVITATIONAL FIELD (5R)

We now turn to the notes on gravitation that start from the back of the notebook,
5R beginning with p. 5R. The calculations on these pages are not a direct continuation of those on pp. 39L–41R examined in sec. 2. A clear indication that p. 5R ff. are later is that the calculations on these pages are much more sophisticated mathematically than those on pp. 39L–41R.

On p. 5R, Einstein derived the equation for the energy-momentum balance for pressureless dust in the presence of a gravitational field,[52] an argument that later appeared in (Einstein and Grossmann 1913, secs. 2 and 4). It starts with the derivation of the equations of motion of a point particle in a metric field from an action principle, where the action integral is just the proper length of the particle's worldline. Expressions for the particle's momentum and the gravitational force acting on the particle are read off from the resulting Euler-Lagrange equations. Einstein generalized these results to expressions for the momentum density and the force density in the case of pressureless dust, or, as it is described in (Einstein and Grossmann 1913, 9), "continuously distributed incoherent masses" ("kontinuierlich verteilter inkohärenter Massen"). He identified the expression for momentum density as part of the stress-energy tensor for pressureless dust. Inserting this stress-energy tensor and a similar expression for the density of the force acting on the pressureless dust into the equations of motion that he started from, Einstein arrived at a plausible candidate for the law of energy-momentum conservation in the presence of a gravitational field, or, more accurately, an equation for the energy-momentum balance between matter and gravitational field.[53] What made the candidate all the more promising were its transformation properties. In fact, the equation is generally covariant. As was shown explicitly by Grossmann in his part of (Einstein and Grossmann 1913*;* part II, sec. 4), it expresses the vanishing of the covariant four-divergence of the stress-energy tensor. On p. 5R, Einstein made a similar claim, namely that the result of the expression found at the bottom of the page is always a vector. He performed a calculation for a special case that at least made this claim plausible. The equation thus looked like a promising generalization of the vanishing of the ordinary four-divergence of the stress-energy tensor, which Laue (1911a, b) had made the fundamental equation of relativistic mechanics. Like Laue, Einstein presumably wanted to generalize it from the special case of pressureless dust to arbitrary physical systems.[54]

Einstein's analysis on p. 5R thus provides an excellent example of how physical and mathematical considerations complement each other in the course of Einstein's

---

52 Essentially the same calculation can be found on p. 20R. For discussion see sec. 5.4.4. P. 5R is also discussed in (Norton 2000, appendices A through C).

53 On p. 43LB, Einstein performed a calculation that is the "inverse" of that on p. 5R. Rather than deriving the law of energy-momentum balance from the equations of motion of a point particle, he derived the equations of motion from the law of energy-momentum balance. This calculation will be discussed in sec. 5.5.10.

development of the general theory of relativity. First, Einstein derived an equation for the energy-momentum balance between matter and gravitational field on the basis of physical arguments centering on the special case of pressureless dust. Then, he confirmed its generalizability to arbitrary physical systems on the basis of mathematical arguments.

We shall now examine Einstein's calculations on p. 5R in detail. He started by writing down the line element,

$$g_{11}dx^2 + \ldots + g_{44}dt^2 = ds^2 . \tag{54}$$

Next to it, he noted that this expression is "always positive for a point" ("immer positiv für Punkt"), i.e., the worldline of a material particle is time-like.[55]

As he had first done in a note added in proof to the paper presenting his second static theory (Einstein 1912b, 458), Einstein used this line element to define the Lagrangian—or "Hamiltonian function" ("Hamiltonsche Funktion") as he called it (Einstein and Grossmann 1913, 7)—for a point particle moving in a given metric field:[56]

$$\frac{ds}{dt} = H . \tag{55}$$

He then wrote down the corresponding Euler-Lagrange equations, the "equations of motion" ("Bewegungsgleichungen"):

$$\frac{d}{dt}\left(\frac{\partial H}{\partial \dot{x}}\right) - \frac{\partial H}{\partial x} = 0 . \tag{56}$$

The equation initially had a plus rather than a minus sign on the left-hand side. This sign error is carried through all the way to the end of the calculation. Einstein only corrected it after he discovered that it led to an unacceptable end result (see the discussion following equation (75)). The equation,

---

54 Einstein made this generalization explicit in his paper with Grossmann: "We ascribe to equation (10) a range of validity that goes far beyond the special case of the flow of incoherent masses. The equation represents in general the energy balance between the gravitational field and an arbitrary material process" ("Der Gleichung (10) schreiben wir einen Gültigkeitsbereich zu, der über den speziellen Fall der Strömung inkohärenter Massen weit hinausgeht. Die Gleichung stellt allgemein die Energiebilanz zwischen dem Gravitationsfelde und einem beliebigen materiellen Vorgang dar …;" Einstein and Grossmann 1913, p. 11). Statements almost verbatim the same as this one can be found in the printed text of Einstein's lecture on the problem of gravitation in Vienna in September 1913 (Einstein 1913, 1253 and 1257).

55 From this we can infer that at this point, Einstein's sign convention is such that the Minkowski metric in its standard diagonal form is $\mathrm{diag}(-1, -1, -1, c^2)$. This same convention was used in the *Nachtrag* to (Einstein 1912b) and in (Einstein and Grossmann 1913). On p. [39L], however, the Minkowski metric in its standard diagonal form was given as $\mathrm{diag}(1, 1, 1, -c^2)$ (if we switch from an imaginary to a real time coordinate).

56 Given the sign convention on this page (see the preceding footnote), H should be $-H$ (cf. Einstein and Grossmann 1913, 7; Einstein 1912b, 458).

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}}\right) = -\frac{\partial \Phi}{\partial x}\,, \tag{57}$$

written to the right of equation (56), was presumably added at that point and helped Einstein correct his sign error. It is easy to see how equation (57) could play that role. When the Lagrangian can be written as $H = L - \Phi$, where Φ is a potential energy term and L (which stands for "lebendige Kraft") is the kinetic energy term, equation (56) is equivalent to equation (57). The latter equation is readily recognized as a generalization of Newton's second law. This then confirmed that the left-hand side of equation (56) must indeed have a minus sign.

Underneath equation (56), Einstein wrote down the partial derivative of H with respect to $\dot{x}$:[57]

$$\frac{\partial H}{\partial \dot{x}} = \frac{g_{11}\dot{x} + g_{12}\dot{y} + . + g_{14}}{ds/dt}\,. \tag{58}$$

This is the x-component of the momentum of a particle of unit mass. He then generalized this result to an expression for the x-component of the momentum density of pressureless dust:[58]

$$\rho_0\sqrt{G}\left(g_{11}\frac{dx}{ds}\frac{dt}{ds} + g_{12}\frac{dy}{ds}\frac{dt}{ds} + \cdot + \cdot\right) \tag{59}$$

The derivation of this expression is not given in the notebook, but can easily be reconstructed (see Einstein and Grossmann 1913, secs. 2 and 4). Using equation (58) for the x-component of momentum J_x and dividing it by the volume V in the same coordinate system, one arrives at an expression for the x-component of momentum density

$$\frac{J_x}{V} = \frac{1}{V}\frac{\partial H}{\partial \dot{x}} = \frac{1}{V}\frac{g_{11}\dot{x} + g_{12}\dot{y} + g_{13}\dot{z} + g_{14}}{ds/dt}\,. \tag{60}$$

To obtain equation (59), the coordinate volume V has to be replaced by the proper volume V_0. The relation between the two is given by (Einstein and Grossmann 1913, 10):[59]

$$V = V_0\frac{ds}{dt}\frac{1}{\sqrt{G}}\,, \tag{61}$$

where G is minus the determinant of the metric tensor. Substituting this expression for V on the right-hand side of equation (60), one arrives at equation (59) for what, as

---

57 $\frac{\partial H}{\partial \dot{x}} = \frac{\partial}{\partial \dot{x}}\left(\sqrt{\frac{ds^2}{dt^2}}\right) = \frac{1}{2}\frac{dt}{ds}\frac{\partial}{\partial \dot{x}}\left(g_{\mu\nu}\frac{dx_\mu}{dt}\frac{dx_\nu}{dt}\right) = \frac{dt}{ds}g_{1\nu}\dot{x}^\nu$.

58 Given the sign convention on this page (see footnote 55), G stands for minus the determinant of the metric.

59 This relation can also be found on p. 20R (see equation (524)) and at the bottom of p. 27L.

Einstein wrote under it, "is the momentum per unit volume" ("ist Bewegungsgrösse pro Volumeinheit"):

$$
\begin{aligned}
\frac{J_x}{V} &= \frac{1}{V_0}\sqrt{G}\frac{dt}{ds}\frac{dt}{ds}(g_{11}\dot{x} + g_{12}\dot{y} + g_{13}\dot{z} + g_{14}) \\
&= \frac{1}{V_0}\sqrt{G}\left(g_{11}\frac{dx}{dt}\frac{dt}{ds}\frac{dt}{ds} + g_{12}\frac{dy}{dt}\frac{dt}{ds}\frac{dt}{ds} + . + g_{14}\frac{dt}{ds}\frac{dt}{ds}\right) \\
&= \rho_0\sqrt{G}\left(g_{11}\frac{dx}{ds}\frac{dt}{ds} + g_{12}\frac{dy}{ds}\frac{dt}{ds} + \cdot + \cdot\right),
\end{aligned}
\tag{62}
$$

where $\rho_0 = 1/V_0$ is the proper density of a unit mass of dust. The energy density E/V of the dust can be found through a similar argument:[60]

$$
\frac{E}{V} = \rho_0\sqrt{G}\sum g_{4\nu}\frac{dx_\nu}{ds}\frac{dx_4}{ds}. \tag{63}
$$

Equation (62) for the x-component of momentum density, similar expressions for its y- and z-components, and equation (63) for the energy density, all divided by $\sqrt{G}$, give four components of the contravariant stress-energy tensor for pressureless dust, which is written down immediately below equation (59):[61]

$$
T^b_{ik} = \rho_0\frac{dx_i}{ds}\frac{dx_k}{ds}. \tag{64}
$$

Einstein called this quantity the "tensor of the motion of masses" ("Tensor der Bewegung der Massen").[62] Contracting this tensor with the metric and multiplying by $\sqrt{G}$, one recovers the momentum and energy densities (cf. equations (62)–(63))

$$
\frac{J_x}{V} = \sqrt{G}\sum g_{1i}T^b_{i4}, \quad \frac{E}{V} = \sqrt{G}\sum g_{4i}T^b_{i4}. \tag{65}
$$

Einstein thus introduced what he called the "tensor of momentum and energy" ("Tensor der Bewegungsgröße u[nd] Energie.").

$$
T_{mn} = \left|\sum\sqrt{G}g_{m\nu}T^b_{\nu n}\right|. \tag{66}
$$

In fact, this is not a tensor but a (mixed) tensor density.

---

60 Cf. footnote 57:

$$
\frac{E}{V} = \left(\frac{1}{V_0}\sqrt{G}\frac{dt}{ds}\right)\frac{\partial H}{\partial\dot{x}_4} = \left(\frac{1}{V_0}\sqrt{G}\frac{dt}{ds}\right)\left(\frac{dt}{ds}g_{4\nu}\dot{x}^\nu\right) = \rho_0\sqrt{G}\sum g_{4\nu}\frac{dx_\nu}{dt}\frac{dt}{ds}\frac{dt}{ds} = \rho_0\sqrt{G}\sum g_{4\nu}\frac{dx_\nu}{ds}\frac{dx_4}{ds}.
$$

61 Note that the indices are written "downstairs" even though they refer to contravariant components. Einstein's convention here and elsewhere in the notebook is just the opposite of the one adopted in (Einstein and Grossmann 1913), where all contravariant quantities are indicated by Greek and all covariant ones by Latin characters.

62 The superscript 'b' stands for "Bewegung" ("motion").

With the help of equation (64), Einstein could begin to rewrite the Euler-Lagrange equations (56) in terms of a stress-energy tensor, which would allow him to generalize the equation from the special case of pressureless dust to any physical system described by a stress-energy tensor. The second term of equation (56) represents the gravitational force on a particle. Einstein generalized this term to an expression for the "negative ponderomotive force per volume unit" ("Negative Ponderomotorische Kraft pro Volumeinheit")[63] on pressureless dust:[64]

$$\frac{1}{2}\sqrt{G}\sum\frac{\partial g_{\mu\nu}}{\partial x_m}T^b_{\mu\nu}\,. \tag{67}$$

As with equation (59), the derivation of this expression is not given in the notebook, but can easily be reconstructed (see Einstein and Grossmann 1913, secs. 2 and 4). Defining the x-component of the gravitational force $\mathfrak{K}$ as $\partial H/\partial x$ and using the definition of the Lagrangian H, one can write:

$$\frac{\mathfrak{K}_1}{V} = \frac{1}{V}\frac{\partial H}{\partial x_1} = \frac{1}{V}\frac{\partial}{\partial x_1}\left(\frac{ds}{dt}\right). \tag{68}$$

Using the definition of the line element, one can rewrite this as:

$$\frac{\mathfrak{K}_1}{V} = \frac{1}{V}\frac{\frac{\partial}{\partial x_1}\sqrt{g_{\mu\nu}dx_\mu dx_\nu}}{dt} = \frac{1}{V}\frac{1}{2}\frac{\sum\limits_{\mu\nu}\frac{\partial g_{\mu\nu}}{\partial x_1}dx_\mu dx_\nu}{ds dt} = \frac{1}{V}\frac{1}{2}\frac{\sum\limits_{\mu\nu}\frac{\partial g_{\mu\nu}}{\partial x_1}\frac{dx_\mu}{dt}\frac{dx_\nu}{dt}}{ds/dt} \tag{69}$$

With the help of equation (61), the coordinate volume V can be replaced by the proper volume V_0:

$$\frac{\mathfrak{K}_1}{V} = \frac{1}{V_0}\sqrt{G}\frac{1}{2}\frac{dt}{ds}\frac{dt}{ds}\sum_{\mu\nu}\frac{\partial g_{\mu\nu}}{\partial x_1}\frac{dx_\mu}{dt}\frac{dx_\nu}{dt} = \frac{1}{2}\rho_0\sqrt{G}\sum_{\mu\nu}\frac{\partial g_{\mu\nu}}{\partial x_1}\frac{dx_\mu}{ds}\frac{dx_\nu}{ds}\,. \tag{70}$$

Inserting $T^b_{\mu\nu}$, the stress-energy tensor for pressureless dust, for $\rho_0(dx_\mu/ds)(dx_\nu/ds)$, one arrives at the $m = 1$-component of Einstein's equation (67). As he had done before (see equation (66)), Einstein assumed that this result, derived for the special case of the stress-energy tensor of pressureless dust, will hold for the stress-energy tensor of any matter.

On the next line Einstein wrote down the tensorial generalization of the Euler-Lagrange equations (56), using equations (66) and (67) and replacing d/dt by $\partial/\partial x_n$:[65,66]

$$\sum_{\nu n}\frac{\partial}{\partial x_n}(\sqrt{G}g_{m\nu}T_{\nu n}) - \frac{1}{2}\sqrt{G}\sum\frac{\partial g_{\mu\nu}}{\partial x_m}T_{\mu\nu} = 0\,. \tag{71}$$

---

63 The word "Negative" was later added.

64 In the expression below, G was corrected from D (which presumably stands for "Determinante"), the subscript m was corrected from some illegible character, and a factor ρ_0 was deleted.

At this point he dropped the index 'b' since he expected the equation to hold for all matter, not just for pressureless dust.

Einstein now tried to gain some insight into the transformational behavior of this equation by rewriting its left-hand side as the action of a generalized derivative on an arbitrary tensor. First, he introduced the (contravariant) tensor density corresponding to the (contravariant) stress-energy tensor $T_{\mu\nu}$ in equation (71). He wrote: "If we set" ("Setzen wir")[67]

$$\sqrt{G}T_{\mu\nu} = \Theta_{\mu\nu}, \tag{72}$$

then equation (71) can be rewritten as:

$$\sum_{\nu n}\frac{\partial}{\partial x_n}(g_{m\nu}\Theta_{\nu n}) - \frac{1}{2}\sum\frac{\partial g_{\mu\nu}}{\partial x_m}\Theta_{\mu\nu} = 0. \tag{73}$$

To write the left-hand side as a differential operator acting on a tensor, Θ should appear with the same indices in both terms. Therefore Einstein relabeled the summation indices in the first term in order to conform with the occurrence of $\Theta_{\mu\nu}$ in the second term, made the assumption that $\Theta_{\mu\nu}$ is symmetric, and rewrote the equation as:[68]

$$\sum_{\mu\nu}\frac{\partial}{\partial x_\mu}(g_{m\nu}\Theta_{\mu\nu}) - \frac{1}{2}\sum\frac{\partial g_{\mu\nu}}{\partial x_m}\Theta_{\mu\nu} = 0. \tag{74}$$

To the right of this equation, he wrote: "In general associated vector" ("Im Allgemeinen zugeordneter Vektor"), and below it "Valid for every symmetric tensor, e.g., $\sqrt{G}\gamma_{\mu\nu}$" ("Gilt für jeden symm[etrischen] Tensor z. B. $\sqrt{G}\gamma_{\mu\nu}$").[69] "Associated" ("zugeordnet") means "covariant" in this context, as is clear from its usage on pp. 6L–R. The notation $\gamma_{\mu\nu}$ for the contravariant components of the metric (see Einstein and Grossmann 1913, 12, note 4) occurs here for the first time in the notebook. Einstein's

65 The expression in parentheses is underlined once, and $T_{\mu\nu}$ in the second term is underlined twice. The equation also contains a number of corrections. The minus sign was corrected from a plus sign. This error stems from the sign error in equation (56). In the first term, the index n was corrected from m and the factor $\sqrt{G}$ was inserted later. In the second term, the factor $1/2$ was deleted once and rewritten. This correction is related to a correction in equation (75) below. Moreover, G was corrected from D (as in equation (67)). Note that $T_{\mu\nu}$ in equation (71) stands for the contravariant stress-energy tensor, whereas in equation (66) it stands for the corresponding mixed tensor density.

66 The same equation appears in slightly different versions and in slightly different notation at various other places in the notebook (see pp. 24L, 26L, 28L, and 43LB) and in (Einstein and Grossmann 1913, p. 10, equation (10)).

67 The notation in the notebook (see also pp. 20R, 24L, 24R, 26L, and 43LB) is different from the notation used in (Einstein and Grossmann 1913), where the contravariant stress-energy tensor is denoted by $\Theta_{\mu\nu}$ and the covariant form by $T_{\mu\nu}$.

68 The equation originally had a plus rather than a minus sign, an error it inherited from equation (56). The factor $1/2$ was deleted once and rewritten. This correction is related to a correction in equation (75) below. The index ν in $g_{\mu\nu}$ is corrected from μ. Both characters Θ are corrected from T's.

69 The word "symm[etric]" is interlineated. Recall that Einstein had to assume symmetry in going from equation (73) to equation (74).

claim is that the quantity on the left-hand side of equation (74) is a vector as long as $\Theta_{\mu\nu}$ is a symmetric tensor. This claim is correct. The expression is just the covariant divergence of a symmetric tensor, as is shown in (Einstein and Grossmann 1913). The result appears in Grossmann's part, which suggests that Einstein may have learnt it from him.

Einstein checked his claim for the specific example $\sqrt{G}\gamma_{\mu\nu}$ mentioned above. Insertion of this tensor density into equation (74) produces:[70]

$$\sum_{\mu\nu}\frac{\partial}{\partial x_\mu}(\sqrt{G}g_{m\nu}\gamma_{\mu\nu})-\frac{1}{2}\sum_{\mu\nu}\left(\sqrt{G}\frac{\partial g_{\mu\nu}}{\partial x_m}\gamma_{\mu\nu}\right) = 0\,, \tag{75}$$

Since $\sum g_{m\nu}\gamma_{\mu\nu} = \delta_{m\mu}$, with $\delta_{m\mu}$ the Kronecker delta, the first term is equal to

$$\frac{\partial\sqrt{G}}{\partial x_m}\,, \tag{76}$$

which Einstein wrote underneath the first term of equation (75). Since[71]

$$\sum_{\mu\nu}\frac{\partial g_{\mu\nu}}{\partial x_m}\gamma_{\mu\nu} = \frac{1}{G}\frac{\partial G}{\partial x_m}\,, \tag{77}$$

the expression in parentheses in the second term is equal to

$$\frac{1}{\sqrt{G}}\frac{\partial G}{\partial x_m}\,, \tag{78}$$

which Einstein wrote underneath the second term of equation (75).[72]

Inserting expressions (76) and (78) into equation (75), one sees that the latter is indeed satisfied.

It is at this point that Einstein recognized two errors that had found their way into his calculations. First, equation (75) inherited the erroneous minus sign from the Euler-Lagrange equations (56) that formed the starting point of this whole calculation. Second, Einstein omitted a factor $1/2$ when he went from equation (74) to equation (75). Einstein probably only caught these errors when the result of his calculations did not meet this expectation, i.e., when he saw that equation (75) with a plus rather than a minus and without the factor $1/2$ does not hold. Einstein probably corrected the sign error first, tracing it all the way back to the Euler-Lagrange equations (56). Equation (57) next to these equations was probably added in this context. As to the omitted factor $1/2$, Einstein originally seems to have been under the impression that expression (76) is equal to expression (78) without the factor $1/2$ multiplying the latter in equation (75). It is probably for this reason that he deleted this

70 The equation originally had a plus rather than a minus sign, an error it inherited from equation (56). Einstein initially omitted the factor of $1/2$. The factor $\sqrt{G}$ was inserted later.

71 See p. 6L for a derivation of this relation.

72 The index m is corrected from μ.

factor in equations (74) and (71), respectively, only to put them back in once he realized that this factor was needed in equation (75) after all.

After making these corrections, Einstein wrote at the bottom of the page: "correct" ("stimmt"). Einstein's trial calculation supported the claim that his physically motivated expression for energy-momentum balance between matter and gravitational field does indeed lead to a differential operator that acts on a symmetric tensor to produce a vector.[73]

4. EXPLORATION OF THE BELTRAMI INVARIANTS AND THE CORE OPERATOR (6L–13R, 41L–R)

4.1 *Introduction* (6L–13R, 41L–R)

Einstein returned to the question of finding field equations for the metric field on p. 6L. He had meanwhile become more sophisticated mathematically. For example, at this point he knew about the Beltrami invariants and carefully distinguished between covariant and contravariant tensors. However, there were still large gaps in his knowledge. He still did not know about the Riemann tensor or covariant differentiation, which severely handicapped his search for satisfactory field equations. Most of the calculations on pp. 6L–13L are investigations of the covariance properties of various expressions that might either be part of the field equations or play a role in their construction. These calculations did not lead to any promising candidates for the left-hand side of the field equations, but they led to several clusters of important results, ideas, and techniques that Einstein was able to put to good use once he learned about the Riemann tensor (see p. 14L). 6L

First of all, one can begin to discern the double strategy discussed in sec. 1.2. On p. 6L Einstein started with two generally-covariant operators acting on a scalar function, known as the Beltrami invariants. The second Beltrami invariant is a generally-covariant generalization of the Laplacian of a scalar function and as such provided a natural point of departure in Einstein's search for gravitational field equations. The basic challenge was to get from an operator acting on a scalar function to an operator acting on the metric tensor. Einstein did not immediately see how to achieve this goal. On p. 7L, he therefore temporarily abandoned his mathematically-oriented approach for a physically-oriented one. He wrote down a version of what we call the "core operator," an expression constructed out of the metric tensor and its coordinate derivatives in such a way that for weak fields it reduces to the d'Alembertian acting on the metric. The problem with this core operator is that its transformation properties are unclear. Einstein addressed this problem by trying to relate the physically well-understood core operator to the mathematically well-understood Beltrami invariants. 7L

73 On p. 8R, Einstein found that the generally-covariant generalization of the exterior derivative operator acting on the metric also vanishes (see the discussion following equation (182)).

Two key components of Einstein's mathematical strategy first emerge in the course of the calculations documented on these pages. The first is the idea is to start from expressions with a well-defined covariance group (either generally covariant or covariant under unimodular coordinate transformations) and then to extract candidates for the left-hand side of the field equations by imposing additional conditions, such as the condition that the correspondence principle be satisfied (i.e., that the field equations reduce to the Poisson equation of Newtonian theory for weak static fields).

The other key component is the use of what Einstein would later call "non-autonomous transformations" ("unselbständige Transformationen").[74] To investigate the covariance properties of various expressions constructed out of the metric tensor and its derivatives, he wrote out the transformation law for such an expression under general coordinate transformations and identified those terms that would have to vanish if the expression were to transform as a tensor. The vanishing of these terms gives conditions on the transformation matrices that depend explicitly on the components of the
7L-R, metric (see pp. 7L-R, 9R, 10L). Because of this dependence, such transformations are
9R, 10L called "non-autonomous." Initially, Einstein investigated the behavior of his candidate field equations under such non-autonomous coordinate transformations. Eventually he realized that the complexity of the relevant calculations could be reduced considerably by combining the notion of non-autonomous transformations with the basic idea of the mathematical strategy, namely to extract candidate field equations from expressions with well-known covariance properties by imposing additional conditions. The covariance properties of field equations constructed in this fashion are determined by the covariance properties of these additional conditions, which typically will be much simpler than the field equations themselves. Einstein could thus focus on determining the class of non-autonomous transformations under which these simpler additional conditions transform tensorially.[75]

An important example of such an additional condition is what we shall call the "Hertz restriction," in which the so-called "Hertz expression," $\sum \partial\gamma_{\mu\nu}/\partial x_\nu$, is set
10L–11L equal to zero.[76] On pp. 10L–11L, Einstein investigated whether the class of non-autonomous transformations under which the Hertz expression transforms as a tensor includes the transformations from quasi-Cartesian coordinates to rotating and linearly accelerating frames in the special case of Minkowski spacetime. These transformations were crucially important to Einstein's attempts to extend the relativity principle from uniform to non-uniform motion and to establish the equivalence of rotation and acceleration in Minkowski spacetime to corresponding gravitational fields.

Later in the notebook, Einstein applied the same strategy to candidate field equations constructed out of the Riemann tensor. As before, the covariance properties of the additional condition(s) determine the covariance properties of the candidate field

74 See Einstein to H. A. Lorentz, 14 August 1913 (CPAE 5, Doc. 467). See sec. 4.3 for further discussion.

75 See p. 9R and the discussion in footnote 134 for the first example of this type of argument.

76 Named after Paul Hertz (see footnote 22). The Hertz restriction reappears in the course of Einstein's exploration of the Riemann tensor on p. 22R (see sec. 5.5.2).

equations. Einstein regarded such conditions as essential parts of the theory, restricting the class of admissible coordinate systems. We therefore refer to these conditions as "coordinate restrictions." They should be distinguished from "coordinate conditions." Mathematically, one and the same equation can express a coordinate restriction or a coordinate condition, but the two have a very different status. Coordinate conditions may always be imposed on generally-covariant equations to choose a suitable class of coordinate systems for some particular purpose. Consequently, different coordinate conditions may be used for different purposes, just as different gauge conditions can be used for different purposes. In contrast, coordinate restrictions are an integral part of the theory, imposing a limitation on the allowed class of coordinate systems in which the theory is expected to hold. With coordinate restrictions, one does not have the freedom to pick different conditions for different purposes.

On p. 9L, building on the energy-momentum balance equation derived on p. 5R, 9L
Einstein arrived at two important insights related to energy-momentum conservation. First, drawing on his experience with the 1912 static theory, he realized that for the field equations to be compatible with energy-momentum conservation their left-hand side should be the sum of a core operator and a quantity representing the stress-energy of the gravitational field. Secondly, he found a way to construct a candidate for this stress-energy tensor out of the first Beltrami invariant.

Einstein's main concern in this part of the notebook, however, was not how to satisfy the conservation principle but how to satisfy simultaneously his other three heuristic requirements, the correspondence principle, the relativity principle, and the equivalence principle.[77] By the time he made the entries at the bottom of p. 12R, Einstein had established a number of results related to the latter three principles. He had
(*i*) introduced the core operator (p. 7L); (*ii*) found its relation to the second Beltrami 7L–12R,
invariant with the determinant of the metric playing the role of the scalar function in 41L–R
the latter's definition (pp. 8R–9R); (*iii*) investigated unimodular transformations since the determinant of the metric transforms as a scalar under this restricted class of transformations only; (*iv*) recognized the importance of the Hertz restriction in getting from the second Beltrami invariant first to the core operator and then to weak field equations with the d'Alembertian acting on the metric (p. 10L); (*v*) derived conditions for the classes of non-autonomous transformations under which the weak field equations and the Hertz expression transform as a tensor and a vector, respectively (pp. 10L–11R); and (*vi*) developed a strategy for checking whether such non-autonomous transformations include the important special cases of the transformation to rotating and accelerating frames in Minkowski spacetime (pp. 11L, 12L–R, 41L–R). However, serious difficulties on all these counts remained.

On pp. 13L–R, Einstein therefore bracketed the question of the covariance proper- 13L–R
ties of his candidate field equations and turned to the compatibility with energy-momentum conservation instead, provisionally requiring only covariance under autonomous unimodular linear transformations. On p. 13R he returned to the physical

---

77 See sec. 1.1 for discussion of these requirements.

strategy trying to find the left-hand side of the field equations by starting from the weak field equations and imposing energy-momentum conservation. An ingenious simpler version of this strategy was used on pp. 26L–R to derive the *Entwurf* field equations. On p. 13R Einstein did not yet see how to generate field equations in this way. Given this impasse and the difficulties he had run into in his investigation of the covariance properties on pp. 6L–12R, one can well imagine Einstein turning to his mathematician friend Marcel Grossmann for fresh ideas.[78] On the next page, p. 14L, the Riemann curvature tensor makes its first appearance in the notebook with Grossmann's name written next to it. Einstein only returned to the physical strategy emerging on p. 13R after a series of failed attempts to extract field equations from the Riemann tensor (see pp. 14L–25R, 42L–43L covered in sec. 5).

6L–7L

4.2 *Experimenting with the Beltrami Invariants (6L–7L)*

Beltrami's two differential invariants (more precisely "differential parameters") are generally-covariant scalars constructed out of the metric, its first- and second-order derivatives, and some arbitrary, at least twice-differentiable scalar function φ.[79] The first Beltrami invariant can be defined as

$$\Delta_1\varphi = \gamma_{\mu\nu}\frac{\partial\varphi}{\partial x_\mu}\frac{\partial\varphi}{\partial x_\nu}, \tag{79}$$

the second as[80]

$$\Delta_2\varphi = \frac{1}{\sqrt{G}}\frac{\partial}{\partial x_\mu}\left(\gamma_{\mu\nu}\sqrt{G}\frac{\partial\varphi}{\partial x_\nu}\right). \tag{80}$$

6L

At the top of p. 6L, Einstein wrote down the contraction of the contravariant metric with the gradient of some scalar function φ

$$\sum\gamma_{\mu\nu}\frac{\partial\varphi}{\partial x_\nu}. \tag{81}$$

---

78 As is related in (Kollros 1956, 278), Einstein allegedly turned to Grossmann at one point and said: "Grossmann, you have to got to help me, otherwise I am going mad" ("Grossmann, Du mußt mir helfen, sonst werd' ich verrückt!").

79 See (Bianchi 1910, secs. 22–24) or (Wright 1908, sec. 53).

80 Equations (79) and (80) are equivalent to (Bianchi 1910, sec. 23, eq. (8), and sec. 24, eq. (19)), respectively. The second Beltrami invariant is alluded to in the 1914 review article on the *Entwurf* theory. In sec. 8 of this paper, Einstein points out that the covariant divergence of the contravariant vector $g^{\mu\nu}\partial\varphi/\partial x_\nu$ is equal to "the well-known generalization of the Laplacian $\Delta\varphi$, $\Phi = \sum_{\mu\nu}\frac{1}{\sqrt{G}}\frac{\partial}{\partial x_\mu}\left(\sqrt{g}g^{\mu\nu}\frac{\partial\varphi}{\partial x_\nu}\right)$" ("die bekannte Verallgemeinerung des Laplaceschen $\Delta\varphi$ …;" Einstein 1914b, 1051–1052; see also Einstein 1916, 797). This generalization is nothing but the second Beltrami invariant. Following (Weyl 1918), Einstein also used the second Beltrami invariant, again without identifying it by name, in lectures in 1919 (see CPAE 7, Doc. 19, [p. 22], and Doc. 20, [p. 1]).

Next to this expression he wrote: “vector” (“Vektor”).[81]

Einstein now applied a differential operator to this vector. In this way, he obtained the second Beltrami invariant

$$\frac{1}{\sqrt{G}}\frac{\partial}{\partial x_\mu}\left(\gamma_{\mu\nu}\sqrt{G}\frac{\partial\varphi}{\partial x_\nu}\right). \tag{82}$$

Next to this expression, he wrote: “scalar” (“Skalar”).

Einstein then investigated a condition, under which expression (82) would reduce to the ordinary Laplacian acting on a scalar. He called this condition a “plausible hypothesis” (“Naheliegende Hypothese”):

$$\sum\frac{\partial}{\partial x_\nu}(\sqrt{G}\gamma_{\mu\nu}) = 0. \tag{83}$$

This condition determines so-called “isothermal coordinates,”[82] or, as they are now called, harmonic coordinates. The corresponding harmonic coordinate restriction came to play an important role later in the notebook in Einstein’s analysis of the Ricci tensor (see pp. 19L–20L and the discussion following equation (471) in sec. 5.4.1).

In the last four lines of p. 6L, Einstein rewrote equation (83) to eliminate the derivative of the determinant G, leaving only derivatives of components of the metric:

$$\sqrt{G}\frac{\partial\gamma_{\mu\nu}}{\partial x_\mu} + \gamma_{\mu\nu}\frac{1}{2\sqrt{G}}\frac{\partial G}{\partial x_\mu} = 0. \tag{84}$$

Dividing by $\sqrt{G}$, he arrived at:

$$\sum\left(\frac{\partial\gamma_{\mu\nu}}{\partial x_\mu} + \frac{1}{2}\gamma_{\mu\nu}\frac{1}{G}\frac{\partial G}{\partial x_\mu}\right) = 0. \tag{85}$$

He then rewrote the second term using that the components of the contravariant metric are defined as

$$\gamma_{\mu\nu} = \frac{\Gamma_{\mu\nu}}{G} \tag{86}$$

where $\Gamma_{\mu\nu}$ is the minor of the $\mu\nu$-component of the covariant metric. Einstein thus rewrote the second term of equation (85) as:

$$\frac{1}{G}\frac{\partial G}{\partial x_\mu} = \sum_{\rho\sigma}\frac{\partial g_{\rho\sigma}}{\partial x_\mu}\frac{\Gamma_{\rho\sigma}}{G} = \sum_{\rho\sigma}\frac{\partial g_{\rho\sigma}}{\partial x_\mu}\gamma_{\rho\sigma}. \tag{87}$$

---

81 He initially wrote and then deleted “associated vector” (“zugeordneter Vekt”). “Associated” stands for “covariant” in this context. Expression (81) is a contravariant vector. Cf. the use of the term “*zugeordnet*” on p. 5R (see the discussion following equation (74) in sec. 3) and on p. 6R (see the remarks following equation (88)).

82 See, e.g., (Bianchi 1910, sec. 36-37) or (Wright 1908, sec. 57).

Here the calculation breaks off. A possible explanation is that Einstein wanted to check whether the spatially flat static metric, which reappears on the next page, satisfies the "plausible hypothesis" (83). The first term of equation (85) clearly vanishes for this metric, but the second term does not.

6R On p. 6R Einstein made a fresh start, denoting the gradient of the scalar function introduced at the top of p. 6L by:

$$\alpha_\nu = \frac{\partial\varphi}{\partial x_\nu}. \tag{88}$$

This is a covariant vector as is indicated by the label "associated vector" ("zugeordneter Vektor") written next to it. Recall that on the facing page 6L, Einstein started to write and then deleted this same label next to expression (81), the contravariant form of this vector.

Writing the contravariant form as $a_\nu = \gamma_{\nu\mu}\alpha_\mu$, Einstein wrote down the second Beltrami invariant (cf. equation (82)) as:

$$\sum \frac{1}{\sqrt{G}}\frac{\partial}{\partial x_\nu}(\sqrt{G}\,a_\nu). \tag{89}$$

Einstein now replaced all covariant elements in this expression by their contravariant counterparts and vice versa. To this end, he defined $\xi_\mu = \Sigma g_{\mu\nu}x_\nu$ (with inverse $x_\nu = \Sigma\gamma_{\mu\nu}\xi_\mu$).[83] He thus arrived at:

$$\sum \frac{1}{\sqrt{\Gamma}}\frac{\partial}{\partial \xi_\nu}(\sqrt{\Gamma}\,\alpha_\nu), \tag{90}$$

where Γ is the determinant of the contravariant metric and

$$\frac{\partial}{\partial\xi_\nu} = \sum \gamma_{\nu\sigma}\frac{\partial}{\partial x_\sigma}. \tag{91}$$

If Einstein thought that in going from the second Beltrami invariant (89) to the expression (90) he had constructed another scalar invariant, he was wrong. But he proceeded to rewrite expression (90) in terms of $\partial/\partial x_\nu$ and G in order to compare it with the second Beltrami invariant:

$$\sum \sqrt{G}\gamma_{\nu\sigma}\frac{\partial}{\partial x_\sigma}\left(\frac{1}{\sqrt{G}}\frac{\partial\varphi}{\partial x_\nu}\right). \tag{92}$$

Next to this expression, he noted that it is a "scalar" ("Skalar"). It is not.

On the next line, Einstein expanded expression (92) to:

---

83 Note that in these definitions, Einstein used finite coordinates x_ν instead of infinitesimals dx_ν, as he had done on p. 39L (cf. equations (6) and (7)). It seems doubtful whether, at this time, Einstein realized that only coordinate differentials are proper vectors, and that these formal definitions have no invariant significance for finite coordinates.

$$\sum_{\nu\sigma} \gamma_{\nu\sigma} \frac{\partial^2 \varphi}{\partial x_\nu \partial x_\sigma} + \gamma_{\nu\sigma} \frac{\partial \varphi}{\partial x_\nu} \frac{1}{2G} \frac{\partial G}{\partial x_\sigma} \quad , \tag{93}$$

losing a minus sign in the differentiation of $1/\sqrt{G}$. The plus sign should thus be a minus sign.

To compare this expression with the second Beltrami invariant, he expanded expression (82) on p. 6L for the latter to

$$\sum_{\nu\sigma} \gamma_{\nu\sigma} \frac{\partial^2 \varphi}{\partial x_\nu \partial x_\sigma} + \gamma_{\nu\sigma} \frac{\partial \varphi}{\partial x_\nu} \frac{1}{2G} \frac{\partial G}{\partial x_\sigma} + \frac{\partial \varphi}{\partial x_\nu} \frac{\partial \gamma_{\sigma\nu}}{\partial x_\sigma} \quad . \tag{94}$$

In the notebook there is actually a line connecting expression (82) on p. 6L to expression (94) on p. 6R.

Comparing the expressions (93) and (94), Einstein noted: "Should there only be *one* such scalar, it has to be the case that

$$\sum_{\sigma} \frac{\partial \gamma_{\sigma\nu}}{\partial x_\sigma} = 0 \text{''} \tag{95}$$

("Soll es nur *einen* derartigen Skalar geben so muss $\sum_{\sigma} \partial \gamma_{\sigma\nu} / \partial x_\sigma = 0$"). Recall, however, that expression (90) is *not* a scalar and that there is a sign error in the subsequent expression (93) for this alleged scalar. Einstein did not discover these errors until the top of p. 7L. On p. 6R, he was under the impression that the second Beltrami invariant and expression (92), which he took to be a scalar, were identical once condition (95), which we shall call the "Hertz restriction,"[84] was imposed. Adding this condition, he now proceeded as if he had two expressions—the second Beltrami invariant (89) and quantity constructed out of it in expression (92)—for one and the same scalar invariant.

To turn this scalar invariant into a candidate for the left-hand side of the field equations, Einstein substituted $\sqrt{G}$ for the scalar φ in expression (92). The resulting field equations will be invariant under unimodular transformations since $\sqrt{G}$ transforms as a scalar under such transformations. In a separate box to the right of equations (92)–(94), Einstein wrote down how G and Γ transform under a coordinate transformation with determinant P:[85]

$$\begin{aligned} G' &= P^2 G, \\ \Gamma' &= \frac{1}{P^2} \Gamma. \end{aligned} \tag{96}$$

If $P = 1$, i.e., for unimodular transformations, G and Γ are indeed scalars. At the bottom of p. 6L, Einstein had just gone through the derivation of another result for G:

84 The reason for naming it after Paul Hertz is explained in footnote 22.
85 This follows from the multiplication theorem for determinants.

$$\frac{1}{2\sqrt{G}}\frac{\partial\sqrt{G}}{\partial x_\nu} = \gamma_{\rho\sigma}\frac{\partial g_{\rho\sigma}}{\partial x_\nu}. \tag{97}$$

Inserting $\sqrt{G}$ for φ in expression (92) and using equation (97), Einstein arrived at the following candidate field equations:

$$\sum_{\mu\nu\rho\sigma}\gamma_{\mu\nu}\frac{\partial}{\partial x_\mu}\left(\gamma_{\rho\sigma}\frac{\partial g_{\rho\sigma}}{\partial x_\nu}\right) = 0 \tag{98}$$

"or" ("oder"), equivalently,[86]

$$\sum_{\mu\nu\rho\sigma}\gamma_{\mu\nu}\frac{\partial}{\partial x_\mu}\left(g_{\rho\sigma}\frac{\partial \gamma_{\rho\sigma}}{\partial x_\nu}\right) = 0\,. \tag{99}$$

Note that Einstein omitted a factor $\sqrt{G}/2$ in the expressions on the left-hand sides of both these equations. These expressions, however, retain the transformation behavior under unimodular transformations of the supposedly-invariant expression (92) from which Einstein constructed them. Since $\sqrt{G}$ is a scalar under unimodular transformation, a scalar divided by $\sqrt{G}$ remains a scalar.

Einstein now turned to "special cases $\mu = \nu$ and $\rho = \sigma$" ("Spezialfälle $\mu = \nu$ $\rho = \sigma$") of diagonal metric tensors. He quickly focused on one such case, namely the spatially flat static metric first introduced on p. 39L. He accordingly corrected "special cases" ("Spezialfälle") to "special case" ("Spezialfall") and wrote down the diagonal components of the static metric:[87]

$$\begin{aligned} &g_{11} = g_{22} = g_{33} = -1 \quad g_{44} = c^2 \\ &\gamma_{11} = \gamma_{22} = \gamma_{33} = +1 \;\; \gamma_{44} = -\frac{1}{c^2}, \end{aligned} \tag{100}$$

as well as its determinant $G = -c^2$, which he could read off from the matrix $\mathrm{diag}(-1, -1, -1, c^2)$ written directly above this expression.

Einstein now inserted the static metric into equations (98) and (99) for the index combinations $\rho = \sigma = 4$ and $\mu = \nu = 1, 2, 3$, the only combinations giving a non-vanishing contribution. He found

$$\sum_\nu\frac{\partial}{\partial x_\nu}\left(+\frac{1}{c^2}\frac{\partial c^2}{\partial x_\nu}\right) = 0\,, \tag{101}$$

and

86 Note that $0 = \frac{\partial}{\partial x_\nu}(g_{\rho\sigma}\gamma_{\rho\sigma}) = \gamma_{\rho\sigma}\frac{\partial g_{\rho\sigma}}{\partial x_\nu} + g_{\rho\sigma}\frac{\partial \gamma_{\rho\sigma}}{\partial x_\nu}$.

87 In the notebook, the "-1" seems to be corrected from "0." The contravariant components have the wrong sign. Einstein originally wrote them down correctly, but then changed the signs.

$$\sum \frac{\partial}{\partial x_\nu}\left(c^2 \frac{\partial}{\partial x_\nu}\frac{1}{c^2}\right) = 0\,, \tag{102}$$

"respectively" ("b[e]z[iehungs]w[eise]").[88]

Both equations are equivalent to[89]

$$\frac{\partial^2 \log c}{\partial x_\nu^2} = 0\,. \tag{103}$$

This equation has the form of the four-dimensional generalization of the Poisson equation and therefore looks promising as a component of candidate field equations.

Next to equation (103) wrote: "In this way not distinguishable" ("So nicht unterscheidbar."). Presumably, this refers to Einstein's query in the middle of the page concerning the relation between expression (94) for the second Beltrami invariant and expression (93) for what Einstein took to be an alternative scalar invariant. Since the spatially flat static metric satisfies condition (95), which reduces the former expression to the latter, it is clear this special case cannot be used to distinguish the two.

At the top of p. 7L, Einstein returned to the more general considerations above the horizontal line on p. 6R, where he had written down expression (93) for what he thought was an alternative scalar invariant. In fact, it is not a scalar and contains a sign error. Einstein had compared this expression to expression (94) for the second Beltrami invariant and had noticed that the former reduces to the latter if the Hertz condition (95) is imposed. 7L

Einstein began his considerations on p. 7L by writing down an expression that can be obtained from the second Beltrami invariant (82) on p. 6L by imposing the Hertz condition (95):[90]

$$\frac{\gamma_{\mu\nu}}{\sqrt{G}}\frac{\partial}{\partial x_\mu}\left(\sqrt{G}\frac{\partial\varphi}{\partial x_\nu}\right) \tag{104}$$

and wrote next to it that this is a "scalar" ("Skalar"). In the next two lines, Einstein tried to construct a vector out of the scalar quantities he had formed on the preceding two pages.

---

88 Inserting the components (100) of the static metric (with or without correcting the sign error noted in the preceding note) into the left-hand sides of equations (98) and (99), one finds the left-hand sides of equations (101) and (102) with the opposite sign. Probably, Einstein wrote down equation (101) using $\gamma_{11} = \gamma_{22} = \gamma_{33} = -1$ and $\gamma_{44} = 1/c^2$, then adjusted the result once he changed the sign of $\gamma_{\mu\nu}$.

89 For equations (101) and (102), one finds $\frac{\partial}{\partial x_\nu}\left(\frac{1}{c^2}\frac{\partial c^2}{\partial x_\nu}\right) = \frac{\partial}{\partial x_\nu}\left(\frac{2}{c}\frac{\partial c}{\partial x_\nu}\right) = 2\frac{\partial^2 \log c}{\partial x_\nu^2} = 0$ and $\frac{\partial}{\partial x_\nu}\left(c^2\frac{\partial}{\partial x_\nu}\frac{1}{c^2}\right) = \frac{\partial}{\partial x_\nu}\left(-\frac{2}{c}\frac{\partial c}{\partial x_\nu}\right) = -2\frac{\partial^2 \log c}{\partial x_\nu^2} = 0$, respectively.

90 The labeling of the indices suggests that this expression was obtained directly from expression (82) on p. 6L.

On the assumption that Einstein meanwhile caught the sign error in expression (93), one can readily understand these two lines: "If one forms [the second Beltrami invariant] $\Delta_2\varphi$ in two ways, it follows that

$$\sum_\nu \frac{1}{G}\frac{\partial}{\partial x_\nu}(\gamma_{\mu\nu}G) \tag{105}$$

[is] a vector" ("Bildet man $\Delta_2\varphi$ auf zwei Arten, so folgt $\sum_\nu \frac{1}{G}\frac{\partial}{\partial x_\nu}(\gamma_{\mu\nu}G)$ ein Vektor"). The conclusion was subsequently deleted.

If the error in expression (93) is corrected, then the difference between expressions (93) and (94) becomes:

$$\frac{\partial\varphi}{\partial x_\nu}\left(\gamma_{\nu\sigma}\frac{1}{G}\frac{\partial G}{\partial x_\sigma} + \frac{\partial\gamma_{\sigma\nu}}{\partial x_\sigma}\right) = \frac{\partial\varphi}{\partial x_\nu}\left(\frac{1}{G}\frac{\partial}{\partial x_\sigma}(\gamma_{\nu\sigma}G)\right). \tag{106}$$

Except for the labeling of its indices, the term in parentheses on the right-hand side is exactly equal to Einstein's expression (105) above. This explains why Einstein initially expected expression (105) to be a vector. Since expression (105) contracted with the covariant vector $\partial\varphi/\partial x_i$ (for arbitrary φ) is the difference between (what Einstein took to be) two scalars, the contraction must also be a scalar and expression (105) itself must be a contravariant vector. However, Einstein presumably recognized that expression (105) is in fact not a vector. The expression is equal to

$$\frac{1}{G}\frac{\partial\gamma_{\mu\nu}}{\partial x_\nu} + \frac{1}{G}\frac{\partial G}{\partial x_\nu}. \tag{107}$$

The second term is a vector. For the entire expression to be a vector, the first term would have to be a vector as well. This, however, is not the case, as Einstein presumably knew. One can thus understand why he eventually deleted the claim that expression (105) is a vector. This immediately told Einstein that the starting point of this entire line of reasoning, the assumption that expression (90) is a scalar, had to be mistaken.

The reconstruction given above leaves one question unanswered: how did Einstein discover the sign error in expression (93)? There is a plausible answer to this question. Suppose Einstein went through the same argument that we just described before he corrected this sign error. He would then have arrived at the conclusion that

$$\frac{\partial\varphi}{\partial x_\nu}\frac{\partial\gamma_{\nu\sigma}}{\partial x_\sigma} \tag{108}$$

is a scalar and that

$$\frac{\partial\gamma_{\nu\sigma}}{\partial x_\sigma} \tag{109}$$

therefore has to be a vector. Einstein already knew this to be false (see p. 39R). This then might well have alerted him to the sign error in expression (93). He may then have repeated the construction of a vector out of the difference between his two scalars (82) and (92), using the corrected version of the former.[91]

4.3 *Investigating the Core Operator (7L–8R)* 7L–8R

On pp. 6L–6R Einstein had tried to construct field equations for the metric tensor using as his starting point a familiar object from differential geometry, the generally-covariant second Beltrami invariant. The calculations documented on these pages had not produced any promising results. On pp. 7L–8R, he tried a different approach inspired by physical rather than mathematical considerations. He generalized the Laplace operator acting on the scalar potential of Newtonian gravitational theory to an operator acting on the tensorial potential $\gamma_{\mu\nu}$. We shall call this object the "core operator." It is a combination of two simpler operators, explicitly called "divergence" ("Divergenz" [p 7R]) and "exterior derivative" ("Erweiterung" [p. 8L]) in the notebook.[92] Einstein first examined the covariance properties of the core operator as a whole (pp. 7L-R). He then switched to generalizing the two constituent operators. The aim of both investigations was to see whether the core operator would provide him with a basis for extending the relativity principle from uniform to non-uniform motion.

One can discern two strategies with which Einstein tried to achieve his aim. The first strategy was to find a special type of non-linear coordinate transformations under which the core operator transforms as a tensor. The second strategy was to generalize the core operator to a differential operator that transforms as a tensor under ordinary non-linear transformations.

We introduce some special notation to facilitate a concise discussion of these two strategies. Consider an object, $\mathbf{O}(\partial, g)$ —constructed out of the metric, $g_{\mu\nu}$, and the derivative operator, $\partial / \partial x_\alpha$ —that transforms as a tensor under arbitrary *linear* transformations, $T_{\text{linear}} : x \rightarrow x'$. The transformation law for $\mathbf{O}$ under the inverse transformation, T_{linear}^{-1}, can schematically be written as:

$$\mathbf{O}(\partial, g) = T_{\text{linear}}^{-1} \mathbf{O}(\partial', g') . \tag{110}$$

If one now expresses ∂' and g' on the right-hand side in terms of ∂ and g with the help of T_{linear}, one, of course, just reproduces $\mathbf{O}$ in unprimed coordinates:

$$T_{\text{linear}}^{-1} \mathbf{O}(T_{\text{linear}}(\partial), T_{\text{linear}}(g)) = \mathbf{O}(\partial, g) . \tag{111}$$

---

91 It remains unclear exactly what Einstein took the relation between expressions (82) and (92) to be. His comment on p. 6R ("Should there only be one such scalar, it has to be the case that …") suggests that he thought of them as two different quantities that coincide only if an additional condition is imposed. However, his comment on p. 7L ("If one forms $\Delta_2\varphi$ in two ways …") suggests that he thought of them as different expressions of the same quantity. In that case, however, expression (90) should vanish, whereas Einstein only says that it is a vector.

92 Cf. (Einstein and Grossmann 1913, Part I, sec. 5).

Einstein's strategy for dealing with arbitrary *non-linear* transformations, T: $x \rightarrow x'$, involves two steps analogous to the two steps above. In general, $\mathbf{O}$ will *not* transform as a tensor under T. The first step is to introduce the object, $\tilde{\mathbf{O}}(\partial, g)$, that one obtains when applying the inverse transformation T^{-1} to $\mathbf{O}(\partial', g')$ *as if* $\mathbf{O}$ *does transform as a tensor under* T, more specifically, as if $\mathbf{O}$ transforms under T the same way it transforms under T_{linear}:

$$\tilde{\mathbf{O}}(\partial, g) \equiv \tilde{T}^{-1}\mathbf{O}(\partial', g') \ . \tag{112}$$

The tilde on $\tilde{T}^{-1}$ merely indicates this special use of the transformation rules. The second step is to express ∂' and g' in terms of ∂ and g with the help of T. Since in general $\mathbf{O}$ does not transform as a tensor under T, this operation will in general not just reproduce $\mathbf{O}$. In addition to $\mathbf{O}$, it will produce a (sum of) term(s), $\mathbf{C}$, constructed out of ∂, g, and the transformation matrices p and π for T and T^{-1}:[93]

$$\tilde{T}^{-1}\mathbf{O}(T(\partial), T(g)) = \mathbf{O}(\partial, g) + \mathbf{C}(\partial, g, p, \pi) \tag{113}$$

This two-step procedure is common to both strategies distinguished above. The two strategies differ in the way they make use of equation (113).

In the first strategy, Einstein uses equations such as equation (113) to read off the condition on the transformation matrices p and π for T that needs to be satisfied for $\mathbf{O}$ to transform as a tensor under T. This will be the case if $\tilde{\mathbf{O}}(\partial, g) = \mathbf{O}(\partial, g)$, i.e., in view of equations (112)–(113), if

$$\mathbf{C}(\partial, g, p, \pi) = 0 \, . \tag{114}$$

This condition gives a set of differential equations involving the transformation matrices, the components of the metric, and derivatives of both. Inserting a specific metric into equation (114) and solving the resulting equation for the transformation matrices, $p(g)$ and $\pi(g)$, one arrives at a special type of coordinate transformations. In the case of ordinary coordinate transformations, the transformation matrices are functions only of the coordinates. In the case of these special coordinate transformations, however, they depend both on the coordinates and on the metric. Following a suggestion by Paul Ehrenfest, Einstein later introduced the term "non-autonomous transformations" ("unselbständige Transformationen") for such transformations.[94] The transformation rule for such non-autonomous transformations can be written as

$$\mathbf{O}(\partial', g') = T(g)\mathbf{O}(\partial, g) , \tag{115}$$

where the matrices p and π for the transformation $T(g)$ must satisfy a condition for non-autonomous transformations of the form of equation (114).

In the case of the second strategy, Einstein looked upon the right-hand side of equation (113) as a generally-covariant expression that in the special case of a diagonal Minkowski metric reduces to the object $\mathbf{O}$ he started from. One can then set $g'_{\mu\nu} = \mathrm{diag}(-1, -1, -1, 1)$ in the transformation laws $g_{\mu\nu} = p_{\alpha\mu}p_{\beta\nu}g'_{\alpha\beta}$ and $\gamma_{\mu\nu} = \pi_{\alpha\mu}\pi_{\beta\nu}\gamma'_{\alpha\beta}$ and express the transformation matrices p and π in terms of $g_{\mu\nu}$

93 For the definition of these transformation matrices, see equations (119)–(122) below.

and its derivatives. Substituting the resulting expressions $p(g)$ and $\pi(g)$ into $\mathbf{C}$ in equation (113), one arrives at a new expression $\mathbf{D}$ that depends only on the metric and its derivatives:

$$\mathbf{D}(\partial, g) \equiv \mathbf{C}(\partial, g, p(g), \pi(g)) . \tag{116}$$

One can now define the generalization $\mathbf{O}_{\text{gen}}$ of the original object $\mathbf{O}$:

$$\mathbf{O}_{\text{gen}}(\partial, g) \equiv \mathbf{O}(\partial, g) + \mathbf{D}(\partial, g) . \tag{117}$$

$\mathbf{O}_{\text{gen}}$ reduces to $\mathbf{O}$ in the special case of the Minkowski metric in pseudo-Cartesian coordinates. The construction of $\mathbf{O}_{\text{gen}}$ only guarantees that $\mathbf{O}_{\text{gen}}$ transforms as a tensor under arbitrary transformations in Minkowski spacetime. Einstein, however, expected and made an attempt to prove (at the top of p. 8L), that $\mathbf{O}_{\text{gen}}$ would transform as a tensor under arbitrary transformations for any metric field.

Einstein applied the first of the two strategies described above to find non-autonomous transformation under which the core operator as a whole transforms as a tensor. He then applied the second strategy to generalize the two constituent operators of the core operator and thereby the core operator itself to expressions that transform as tensors under arbitrary transformations. He did not see these calculations through to the end. He came to realize that the generalized operators produced by this strategy degenerate when applied to the metric tensor. Einstein thereupon abandoned this second strategy altogether. The first strategy and the concept of non-autonomous transformations, however, continue to play an important role in the notebook.

4.3.1 *Covariance of the Core Operator under Non-autonomous Transformations (7L–R)*

Underneath the horizontal line on p. 7L, Einstein wrote down the core operator: 7L

---

94 See Einstein to H. A. Lorentz, 14 August 1913 (CPAE 5, Doc. 467): "One can consider two fundamentally different possibilities. 1) Transformations which are independent of the existing $g_{\mu\nu}$-field, which Ehrenfest designated as 'autonomous transformations;' according to my knowledge group theory has only dealt with this kind of transformations. 2) Transformations whose [matrices] would have to be determined by differential equations for the $g_{\mu\nu}$-field considered as given, which hence have to be adapted to the existing $g_{\mu\nu}$-field. Such transformations have—as far as I know—not yet been systematically studied. ('non-autonomous transformations')" ("Zwei Möglichkeiten prinzipiell verschiedener Art kommen da in Betracht. 1) Transformationen, welche von dem vorhandenen $g_{\mu\nu}$-Feld unabhängig sind, welche Ehrenfest als 'selbständige Transformationen' bezeichnete; nur mit solchen hat sich meines Wissens bisher die Gruppentheorie beschäftigt. 2) Transformationen, deren [...] erst durch Differentialgleichungen zu dem als gegebenen zu betrachtenden $g_{\mu\nu}$-Feld zu bestimmen wären, die also dem vorhandenen $g_{\mu\nu}$-Feld angepasst werden müssen. Solche Transformationen sind—soviel ich weiss—noch nicht systematisch untersucht worden. (‚unselbständige Transformationen')"). For further discussion of non-autonomous transformations—or, as Einstein later called them "justified" ("berechtigte") transformations between "adapted" ("angepaßte") coordinates" (Einstein and Grossmann 1914, 221; Einstein 1914b, 1070)—see "Untying the Knot ..." sec. 3.3 (in this volume). For a modern discussion of such transformations, see (Bergmann and Komar 1972).

$$\sum \frac{\partial}{\partial x_\mu}\left(\gamma_{\mu\nu}\frac{\partial \gamma_{i\kappa}}{\partial x_\nu}\right). \tag{118}$$

For weak static fields represented by a metric tensor of the form $\mathrm{diag}(-1, -1, -1, c^2(x, y, z))$ the 44-component of this operator reduces to minus the Laplacian acting on c^2, the square of the gravitational potential of Einstein's 1912 theory.

On pp. 7L–8R, Einstein studied the transformation properties of the core operator using transformation matrices $p_{\mu\nu}$ and $\pi_{\mu\nu}$, which are defined as follows. Under a transformation from coordinates x_μ to x'_μ, a contravariant vector a_μ transforms as

$$a'_\mu = \sum p_{\mu\nu} a_\nu \quad \text{with } p_{\mu\nu} \equiv \frac{\partial x'_\mu}{\partial x_\nu}, \tag{119}$$

while a covariant vector transforms as[95]

$$\alpha'_\mu = \sum \pi_{\mu\nu} \alpha_\nu \quad \text{with } \pi_{\mu\nu} \equiv \frac{\partial x_\nu}{\partial x'_\mu}. \tag{120}$$

The inverse transformation of a contravariant vector is given by:

$$a_\mu = \sum \pi_{\nu\mu} a'_\nu \text{ with } \pi_{\nu\mu} = \frac{\partial x_\mu}{\partial x'_\nu}; \tag{121}$$

the inverse transformation of a covariant vector by:

$$\alpha_\mu = \sum p_{\nu\mu} \alpha'_\nu \text{ with } p_{\nu\mu} \equiv \frac{\partial x'_\nu}{\partial x_\mu}. \tag{122}$$

It follows that

$$\sum p_{\mu\alpha}\pi_{\nu\alpha} = \delta_{\mu\nu} \tag{123}$$

and that

$$\sum p_{\alpha\mu}\pi_{\alpha\nu} = \delta_{\mu\nu}, \tag{124}$$

where we availed ourselves of the Kronecker delta, which Einstein does not use in the notebook.

Einstein considered the transformation of the core operator from x'_μ to x_μ. Substituting $\gamma'_{\mu\nu} = p_{\mu\alpha} p_{\nu\beta} \gamma_{\alpha\beta}$ and $\partial/\partial x'_\mu = \pi_{\mu\alpha}(\partial/\partial x_\alpha)$ into[96]

---

95 Notice that the definitions of $p_{\mu\nu}$ and $\pi_{\mu\nu}$ differ from the definitions of these quantities in (Einstein and Grossmann 1913, 24), where they are defined as $p_{\mu\nu} \equiv \partial x_\mu / \partial x'_\nu$ and $\pi_{\mu\nu} \equiv \partial x'_\nu / \partial x_\mu$. In other words, the roles of p and π are interchanged as are the indices μ and ν. This is related to the fact that in the Zurich Notebook contravariant quantities (with some exceptions such as the contravariant components $\gamma_{\mu\nu}$ of the metric) are generally denoted by Latin letters and covariant quantities by Greek ones, whereas in (Einstein and Grossmann 1913) this is just the other way around.

96 Expression (125) is a concrete example of $\mathbf{O}(\partial', g')$ introduced in equation (112).

$$\sum_{\mu\nu} \frac{\partial}{\partial x'_\mu}\left(\gamma'_{\mu\nu}\frac{\partial \gamma'_{i\kappa}}{\partial x'_\nu}\right), \tag{125}$$

one arrives at the equation written directly under expression (118) for the core operator in the notebook:[97]

$$\sum_{\alpha\beta\mu\nu\rho\sigma} \pi_{\mu\alpha}\frac{\partial}{\partial x_\alpha}\left(p_{\mu\rho}p_{\nu\sigma}\gamma_{\rho\sigma}\pi_{\nu\beta}\frac{\partial}{\partial x_\beta}(p_{il}p_{\kappa m}\gamma_{lm})\right). \tag{126}$$

The contraction of $p_{\nu\sigma}$ and $\pi_{\nu\beta}$ gives $\delta_{\beta\sigma}$. Einstein also set

$$\pi_{\mu\alpha}\frac{\partial}{\partial x_\alpha}(p_{\mu\rho}\ldots) = \frac{\partial}{\partial x_\rho}(\ldots), \tag{127}$$

tacitly making the erroneous assumption that $\pi_{\mu\alpha}\partial p_{\mu\rho}/\partial x_\alpha = 0$, possibly on the basis of the following incorrect application of the chain rule:[98]

$$\pi_{\mu\alpha}\frac{\partial p_{\mu\rho}}{\partial x_\alpha} = \frac{\partial x_\alpha}{\partial x'_\mu}\,\frac{\partial^2 x'_\mu}{\partial x_\alpha \partial x_\rho} = \frac{\partial^2 x'_\mu}{\partial x'_\mu \partial x_\rho} = 0. \tag{128}$$

With these simplifications, Einstein arrived at the following expression for the core operator in x'_μ-coordinates in terms of quantities in x_μ-coordinates:[99]

$$\sum_{\mu\nu} \frac{\partial}{\partial x'_\mu}\left(\gamma'_{\mu\nu}\frac{\partial \gamma'_{i\kappa}}{\partial x'_\nu}\right) = \sum \frac{\partial}{\partial x_\rho}\left(\gamma_{\rho\sigma}\frac{\partial}{\partial x_\sigma}(p_{il}p_{\kappa m}\gamma_{lm})\right). \tag{129}$$

To further investigate the covariance properties of the core operator, Einstein adopted the two-step procedure described in the introduction of sec. 4.3. The first step is to write down the law according to which the core operator would transform if it transformed as a tensor under the transformation from x'^μ to x^μ. Adopting Einstein's notation

$$(\gamma)_{\alpha\beta} \equiv \sum \frac{\partial}{\partial x_\mu}\left(\gamma_{\mu\nu}\frac{\partial \gamma_{\alpha\beta}}{\partial x_\nu}\right) \tag{130}$$

for the core operator, one can write this tensorial transformation law as[100]

$$(\gamma)_{\alpha\beta} = \sum \pi_{i\alpha}\pi_{\kappa\beta}(\gamma)'_{i\kappa}. \tag{131}$$

The second step of the procedure is to rewrite $(\gamma)'_{i\kappa}$, the core operator in the primed coordinate system, in terms of quantities in the unprimed coordinate system. Inserting

---

97 There is also a summation over m and l.

98 This error was committed repeatedly by Einstein. See, e.g., pp. 7R, 8L. It was eventually discovered on p. 10L (see sec. 4.5.1 below).

99 The notebook has $\partial/\partial\rho$ and $\partial/\partial\sigma$ instead of $\partial/\partial x_\rho$ and $\partial/\partial x_\sigma$.

100 Equations (131) and (132) form a concrete example of a combination of equations (112) and (113): $\tilde{\mathbf{O}}(\partial, g) \equiv \tilde{T}^{-1}\mathbf{O}(\partial', g') = \tilde{T}^{-1}\mathbf{O}(T(\partial), T(g))$.

(the erroneous) expression (129) for $(\gamma)'_{i\kappa}$ found earlier, one arrives at the equation given in the notebook at this point:

$$(\gamma)_{\alpha\beta} = \sum \pi_{i\alpha}\pi_{\kappa\beta}\frac{\partial}{\partial x_\rho}\left(\gamma_{\rho\sigma}\frac{\partial}{\partial x_\sigma}(p_{il}p_{\kappa m}\gamma_{lm})\right). \tag{132}$$

If $(\gamma)_{\alpha\beta}$ transforms as a tensor under the transformation from x^μ to x'^μ and vice versa, the right-hand side of this equation should be equal to the core operator

$$\sum \frac{\partial}{\partial x_\rho}\left(\gamma_{\rho\sigma}\frac{\partial \gamma_{\alpha\beta}}{\partial x_\sigma}\right). \tag{133}$$

This is trivially true for linear transformations.[101] Einstein wanted to find more general transformations for which the right-hand side of equation (132) reduces to expression (133). For such transformations the sum of all terms on the right-hand side of equation (132) that involve derivatives of the components of the transformation matrix $p_{\mu\nu}$ should vanish. Einstein set out to collect such terms.

First, he considered the differential operator $\partial/\partial x_\sigma$ in front of the innermost set of parentheses on the right-hand side of equation (132). He rewrote this part of the equation as:

$$p_{il}p_{\kappa m}\frac{\partial \gamma_{lm}}{\partial x_\sigma} + \gamma_{lm}\frac{\partial}{\partial x_\sigma}(p_{il}p_{\kappa m}). \tag{134}$$

Then he turned to the differential operator $\partial/\partial x_\rho$ in front of the outermost set of parentheses on the right-hand side of equation (132). First, using equation (134), he wrote down the term that comes from having $\partial/\partial x_\rho$ act on $\gamma_{\rho\sigma}$:

$$\frac{\partial \gamma_{\rho\sigma}}{\partial x_\rho}\left(p_{il}p_{\kappa m}\frac{\partial \gamma_{lm}}{\partial x_\sigma} + \gamma_{lm}\frac{\partial}{\partial x_\sigma}(p_{il}p_{\kappa m})\right). \tag{135}$$

Finally, he wrote down the four terms that result from applying $\partial/\partial x_\rho$ to expression (134) and contracting it with $\gamma_{\rho\sigma}$:

$$\begin{aligned}\gamma_{\rho\sigma}p_{il}p_{\kappa m}\frac{\partial^2 \gamma_{lm}}{\partial x_\rho \partial x_\sigma} + \gamma_{\rho\sigma}\frac{\partial \gamma_{lm}}{\partial x_\sigma}\frac{\partial}{\partial x_\rho}(p_{il}p_{\kappa m}) + \gamma_{\rho\sigma}\frac{\partial \gamma_{lm}}{\partial x_\rho}\frac{\partial}{\partial x_\sigma}(p_{il}p_{\kappa m}) \\ + \gamma_{\rho\sigma}\gamma_{lm}\frac{\partial^2}{\partial x_\rho \partial x_\sigma}(p_{il}p_{\kappa m})\end{aligned} \tag{136}$$

The core operator $(\gamma)_{\alpha\beta}$ transforms as a tensor under transformations with transformation matrices $p_{\mu\nu}$ if the components of the transformation matrices satisfy the con-

101 For linear transformations, equation (132) reduces to:

$$(\gamma)_{\alpha\beta} = \sum \pi_{i\alpha}\pi_{\kappa\beta}p_{il}p_{\kappa m}\frac{\partial}{\partial x_\rho}(\gamma_{\rho\sigma}\frac{\partial \gamma_{lm}}{\partial x_\sigma}) = \sum \delta_{\alpha l}\delta_{\beta m}\frac{\partial}{\partial x_\rho}(\gamma_{\rho\sigma}\frac{\partial \gamma_{lm}}{\partial x_\sigma}) = \sum \frac{\partial}{\partial x_\rho}\left(\gamma_{\rho\sigma}\frac{\partial \gamma_{\alpha\beta}}{\partial x_\sigma}\right).$$

dition that the sum of all terms in expressions (135) and (136) that involve derivatives of $p_{\mu\nu}$ vanish for a given metric field with contravariant components $\gamma_{\mu\nu}$. Collecting such terms, one arrives at the condition:[102]

$$\begin{aligned}&\frac{\partial\gamma_{\rho\sigma}}{\partial x_\rho}\left(\gamma_{lm}\frac{\partial}{\partial x_\sigma}(p_{il}p_{\kappa m})\right)+\gamma_{\rho\sigma}\frac{\partial\gamma_{lm}}{\partial x_\sigma}\frac{\partial}{\partial x_\rho}(p_{il}p_{\kappa m})\\&\gamma_{\rho\sigma}\frac{\partial\gamma_{lm}}{\partial x_\rho}\frac{\partial}{\partial x_\sigma}(p_{il}p_{\kappa m})+\gamma_{\rho\sigma}\gamma_{lm}\frac{\partial^2}{\partial x_\rho\partial x_\sigma}(p_{il}p_{\kappa m}) = 0\end{aligned}\tag{137}$$

This condition on the transformation matrices $p_{\mu\nu}$ is the condition for what Einstein would later call "non-autonomous" transformations under which the core operator transforms as a tensor (see the introduction to sec. 4.3).[103]

Looking at this condition, one readily sees that it will be satisfied if

$$\frac{\partial}{\partial x_\sigma}(p_{il}p_{\kappa m}) = 0\tag{138}$$

for all index combinations. Except for a meaningless summation sign, which Einstein seems to have left in by mistake, and a relabeling of the indices the left-hand side of this equation is just the expression at the top of p. 7R: 7R

$$\sum\frac{\partial}{\partial x_\sigma}(p_{\alpha l}p_{\beta m})\,.\tag{139}$$

This expression is introduced with the comment: "where p is differentiated at least once. $i = \alpha \quad \kappa = \beta$" ("wobei p mindestens einmal abgeleitet wird. $i = \alpha$ $\kappa = \beta$").[104] The restriction to terms in which "p is differentiated at least once" identifies those terms in expressions (135) and (136) on p. 7L that must vanish if the core operator is to transform as a tensor under the coordinate transformation described by the matrix $p_{\mu\nu}$. Unfortunately, condition (138) restricts the allowed transformations to linear transformations, whereas Einstein was looking for non-linear transformation under which the core operator transforms as a tensor.

Einstein now drew a horizontal line and made a fresh start. He wrote down the contraction of the core operator with the covariant metric:

$$\sum_{i\kappa\mu\nu} g_{i\kappa}\frac{\partial}{\partial x_\mu}\left(\gamma_{\mu\nu}\frac{\partial\gamma_{i\kappa}}{\partial x_\nu}\right)\tag{140}$$

---

102 This is a concrete example of condition (114) in the introduction to sec. 4.3.

103 This condition would still need to be corrected for the error made in going from equation (126) to equation (129).

104 To understand the change in the labeling of the indices—from i and k to α and β—note that the core operator (see equation (132)) is obtained by contracting the sum of expressions (135) and (136) with $\pi_{i\alpha}\pi_{\kappa\beta}$. In first-order approximation, $\pi_{i\alpha}$ and $\pi_{\kappa\beta}$ can be replaced by the Kronecker deltas $\delta_{i\alpha}$ and $\delta_{\kappa\beta}$.

writing "scalar" ("Skalar") next to it. The expression will transform as a scalar for transformations under which the core operator transforms as a tensor. Conversely, Einstein may have expected that transformations under which expression (140) transforms as a scalar are transformations under which the core operator transforms as a tensor. He may have felt that the former were easier to find than the latter.

On the next line, however, Einstein returned once more to the condition derived on p. 7L for the transformation matrices $p_{\mu\nu}$ and $\pi_{\mu\nu}$ (see equations (129)–(137)). We can only make sense of the expression Einstein wrote down if we assume that he now focused on infinitesimal transformations. Consider the right-hand of equation (132):

$$\sum \pi_{i\alpha}\pi_{\kappa\beta}\frac{\partial}{\partial x_\rho}\left(\gamma_{\rho\sigma}\frac{\partial}{\partial x_\sigma}(p_{il}p_{\kappa m}\gamma_{lm})\right) . \tag{141}$$

Following Einstein's notation on p. 10L (see equation (238)) for an infinitesimal transformation, $p_{il} = \delta_{il} + p^x_{il}$, and neglecting terms smaller than those of first order in p^x_{il}, one can write the product in the innermost parentheses in expression (141) as

$$(\delta_{il} + p^x_{il})(\delta_{\kappa m} + p^x_{\kappa m})\gamma_{lm} = \gamma_{i\kappa} + p^x_{il}\gamma_{l\kappa} + p^x_{\kappa m}\gamma_{im} \tag{142}$$

Inserting this last expression into expression (141) and collecting all terms involving derivatives of the transformation matrices, one finds, to first order,

$$\pi_{i\alpha}\pi_{\kappa\beta}\frac{\partial}{\partial x_\rho}\left(\gamma_{\rho\sigma}\frac{\partial}{\partial x_\sigma}(p^x_{il}\gamma_{l\kappa} + p^x_{\kappa m}\gamma_{im})\right) \tag{143}$$

Since only terms involving derivatives of p^x_{il} matter, $\pi_{i\alpha}$ and $\pi_{\kappa\beta}$ can be replaced by Kronecker deltas in first-order approximation. This then gives

$$\frac{\partial}{\partial x_\rho}\left(\gamma_{\rho\sigma}\frac{\partial}{\partial x_\sigma}(p^x_{\alpha l}\gamma_{l\beta} + p^x_{\beta l}\gamma_{\alpha l})\right) . \tag{144}$$

If the core operator transforms as a tensor under the transformation described by $p_{\mu\nu}$, the sum of all terms in expressions (144) that involve derivatives of the components of $p^x_{\mu\nu}$ must vanish (cf. the discussion following equation (136)). This condition can be satisfied by requiring that the sum of all terms in

$$\frac{\partial}{\partial x_\rho}\left(\gamma_{\rho\sigma}\frac{\partial}{\partial x_\sigma}(p^x_{\alpha l}\gamma_{l\beta})\right) \tag{145}$$

"in which p is differentiated at least once" ("wobei p mindestens einmal abgeleitet wird") vanish. Einstein wrote down expression (145)—albeit with $p_{\alpha l}$ rather than with $p^x_{\alpha l}$—with this remark next to it, which supports our reconstruction of the purpose behind it. The condition for infinitesimal transformations resulting from expression (145) is much simpler than condition (137) for finite transformations. Einstein's calculations nonetheless break off at this point.

4.3.2 *Generalizing the Constituent Parts of the Core Operator: Divergence and Exterior Derivative Operators (7R–8R)* 7R–8R

Rather than continuing the search for non-autonomous non-linear coordinate transformations under which the core operator transforms as a tensor, Einstein, on pp. 7R–8R, tried to find a generalization of the core operator that would transform as a tensor under ordinary non-linear transformations. In other words, he switched from the first to the second of the two strategies that we distinguished in the introduction of sec. 4.3. He applied this strategy to the two constituent components of the core operator, the divergence and the exterior derivative. On p. 7R, under the heading "Divergence of a tensor" ("Divergenz des Tensors"), he tried to generalize the divergence operator. On pp. 8L-R, under the heading "Exterior derivative of a tensor" ("Erweiterung des Tensors"), he tried to generalize the exterior derivative operator.

To generalize the divergence operator, Einstein started from the ordinary divergence of a second-rank tensor in primed pseudo-Cartesian coordinates on Minkowski spacetime and then wrote down the expression in unprimed arbitrary coordinates that one would get if the divergence in primed coordinates transformed as a tensor under this transformation. In the primed coordinates the Minkowski metric has the standard diagonal form. Since the components of this metric are constants, there will be a simple relation between the metric in unprimed coordinates and the matrices for the transformation between primed and unprimed coordinates. Using this relation, one can eliminate the transformation matrices from the expression for the divergence transformed from primed to unprimed coordinates as if it were a tensor. This expression will then be entirely in terms of components of the metric in unprimed coordinates and their derivatives. Einstein expected that this procedure would yield the generalized divergence operator that he had found on p. 5R (essentially the covariant divergence), but he was unable to prove this conjecture.

These considerations begin beneath the second horizontal line on p. 7R. Under the heading "Divergence of the Tensor" ("Divergenz des Tensors"), Einstein wrote down the ordinary divergence of a second-rank contravariant tensor in a primed coordinate system[105] 7R

$$a'_{\mu} = \sum \frac{\partial T'_{\mu\nu}}{\partial x'_{\nu}} \qquad \mu\text{-Vektor}\,. \tag{146}$$

In the line above this equation, Einstein characterized the primed coordinate system with the remark: "original system ($'$) shall have constant g, γ." ("Ursprüngliches System ($'$) habe konstante g, γ."). Presumably, what he had in mind was a pseudo-Cartesian coordinate system on Minkowski spacetime.

Einstein now applied the same two-step procedure that we encountered on p. 7L. First, he wrote down how a'_{μ} would transform if it were to transform as a vector under transformations from the special primed to arbitrary unprimed coordinates. He then took expression (146) for the vector in primed coordinates and used the standard trans-

105 The quantity a'_{μ} is a concrete example $\mathbf{O}(\partial', g')$ in equation (112).

formation rules to rewrite the building blocks of this vector in terms of their counterparts in the unprimed coordinate system. He thus arrived at the following expression for the vector a_μ "in the unprimed system" ("Im ungestrichenen System"):[106]

$$a_\sigma = \sum \pi_{\mu\sigma} a'_\mu = \sum \pi_{\mu\sigma} \pi_{\nu\tau} \frac{\partial}{\partial x_\tau} (p_{\mu i} p_{\nu\kappa} T_{i\kappa}) \,. \tag{147}$$

He simplified the right-hand side of equation (147), using the same erroneous relation,

$$\ldots \pi_{\nu\tau} \frac{\partial}{\partial x_\tau} (\ldots p_{\nu\kappa} \ldots) = \ldots \pi_{\nu\tau} p_{\nu\kappa} \frac{\partial}{\partial x_\tau} (\ldots) = \ldots \frac{\partial}{\partial x_\kappa} (\ldots) \,, \tag{148}$$

that he had used earlier (cf. equations (127)–(128)). In this way he obtained[107]

$$a_\sigma = \sum \pi_{\mu\sigma} \frac{\partial}{\partial x_\kappa} (p_{\mu i} T_{i\kappa}) = \sum \frac{\partial T_{\sigma\kappa}}{\partial x_\kappa} + \sum_{\mu i\kappa} T_{i\kappa} \pi_{\mu\sigma} \frac{\partial p_{\mu i}}{\partial x_\kappa} \,. \tag{149}$$

Einstein now indicated how he wanted to generalize the ordinary divergence operator: "This sum is to be expressed by the g resp. γ. In doing so one has to use the fact that the primed g and γ are constant." ("Diese Summe ist durch die g bzw. γ auszudrükken. Dabei ist zu benutzen, dass die gestrichenen g bzw. γ konstant sind."). In other words, using the simple form of the metric in the special primed coordinates, he wanted to express the components of the transformation matrices and their derivatives in the last term of equation (149) in terms of the components of the metric and their derivatives in the arbitrary unprimed coordinates. The final step producing the sought-after generalization of the divergence operator is to replace the components of the Minkowski metric in the unprimed coordinates in the resulting expression by components of an arbitrary metric.

Underneath the sentence explaining the aim of his calculation, indicating the connection to the last term of equation (149) by a vertical line, Einstein wrote:

$$\sum T_{i\kappa} \left(\frac{\partial g_{\sigma\kappa}}{\partial x_i} - \frac{1}{2} \frac{\partial g_{i\kappa}}{\partial x_\sigma} \right) . \tag{150}$$

He thus expected that a_σ could be written as:

$$a_\sigma = \sum \frac{\partial T_{\sigma\kappa}}{\partial x_\kappa} + \sum T_{i\kappa} \left(\frac{\partial g_{\sigma\kappa}}{\partial x_i} - \frac{1}{2} \frac{\partial g_{i\kappa}}{\partial x_\sigma} \right) . \tag{151}$$

---

106 Equation (147) is a concrete example of a combination of equations (112) and (113): $\tilde{\mathbf{O}}(\partial, g) \equiv \tilde{T}^{-1} \mathbf{O}(\partial', g') = \tilde{T}^{-1} \mathbf{O}(T(\partial), T(g))$.

107 Correcting Einstein's mistake in going from equation (147) to equation (149), one finds:

$$a_\sigma = \sum \pi_{\mu\sigma} \pi_{\nu\tau} \frac{\partial}{\partial x_\tau} (p_{\mu i} p_{\nu\kappa} T_{i\kappa}) = \frac{\partial T_{\sigma\tau}}{\partial x_\tau} + T_{i\tau} \pi_{\mu\sigma} \frac{\partial p_{\mu i}}{\partial x_\tau} + T_{\sigma\kappa} \pi_{\nu\tau} \frac{\partial p_{\nu\kappa}}{\partial x_\tau} \,.$$

This is a concrete example of equation (113). The first term on the right-hand side corresponds to $\mathbf{O}(\partial, g)$; the last two terms to $\mathbf{C}(\partial, g, p, \pi)$.

One sees immediately that this cannot be correct since the right-hand side is a sum of a contravariant and a covariant term.

Einstein's expectation derives from his experience with the covariant divergence of the stress-energy tensor on p. 5R. If in equation (71) found on p. 5R one sets $\sqrt{G} = 1$, its left-hand side reduces to:

$$\sum \frac{\partial}{\partial x_n}(g_{m\nu} T_{\nu n}) - \frac{1}{2}\sum \frac{\partial g_{\mu\nu}}{\partial x_m} T_{\mu\nu} . \tag{152}$$

Relabeling indices and using that the stress-energy tensor $T_{\mu\nu}$ is symmetric,[108] one can rewrite expression (152) as:

$$\sum \frac{\partial}{\partial x_i}(g_{\sigma\kappa} T_{\kappa i}) - \frac{1}{2}\sum \frac{\partial g_{i\kappa}}{\partial x_\sigma} T_{i\kappa} = \sum g_{\sigma\kappa} \frac{\partial T_{i\kappa}}{\partial x_i} + \sum T_{i\kappa}\left(\frac{\partial g_{\sigma\kappa}}{\partial x_i} - \frac{1}{2}\frac{\partial g_{i\kappa}}{\partial x_\sigma}\right) . \tag{153}$$

Comparing this last expression with Einstein's expectation for the form of a generalized divergence operator in equation (151), one easily recognizes the basic problem. He expected the generalized divergence of a contravariant tensor to be a contravariant vector, whereas the operator he had found on p. 5R turns a contravariant tensor into a covariant vector (see expression (152)).

Einstein did not recognize the problem at first and tried to show that equation (149) for a_σ can indeed be rewritten as equation (151). This can be inferred from the last two lines on p. 7R. He tried to show that

$$\frac{\partial g_{\sigma\kappa}}{\partial x_i} - \frac{1}{2}\frac{\partial g_{i\kappa}}{\partial x_\sigma} \tag{154}$$

in equation (151) is equal to[109]

$$\pi_{\mu\sigma}\frac{\partial p_{\mu i}}{\partial x_\kappa} \tag{155}$$

in equation (149). To this end he eliminated the metric in unprimed coordinates in expression (154) in favor of the metric in primed coordinates, whose components are constants:

$$\sum \frac{\partial}{\partial x_i}(p_{\mu\sigma} p_{\nu\kappa} g'_{\mu\nu}) - \frac{1}{2}\frac{\partial}{\partial x_\sigma}(p_{\mu i} p_{\nu\kappa} g'_{\mu\nu}) \quad . \tag{156}$$

Finally, he wrote down one contribution coming from the second term in expression (156), the one involving a derivative of $p_{\mu i}$, which also occurs in expression (155):

108 At the beginning of the calculation on p. 7R (see equation (146)), $T_{\mu\nu}$ was an arbitrary tensor. At this point in the calculation, however, it becomes essential that $T_{\mu\nu}$ is symmetric.

109 Since $\pi_{\mu\sigma} = \partial x_\sigma / \partial x'_\mu$, σ is a contravariant index. Expression (155) can thus never be equal to expression (154) in which σ is a covariant index.

$$\frac{1}{2}\frac{\partial p_{\mu i}}{\partial x_\sigma} p_{\nu\kappa} g'_{\mu\nu} . \tag{157}$$

At this point the calculation breaks off.

8L The method by which Einstein tried to generalize the ordinary divergence operator on p. 7R does not guarantee that the result will be a tensor. The main problem comes from the final step in which an expression derived for one special metric, the Minkowski metric, is assumed to transform as a tensor for an arbitrary metric (at least under unimodular transformations because of the restriction to $\sqrt{G} = 1$). The deleted calculation at the top of p. 8L may have been an attempt to prove that the method employed on p. 7R actually does produce an object that transforms as a tensor under arbitrary coordinate transformations. If that is indeed the purpose of this calculation, the strategy chosen by Einstein is clear. He tried to prove that a transformation from arbitrary unprimed coordinates to arbitrary double-primed coordinates can be decomposed into two transformations of the type considered on p. 7R, namely a transformation from the unprimed coordinates to special primed coordinates followed by a transformation from these special primed coordinates to the double-primed coordinates. Since any metric can locally be transformed to a Minkowski metric in standard diagonal form, the form of the metric in the special primed coordinates, this would guarantee that the method of p. 7R does indeed produce a tensor.

These considerations may lie behind Einstein's question at the top of p. 8L, "Do symmetrical transformations form a group?" ("Haben symmetrische Transformationen Gruppeneigenschaft?"), and behind the subsequent investigation of a transformation from unprimed to primed to double-primed coordinates. What remains unclear, however, is why Einstein focused on symmetric transformations.

To determine whether symmetric transformations given by

$$x'_\nu = \sum_\sigma p_{\nu\sigma} x_\sigma , \tag{158}$$

with $p_{\nu\sigma} = p_{\sigma\nu}$ form a group, Einstein considered the components

$$p''_{\lambda\sigma} = \sum p'_{\lambda\nu} p_{\nu\sigma} \tag{159}$$

of the matrix for the composite transformation

$$x''_\lambda = \sum_\nu p'_{\lambda\nu} x'_\nu = \sum_{\nu\sigma} p'_{\lambda\nu} p_{\nu\sigma} x_\sigma \tag{160}$$

and switched the indices of these components:

$$p''_{\sigma\lambda} = \sum_\nu p'_{\sigma\nu} p_{\nu\lambda} . \tag{161}$$

He did not pursue this calculation any further. Either he concluded (correctly) that symmetric transformations do not form a group or he did not see how to settle the question either way.

He now turned to rotational transformations, perhaps as an example of a class of transformations that certainly do form a group, or to check whether such transformations are symmetric. He wrote down transformation laws for a rotation over an angle α as well as for its inverse:

$$\begin{aligned} x' &= x\cos\alpha + y\sin\alpha \\ y' &= -x\cos\alpha + y\sin\alpha \\ x &= x'\cos\alpha - y'\sin\alpha \\ y &= x'\sin\alpha + y'\cos\alpha \end{aligned} \tag{162}$$

These expressions show that rotation does not belong to the class of symmetric transformations. Perhaps this is why he deleted the calculation at the top of p. 7R. Another possibility is that he realized that the restriction to symmetric transformations was not necessary to show that the method of p. 7R produces a tensor.

Under the heading "Exterior derivative of the Tensor" ("Erweiterung des Tensors"), Einstein now turned to the second differential operator relevant to generalizing the core operator. This takes up the remainder of p. 8L and the first two lines on p. 8R. As in the case of generalizing the divergence operator, Einstein's starting point is the ordinary exterior derivative of a contravariant second-rank tensor in special relativity. He wrote: "In ordinary space [i.e., Minkowski spacetime in pseudo-Cartesian coordinates][110]

$$\Delta'_{\mu\nu\rho} = \frac{\partial T'_{\mu\nu}}{\partial x'_{\rho}}, \tag{163}$$

is a tensor of three manifolds [i.e., of third rank]" ("Im gew. [gewöhnlichen] Raum ist … Tensor von 3 Mannigfaltigkeiten"). He then "introduced transformations with constant coefficients" ("Subst[itutionen] von konst. Koeffizienten eingeführt")

$$\Delta'_{\mu\nu\rho}[=]\sum\pi_{\rho\sigma}\frac{\partial}{\partial x_{\sigma}}(p_{\mu\alpha}p_{\nu\beta}T_{\alpha\beta}). \tag{164}$$

and confirmed that "for such transformations $\partial T_{\alpha\beta}/\partial x_{\sigma}$ is also a tensor" ("Für solche Transformationen ist $\partial T_{\alpha\beta}/\partial x_{\sigma}$ auch Tensor"). As indicated by the arrows in equation (164), this is because the components $p_{\mu\nu}$ of the transformation matrix can put in front of the derivative operator $\partial/\partial x_{\sigma}$.

Einstein now tried to generalize the tensor $\Delta'_{\mu\nu\rho}$. He started by asking the question: "What is this tensor called when arbitrary substitutions are admitted?" ("Wie heisst dieser Tensor, wenn bel[iebige] Subst[itutionen] zugelassen werden?"). Clearly he was not familiar with the notion of a covariant derivative at this point. Einstein used the same two-step procedure that he had used on p. 7R (see equation (147)). First, he wrote down how $\Delta'_{\mu\nu\rho}$ would transform if it transformed as a tensor under transfor-

110 The quantity $\Delta'_{\mu\nu\rho}$ is a concrete example of $\mathbf{O}(\partial', g')$ in equation (112).

mations from special primed to arbitrary unprimed coordinates. He then took expression (163) for $\Delta'_{\mu\nu\rho}$ and used the standard transformation rules to rewrite its building blocks in terms of their counterparts in unprimed coordinates. He thus arrived at the following expression for the tensor $\Delta_{\mu\nu\rho}$ in unprimed coordinates[111]

$$\Delta_{\mu\nu\rho} = \sum_{\mu\nu\rho} \pi_{m\mu}\pi_{n\nu}p_{r\rho}\Delta'_{mnr} = \sum \pi_{m\mu}\pi_{n\nu}p_{r\rho}\pi_{r\alpha}\frac{\partial}{\partial x_\alpha}(p_{m\delta}p_{n\varepsilon}T_{\delta\varepsilon}) \tag{165}$$

Einstein initially wrote $\pi_{\rho\alpha}$ but eventually corrected it to $\pi_{r\alpha}$. The right-hand side of equation (165) gives a sum of three terms:[112]

$$\frac{\partial T_{\mu\nu}}{\partial x_\rho} + \sum \pi_{m\mu}\pi_{n\nu}p_{r\rho}\pi_{r\alpha}\frac{\partial p_{n\varepsilon}}{\partial x_\alpha}p_{m\delta}T_{\delta\varepsilon} + \sum \pi_{m\mu}\pi_{n\nu}p_{r\rho}\pi_{r\alpha}p_{n\varepsilon}\frac{\partial p_{m\delta}}{\partial x_\alpha}T_{\delta\varepsilon}\,. \tag{166}$$

Using that $\pi_{m\mu}p_{m\delta} = \delta_{\mu\delta}$, Einstein simplified the second term, still with $\pi_{\rho\alpha}$ instead of $\pi_{r\alpha}$ and omitting $T_{\delta\varepsilon}$:

$$\pi_{n\nu}\frac{\partial p_{n\varepsilon}}{\partial x_\alpha}\pi_{\rho\alpha}p_{r\rho}\,. \tag{167}$$

This expression does not allow for further simplification. This may be what drew Einstein's attention to the error in equation (165), which he then corrected. The correct expressions (165)-(166) can be simplified further. This yields:

$$\frac{\partial T_{\mu\nu}}{\partial x_\rho} + \sum\left(\pi_{n\nu}\frac{\partial p_{n\varepsilon}}{\partial x_\rho}\right)T_{\mu\varepsilon} + \sum\left(\pi_{m\mu}\frac{\partial p_{m\delta}}{\partial x_\rho}\right)T_{\delta\nu}\,. \tag{168}$$

To generalize the exterior derivative of $T_{\mu\nu}$, one proceeds in the same way as in generalizing the divergence of $T_{\mu\nu}$ on the basis of expression (149).

Using that the metric in primed coordinates is the Minkowski metric in standard diagonal form, one first expresses the components of the transformation matrices and their derivatives in the last two terms of equation (168) in terms of the components of the Minkowski metric and their derivatives in the arbitrary unprimed coordinates. In the resulting expression, one then substitutes the components of an arbitrary metric for the components of the Minkowski metric in unprimed coordinates.

8R At the top of p. 8R, Einstein tried to find the relation between the transformation matrices and the Minkowski metric in the unprimed coordinates. This relation is given by the first equation on p. 8R:

$$\frac{\partial g_{\rho\sigma}}{\partial x_\tau} = \sum_{\beta\alpha} g_{\rho\beta}\pi_{\alpha\beta}\frac{\partial p_{\alpha\sigma}}{\partial x_\tau} + \sum_{\alpha\beta} g_{\alpha\sigma}\pi_{\beta\alpha}\frac{\partial p_{\beta\rho}}{\partial x_\tau}\,. \tag{169}$$

---

111 Equation (165) is a concrete example of a combination of equations (112) and (113): $\tilde{\mathbf{O}}(\partial, g) \equiv \tilde{T}^{-1}\mathbf{O}(\partial', g') = \tilde{T}^{-1}\mathbf{O}(T(\partial), T(g))$.

112 Expression (166) is a concrete example of the right-hand side of equation (113): $\mathbf{O}(\partial, g) + \mathbf{C}(\partial, g, p, \pi)$.

Although the derivation of this equation is not recorded in the notebook, it is easily reconstructed. It is basically the same calculation as in equations (163)–(168), only for $g_{\mu\nu}$ instead of $T_{\mu\nu}$. The starting point of the derivation is the observation that in the primed coordinates in which the Minkowski metric takes on its standard diagonal form, the exterior derivative of the metric vanishes:

$$\frac{\partial g'_{\alpha\beta}}{\partial x'_\gamma} = 0 . \tag{170}$$

The contraction of the left-hand side of this equation with the transformation matrices $p_{\alpha\rho}p_{\beta\sigma}p_{\gamma\tau}$ obviously still vanishes:

$$\sum p_{\alpha\rho}p_{\beta\sigma}p_{\gamma\tau}\frac{\partial g'_{\alpha\beta}}{\partial x'_\gamma} = 0 . \tag{171}$$

Expressing the primed quantities in terms of their unprimed counterparts, one finds that

$$\sum p_{\alpha\rho}p_{\beta\sigma}p_{\gamma\tau}\left(\pi_{\gamma\lambda}\frac{\partial}{\partial x_\lambda}(\pi_{\alpha\mu}\pi_{\beta\nu}g_{\mu\nu})\right) = 0 , \tag{172}$$

which can be rewritten as

$$\frac{\partial g_{\rho\sigma}}{\partial x_\tau} + p_{\beta\sigma}\frac{\partial \pi_{\beta\nu}}{\partial x_\tau}g_{\rho\nu} + p_{\alpha\rho}\frac{\partial \pi_{\alpha\mu}}{\partial x_\tau}g_{\mu\sigma} = 0 . \tag{173}$$

Since $p_{\beta\sigma}\pi_{\beta\nu} = \delta_{\sigma\nu}$ and, consequently,

$$0 = \frac{\partial}{\partial x_\mu}(p_{\beta\sigma}\pi_{\beta\nu}) = p_{\beta\sigma}\frac{\partial \pi_{\beta\nu}}{\partial x_\mu} + \pi_{\beta\nu}\frac{\partial p_{\beta\sigma}}{\partial x_\mu} , \tag{174}$$

equation (173) is equivalent to

$$\frac{\partial g_{\rho\sigma}}{\partial x_\tau} - g_{\rho\nu}\pi_{\beta\nu}\frac{\partial p_{\beta\sigma}}{\partial x_\tau} - g_{\mu\sigma}\pi_{\alpha\mu}\frac{\partial p_{\alpha\rho}}{\partial x_\tau} = 0 . \tag{175}$$

Bringing the second and the third term to the right-hand side and relabeling indices ($\nu\beta \to \beta\alpha$ in the second term, $\mu\alpha \to \alpha\beta$ in the third term), one finds an expression for the exterior derivative of the covariant metric,

$$\frac{\partial g_{\rho\sigma}}{\partial x_\tau} = g_{\rho\beta}\pi_{\alpha\beta}\frac{\partial p_{\alpha\sigma}}{\partial x_\tau} + g_{\alpha\sigma}\pi_{\beta\alpha}\frac{\partial p_{\beta\rho}}{\partial x_\tau} , \tag{176}$$

which is just equation (169) given at the top of p. 8R. Underneath this equation, Einstein wrote down a similar equation for the contravariant metric:[113]

113 It looks as if Einstein first started to write down equation (177) next to equation (169) rather than underneath it.

$$\frac{\partial\gamma_{\mu\nu}}{\partial x_\sigma} = \sum_{\alpha\beta}\gamma_{\mu\alpha}p_{\beta\alpha}\frac{\partial\pi_{\beta\nu}}{\partial x_\sigma} + \sum_{\alpha\beta}\gamma_{\alpha\nu}p_{\beta\alpha}\frac{\partial\pi_{\beta\mu}}{\partial x_\sigma}. \tag{177}$$

The derivation of this equation is fully analogous to the derivation of equation (176). The only difference is that the starting point is now $\partial\gamma'_{\alpha\beta}/\partial x'_\gamma = 0$ rather than $\partial g'_{\alpha\beta}/\partial x'_\gamma = 0$.[114]

Einstein did not proceed any further. With hindsight, however, knowing that the generalization that Einstein was looking for is just the covariant derivative, one can easily complete his chain of reasoning. With the help of equation (169) and two equations like it with different permutations of the indices ρ, σ, and τ, one can show that the relation between the transformation matrices and the Minkowski metric in the unprimed coordinates that Einstein was looking for is given by

$$\pi_{\alpha\beta}\frac{\partial p_{\alpha\sigma}}{\partial x_\tau} = \begin{Bmatrix}\beta\\ \sigma\tau\end{Bmatrix} = \frac{1}{2}\gamma_{\beta\alpha}\left(\frac{\partial g_{\alpha\sigma}}{\partial x_\tau} + \frac{\partial g_{\alpha\tau}}{\partial x_\sigma} - \frac{\partial g_{\sigma\tau}}{\partial x_\rho}\right), \tag{178}$$

where the curly brackets represent the Christoffel symbols of the second kind. This can easily be verified by inserting equation (178) back into equation (169).[115] Inserting equation (178) into expression (168), one finds the generally-covariant analogue of the exterior derivative of the contravariant tensor $T_{\mu\nu}$:[116]

$$\begin{aligned}&\frac{\partial T_{\mu\nu}}{\partial x_\rho} + \sum\left(\pi_{n\nu}\frac{\partial p_{n\varepsilon}}{\partial x_\rho}\right)T_{\mu\varepsilon} + \sum\left(\pi_{m\mu}\frac{\partial p_{m\delta}}{\partial x_\rho}\right)T_{\delta\nu}\\ &\qquad = \frac{\partial T_{\mu\nu}}{\partial x_\rho} + \sum\begin{Bmatrix}\nu\\ \varepsilon\rho\end{Bmatrix}T_{\mu\varepsilon} + \sum\begin{Bmatrix}\mu\\ \delta\rho\end{Bmatrix}T_{\delta\nu}\end{aligned} \tag{179}$$

Inserting equation (178) into the right-hand side of the equation in footnote 107, one finds the covariant divergence of $T_{\mu\nu}$:

$$\frac{\partial T_{\sigma\tau}}{\partial x_\tau} + \pi_{\mu\sigma}\frac{\partial p_{\mu i}}{\partial x_\tau}T_{i\tau} + \pi_{\nu\tau}\frac{\partial p_{\nu\kappa}}{\partial x_\tau}T_{\sigma\kappa} = \frac{\partial T_{\sigma\tau}}{\partial x_\tau} + \begin{Bmatrix}\sigma\\ i\tau\end{Bmatrix}T_{i\tau} + \begin{Bmatrix}\tau\\ \kappa\tau\end{Bmatrix}T_{\sigma\kappa}. \tag{180}$$

---

114 Equation (177) can also be obtained by substituting the contravariant metric $\gamma_{\mu\nu}$ for the contravariant tensor $T_{\mu\nu}$ in equation (123), using equation (174) and the fact that $\partial\gamma'_{\alpha\beta}/\partial x'_\gamma = 0$ and by relabeling indices.

115 Inserting equation (178) into the right-hand side of equation (169), one recovers the left-hand side:

$$\sum_{\beta\alpha}g_{\rho\beta}\pi_{\alpha\beta}\frac{\partial p_{\alpha\sigma}}{\partial x_\tau} + \sum_{\alpha\beta}g_{\alpha\sigma}\pi_{\beta\alpha}\frac{\partial p_{\beta\rho}}{\partial x_\tau} = \sum g_{\rho\beta}\begin{Bmatrix}\beta\\ \sigma\tau\end{Bmatrix} + \sum g_{\alpha\sigma}\begin{Bmatrix}\alpha\\ \rho\tau\end{Bmatrix} = \frac{\partial g_{\rho\sigma}}{\partial x_\tau}.$$

116 The final expressions in equations (179) and (180) are concrete examples of the quantities $\mathbf{O}_{\text{gen}}(\partial, g) \equiv \mathbf{O}(\partial, g) + \mathbf{D}(\partial, g)$ defined in equation (117).

These last two equations show how close Einstein came to finding the correct generally-covariant generalizations of the two constituents of the core operator, the divergence and the exterior derivative, using the second one of the two strategies that we distinguished in the introduction of sec. 4.3.

Why did Einstein not pursue this calculation beyond the first two lines on p. 8R? It seems unlikely that the complexity of having to solve equation (60) for $\pi_{\alpha\beta}(\partial p_{\alpha\sigma}/\partial x_\tau)$ would have deterred him. On pp. 14R–18R, we shall see Einstein pursue far more cumbersome calculations with great tenacity. A more plausible answer is that Einstein realized at this point, if not earlier, that the generalization of the exterior derivative he was in the process of constructing cannot be used to build a generalized core operator that could serve as the left-hand side of the gravitational field equations. The problem is that the exterior derivative of the metric vanishes. On p. 5R, Einstein had already found that the covariant divergence of the metric vanishes, but that does not mean that the generalization of the core operator, which is essentially the divergence of the exterior derivative of the metric, vanishes. If the exterior derivative of the metric vanishes, however, the core operator vanishes as well.

Equation (177), the second equation on p. 8R, is, in fact, the statement that the covariant exterior derivative of the contravariant metric vanishes. This can be seen as follows. From equations (174) and (178) it follows that

$$p_{\beta\sigma}\frac{\partial\pi_{\beta\nu}}{\partial x_\mu} = -\pi_{\beta\nu}\frac{\partial p_{\beta\sigma}}{\partial x_\mu} = -\begin{Bmatrix}\nu\\ \sigma\mu\end{Bmatrix}. \tag{181}$$

Substituting this expression into equation (177), one finds,

$$\frac{\partial\gamma_{\mu\nu}}{\partial x_\sigma} = -\sum\gamma_{\mu\alpha}\begin{Bmatrix}\nu\\ \alpha\sigma\end{Bmatrix} - \sum\gamma_{\alpha\nu}\begin{Bmatrix}\mu\\ \alpha\sigma\end{Bmatrix} \tag{182}$$

Comparison with equation (179) shows that equation (182) expresses the vanishing of the covariant exterior derivative of the metric.[117]

Einstein did not have to find the relation between the transformation matrices and the metric and rewrite equation (177) in the form of equation (182) to see that the generalization of the exterior derivative acting on the metric would vanish. It is, in fact, a direct consequence of the method that Einstein used to construct this generalization. Whatever the exact form of the sought-after operator acting on the metric in the arbitrary unprimed coordinates, its form in the special primed coordinates used to construct it is $\partial\gamma'_{\mu\nu}/\partial x'_\rho$ (cf. equation (163)). The primed coordinates were chosen in such a way that the metric—be it the Minkowski metric or an arbitrary metric—takes

117 When Grossmann introduced the Christoffel symbols in his part of the *Entwurf* paper, he added a footnote saying: "On the basis of these formulae one can easily prove that the exterior derivative of the fundamental tensor vanishes identically" ("Auf Grund dieser Formeln beweist man leicht, dass die Erweiterung des Fundamentaltensors identisch verschwindet," Einstein and Grossmann 1913, part 2, sec. 2).

on the form of the standard diagonal Minkowski metric in these coordinates. So the generalized exterior derivative of the metric vanishes in the special primed coordinates. Now this quantity was constructed to transform as a tensor under arbitrary coordinate transformations. So it will vanish in all coordinate systems.[118]

At this point, Einstein had no choice but to abandon his second strategy for finding field equations on the basis of the core operator: generalizing this operator to an expression that transforms as a tensor under arbitrary (autonomous) coordinate transformations does not work. He returned to the first strategy of finding non-autonomous non-linear transformations under which the core operator itself—if necessary with correction terms—transforms as a tensor.

8R–9R

4.4 *Trying to Extract Field Equations and a Gravitational Stress-Energy Tensor from the Beltrami Invariants (8R–9R)*

On pp. 7L-8R Einstein had examined the transformation properties of candidate field equations based on the core operator, the natural analogue of the Poisson equation in a theory in which the gravitational potential is represented by a tensor rather than a scalar. Einstein's approach on these pages had thus been along the lines of what we call the physical strategy. On pp. 8R–9R, Einstein returned to the mathematical strategy, more specifically to the exploration of the Beltrami invariants introduced on p. 6L. He tried to extract the core operator from the second Beltrami invariant, using (some power of) the determinant of the metric as the arbitrary scalar function in the definition of this invariant. The connection between the core operator and the Beltrami invariant might throw light on the covariance properties of the former. Einstein's investigation of the covariance properties of the core operator on pp. 7L–R had remained inconclusive.

On p. 8R, Einstein returned to the basic expression for the first and the second Beltrami invariants. The first Beltrami invariant can be used to find a candidate for the quantity representing gravitational energy-momentum, the second to find a candidate for the left-hand side of the field equations. On p. 9L Einstein managed to write the second Beltrami invariant as a sum of two contributions, the first of which is the contraction of the metric with the core operator. On the bottom half of p. 9L, Einstein tried to relate the second contribution to gravitational energy-momentum. This is the first time in the notebook that Einstein, drawing on his experience with the 1912 static the-

---

118 In the *Entwurf* paper, Einstein mentioned this problem as one of the obstacles to formulating generally-covariant field equations: "These operations [i.e., the divergence and the exterior derivative operators] degenerate if they are applied to the fundamental tensor $g_{\mu\nu}$. From this it seems to follow that the equations sought will be covariant only with respect to a particular group of transformations, which for the time being, however, is unknown to us" ("Aber es degenerieren diese Operationen, wenn sie an dem Fundamentaltensor $g_{\mu\nu}$ ausgeführt werden. Es scheint daraus hervorzugehen, daß die gesuchten Gleichungen nur bezüglich einer gewissen Gruppe von Transformationen kovariant sein werden, welche Gruppe uns aber vorläufig unbekannt ist." Einstein and Grossmann 1913, part 1, sec. 5).

ory,[119] introduces the notion that gravitational energy-momentum should enter the field equations on equal footing with the energy-momentum of matter. Expecting gravitational energy-momentum to be represented by a generally-covariant tensor, Einstein turned to the first Beltrami invariant to find a candidate for the stress-energy tensor of the gravitational field. The contraction of the metric with this supposed gravitational stress-energy tensor, however, turns out to be slightly different from the second contribution to the expression for the second Beltrami invariant, and Einstein abandoned the idea of interpreting this contribution in terms of gravitational energy-momentum.

On p. 9R Einstein tried to find the infinitesimal non-autonomous transformations under which this contribution to the second Beltrami invariant transforms as a scalar. The first contribution, the contraction of the metric and the core operator, would then be the difference between two scalars (for this restricted class of transformations) and therefore be such a scalar itself. This in turn would suggest that the core operator transform as a tensor under these transformations. Since the rationale behind the return to the Beltrami invariants on p. 8R was presumably to avoid non-autonomous transformations, which Einstein had found difficult to handle (see pp. 7L–8R), it is not surprising that the Beltrami invariants no longer explicitly appear in the notebook after these calculations on pp. 8R–9R. Einstein, however, continued to use the restriction to unimodularity in his calculations on the following pages. This suggests that he still hoped to find some connection between the field equations and the Beltrami invariants.

Underneath the first horizontal line on p. 8R, Einstein substituted (some power α 8R
of) the determinant G of the metric for the arbitrary function φ in the definition of the first and the second Beltrami invariants (see equations (79) and (80), respectively):

$$\varphi_1 = \gamma_{\mu\nu}\frac{\partial G}{\partial x_\mu}\frac{\partial G}{\partial x_\nu}\,, \tag{183}$$

$$\varphi_2 = \frac{1}{\sqrt{G}}\frac{\partial}{\partial x_\mu}\left(\gamma_{\mu\nu}\sqrt{G}\frac{\partial}{\partial x_\nu}G^\alpha\right)\,. \tag{184}$$

Since G is a scalar only under unimodular transformations, the Beltrami invariants φ_1 and φ_2 above are no longer generally-covariant scalars but invariants under this restricted class of transformations only. To establish the connection with the core operator,

119 In (Einstein 1912b, sec. 4), it was pointed out that the field equation originally proposed for the theory for static gravitational fields, $\Delta c = kc\sigma$ (where the variable speed of light c doubles as the gravitational potential, k is a constant, and σ is the mass density), is in conflict with the action-equals-reaction principle and thereby with energy-momentum conservation. Einstein remedied the problem by adding the gravitational energy density to the source term: $\Delta c = k\left\{c\sigma + \frac{1}{2k}\frac{\mathrm{grad}^2 c}{c}\right\}$.

$$\sum \frac{\partial}{\partial x_\mu}\left(\gamma_{\mu\nu}\frac{\partial \gamma_{i\kappa}}{\partial x_\nu}\right) \tag{185}$$

(see equation (118)), Einstein rewrote $\partial G/\partial x_\nu$ as

$$\frac{\partial G}{\partial x_\nu} = \sum_{i\kappa}\frac{\partial g_{i\kappa}}{\partial x_\nu}G_{i\kappa} = \sum G\frac{\partial g_{i\kappa}}{\partial x_\nu}\gamma_{i\kappa} = -G\sum g_{i\kappa}\frac{\partial \gamma_{i\kappa}}{\partial x_\nu}, \tag{186}$$

where $G_{i\kappa}$ is the minor of $g_{i\kappa}$. In the second step, Einstein used that $\gamma_{i\kappa} = G_{i\kappa}/G$, and in the third step that $\partial/\partial x_\nu(\sum g_{i\kappa}\gamma_{i\kappa}) = 0$. Next to this expression he wrote "of zeroth power" ("nullter Potenz").[120]

Inserting the first expression for $\partial G/\partial x_\nu$ in equation (186) into equation (183), Einstein arrived at:

$$\varphi_1 = \sum_{i\kappa lm\mu\nu}\gamma_{\mu\nu}\left(\frac{\partial g_{i\kappa}}{\partial x_\mu}\frac{\partial g_{lm}}{\partial x_\nu}G_{i\kappa}G_{lm}\right), \tag{187}$$

"or" ("oder"), as he wrote next to it, using the second expression for $\partial G/\partial x_\nu$ in equation (186),

$$\sum \gamma_{\mu\nu}\left(\frac{\partial g_{i\kappa}}{\partial x_\mu}\frac{\partial g_{lm}}{\partial x_\nu}\gamma_{i\kappa}\gamma_{lm}\right). \tag{188}$$

A factor G^2 is omitted in this last expression. This is inconsequential: since both φ_1 and G are scalars under unimodular transformations, φ_1/G^2 is too.

After drawing a horizontal line, Einstein turned to the second Beltrami invariant,

$$\varphi_2 = \sum \frac{\partial}{\partial x_\mu}\left(\sqrt{G}\gamma_{\mu\nu}\frac{\partial G^\alpha}{\partial x_\nu}\right). \tag{189}$$

Once again, he omitted an inconsequential factor $\sqrt{G}$ (cf. equation (184)).[121] With the help of equation (186), Einstein rewrote this as

120 This is the only place in the notebook where the word "Potenz" occurs. Einstein only wrote "Potenz" after writing and deleting first "of zeroth kind" ("nullter Art") and then "of zeroth order" ("nullter Ordnung"). At the top of p. 9L he switched back to "zeroth order" ("nullter Ordnung"). In (Bianchi 1910, Ch. II, sec. 22), the "order" ("Ordnung") of a "differential parameter" ("Differentialparameter")—i.e., an expression constructed out of the metric and its derivatives and a number of arbitrary functions and their derivatives—is defined as the highest-order derivative of the arbitrary functions occurring in it. Since there are no arbitrary functions in φ_1 and φ_2 in equations (183) and (184), these quantities are of zeroth order in this sense.

121 The expression written next to equation (189), $\partial/\partial x_\mu(\sqrt{G}\gamma_{\mu\nu}\partial\psi/\partial x_\nu)$, is likewise a scalar under unimodular transformations.

$$\alpha\sum\frac{\partial}{\partial x_\mu}\left(\sqrt{G}\gamma_{\mu\nu}G^{\alpha-1}\frac{\partial g_{i\kappa}}{\partial x_\nu}G_{i\kappa}\right)\approx\sum\frac{\partial}{\partial x_\mu}\left(G^{\alpha+\frac{1}{2}}\gamma_{\mu\nu}\gamma_{i\kappa}\frac{\partial g_{i\kappa}}{\partial x_\nu}\right). \tag{190}$$

As indicated by the proportionality sign, the overall factor α is dropped from the expression. Einstein now set $\alpha = -1/2$ and transferred the differentiation operator $\partial/\partial x_\nu$ from the covariant to the contravariant metric using that $\partial/\partial x_\nu(\sum g_{i\kappa}\gamma_{i\kappa}) = 0$. The resulting expression contains the core operator:

$$\sum\frac{\partial}{\partial x_\mu}\left(\gamma_{\mu\nu}\gamma_{i\kappa}\frac{\partial g_{i\kappa}}{\partial x_\nu}\right) = -\sum\frac{\partial}{\partial x_\mu}\left(\gamma_{\mu\nu}g_{i\kappa}\frac{\partial\gamma_{i\kappa}}{\partial x_\nu}\right). \tag{191}$$

At the bottom of the page, Einstein briefly turned to the first Beltrami invariant and wrote down a "Different expression for the above scalar φ_1" ("Anderer Ausdruck für obigen Skalar φ_1"):

$$\sum g_{\mu\mu'}\gamma_{\mu\nu}\frac{\partial G}{\partial x_\nu}\gamma_{\mu'\nu'}\frac{\partial G}{\partial x_{\nu'}} \tag{192}$$

(which is indeed equivalent to equation (183) since $g_{\mu\mu'}\gamma_{\mu\nu}\gamma_{\mu'\nu'} = \gamma_{\nu\nu'}$). The phrase "different expression for φ_1" ("Anderer Ausdruck für φ_1") is repeated two lines farther down. Einstein returned to this expression on p. 9L to extract a candidate for the stress-energy tensor of the gravitational field.

Einstein drew a horizontal line and copied the final expression for φ_2 from equation (190),

$$\sum\frac{\partial}{\partial x_\mu}\left(G^{\alpha+\frac{1}{2}}\gamma_{\mu\nu}\gamma_{i\kappa}\frac{\partial g_{i\kappa}}{\partial x_\nu}\right). \tag{193}$$

Underneath this expression he began to evaluate one of the derivatives in expression (192),

$$\frac{\partial G}{\partial x_\nu} = -\sum g_{i\kappa}\frac{\partial G_{i\kappa}}{\partial x_\nu}, \tag{194}$$

but then deleted the erroneous right-hand side.

Einstein continued his investigation of the Beltrami invariants on p. 9L. He began 9L
by rewriting the first Beltrami invariant (183) modulo a factor G^{-2}. Starting from expression (188) on p. 8R, he transferred the derivative operators from the covariant to the contravariant metric, using that $\partial/\partial x_\mu(\sum g_{i\kappa}\gamma_{i\kappa}) = 0$. He thus arrived at:[122]

122 The quantity φ_1 in equation (195) needs to be multiplied by G^2 to obtain the first Beltrami invariant φ_1 as defined in equation (183). See footnote 120 for a discussion of the term "zeroth order" ("nullter Ordnung") written next to equation (195).

$$\varphi_1 = \sum g_{i\kappa} g_{lm} \gamma_{\mu\nu} \frac{\partial \gamma_{i\kappa}}{\partial x_\mu} \frac{\partial \gamma_{lm}}{\partial x_\nu} . \tag{195}$$

He similarly rewrote expression (190) for the second Beltrami invariant, setting $\alpha = -1/2$ as he had done on p. 8R:

$$\varphi_2 = \sum \frac{\partial}{\partial x_\mu} \left(\cancel{G^{\left(\alpha+\frac{1}{2}\right)}} g_{i\kappa} \gamma_{\mu\nu} \frac{\partial \gamma_{i\kappa}}{\partial x_\nu} \right) . \tag{196}$$

Einstein expanded this expression to

$$\varphi_2 = \sum g_{i\kappa} \frac{\partial}{\partial x_\mu} \left(\gamma_{\mu\nu} \frac{\partial \gamma_{i\kappa}}{\partial x_\nu} \right) + \underbrace{\sum \gamma_{\mu\nu} \frac{\partial g_{i\kappa}}{\partial x_\mu} \frac{\partial \gamma_{i\kappa}}{\partial x_\nu}} . \tag{197}$$

The first term is the core operator (185) contracted with the covariant metric. The second term required further attention. Einstein tried to use the relation

$$0 = \sum \left(\frac{\partial g_{i\kappa}}{\partial x_\mu} \gamma_{i\kappa} + g_{i\kappa} \frac{\partial \gamma_{i\kappa}}{\partial x_\mu} \right) \tag{198}$$

to deal with it. Differentiating this relation, he did indeed recover, up to a contraction with $\gamma_{\mu\nu}$, the second term on the right-hand side of equation (197) but at the price of introducing three other terms:

$$0 = \sum \left(\frac{\partial^2 g_{i\kappa}}{\partial x_\mu \partial x_\nu} \gamma_{i\kappa} + \frac{\partial g_{i\kappa}}{\partial x_\mu} \frac{\partial \gamma_{i\kappa}}{\partial x_\nu} + . + g_{i\kappa} \frac{\partial^2 \gamma_{i\kappa}}{\partial x_\mu \partial x_\nu} \right) \tag{199}$$

Rather than rewriting the second Beltrami invariant with the help of equation (199), Einstein tried to interpret the second term in equation (197) for φ_2 with the help of the first Beltrami invariant modulo a factor G^{-2}, i.e., the quantity φ_1 at the top of p. 9L (see equation (195)).

In the passage in the middle of p. 9L, set off by two horizontal lines, Einstein tried to interpret the difference between φ_2 and the core operator contracted with $g_{i\kappa}$ as minus a candidate gravitational stress-energy tensor contracted with $g_{i\kappa}$. As was explained in the introduction to this subsection, the second Beltrami invariant would then in all likelihood[123] yield a candidate for the left-hand side of the field equations which is (*i*) a tensor under unimodular transformations and (*ii*) equal to the core operator minus the gravitational stress-energy tensor. Setting this candidate equal to the stress-energy tensor for matter, one sees that such field equations satisfy the requirement that all energy-momentum enters the field equations on the same footing. Einstein naturally assumed at this point that gravitational energy-momentum like the

123 In all likelihood, because it obviously cannot be ruled out that the expression is not a tensor even though its contraction with the metric tensor is a scalar. Its contraction with an *arbitrary* second-rank tensor might not be.

energy-momentum of matter would be represented by a tensor.[124] By analogy to both the stress-energy tensor for the electromagnetic field and the one for the gravitational field of his 1912 static theory,[125] he furthermore expected the gravitational stress-energy tensor to be quadratic in first-order derivatives of the metric. Any such object, transforming as a tensor at least under unimodular transformations, would have to be constructed out of the first Beltrami invariant, multiplied possibly by other scalars under unimodular transformations such as the determinant G of the metric. The quantity φ_1 in equation (195) is such an object. Moreover, φ_1 looks very similar to the term in φ_2 to be written as the contraction of $g_{i\kappa}$ with the gravitational stress-energy tensor.[126] Einstein thus tried to extract a gravitational stress-energy tensor from φ_1.

Einstein began this attempt with expression (192) on p. 8R. Dividing this equation by G^2, one arrives at the first equation in the passage set off between two horizontal lines on the bottom half of p. 9L:

$$\varphi_1 = \frac{1}{G^2}\sum g_{\mu\mu'}\,\gamma_{\mu\nu}\frac{\partial G}{\partial x_\nu}\,\gamma_{\mu'\nu'}\frac{\partial G}{\partial x_{\nu'}}. \tag{200}$$

The “presumable gravitation tensor” (“vermutlicher Gravitationstensor”) that can be read off from this expression is:

$$\frac{1}{G^2}\sum_{\nu\nu'}\gamma_{\mu\nu}\frac{\partial G}{\partial x_\nu}\gamma_{\mu'\nu'}\frac{\partial G}{\partial x_{\nu'}} = \sum_{\nu\nu' ii'\kappa\kappa'}\gamma_{\mu\nu}\frac{\partial g_{i\kappa}}{\partial x_\nu}\gamma_{i\kappa}\gamma_{\mu'\nu'}\frac{\partial g_{i'\kappa'}}{\partial x_{\nu'}}\gamma_{i'\kappa'}. \tag{201}$$

(In the second step, Einstein used equation (186).) For the remainder of the argument on p. 9L, however, Einstein used the expression[127]

$$\sum\gamma_{\mu\nu}\gamma_{\mu'\nu'}\frac{\partial\psi}{\partial x_\nu}\frac{\partial\psi}{\partial x_{\nu'}}, \tag{202}$$

defining ψ through

$$\lg G = \psi. \tag{203}$$

---

124 (Einstein and Grossmann 1914, 218, footnote 1) is the first place where Einstein explicitly stated in print that the assumption that gravitational energy-momentum can be represented by a tensor is erroneous. In this footnote he identified this assumption as the flaw in an argument in (Einstein 1914a) that appeared to restrict the covariance of the *Entwurf* field equations to linear transformations. For discussion of this argument and its flaws, see (Norton 1984, sec. 5), “What Did Einstein Know …” sec. 2 and “Untying the Knot …” sec. 3.3 (both in this volume).

125 (Einstein 1912b, 456–457).

126 Using $\frac{\partial g_{i\kappa}}{\partial x_\mu} = -g_{il}g_{\kappa m}\frac{\partial\gamma_{lm}}{\partial x_\mu}$, one can rewrite the second contribution to φ_2 in equation (197) as $\sum\gamma_{\mu\nu}\frac{\partial g_{i\kappa}}{\partial x_\mu}\frac{\partial\gamma_{i\kappa}}{\partial x_\nu} = -\sum\gamma_{\mu\nu}g_{il}g_{\kappa m}\frac{\partial\gamma_{lm}}{\partial x_\mu}\frac{\partial\gamma_{i\kappa}}{\partial x_\nu}$, which closely resembles the expression for φ_1 in equation (195).

127 On p. 8R, Einstein had written the second Beltrami invariant in terms of this function ψ (see footnote 121).

He drew a line from expression (202) to the expression on the left-hand side of equation (201) and encircled these two expressions. On the next line, he explicitly made the claim that this choice of a gravitational stress-energy tensor is unique: "[This] is the only tensor in which we differentiate only once" ("Ist der einzige Tensor, in dem nur einmal diff[erenziert] wird").

To see whether his gravitational stress-energy tensor (202) would be acceptable from a physical point of view as well, Einstein, writing "divergence calculated" ("Divergenz gebildet"), substituted it for the stress-energy tensor $T_{\mu\nu}$ of matter in equation (71),

$$\sum_{\nu n}\frac{\partial}{\partial x_n}(\sqrt{G}g_{m\nu}T_{\nu n})+\frac{1}{2}\sqrt{G}\sum\frac{\partial g_{\mu\nu}}{\partial x_m}T_{\mu\nu}=0\,, \tag{204}$$

for the energy-momentum balance between matter and gravitational field.

On p. 5R, Einstein derived this equation for the stress-energy tensor of pressureless dust. He then postulated the same equation for the stress-energy tensor of any matter. Gravitational energy-momentum, however, plays a special role and one cannot simply substitute the gravitational stress-energy tensor for $T_{\mu\nu}$ in equation (204).[128] On p. 9L, Einstein did not recognize the special status of gravitational energy-momentum and demanded that expression (202), his candidate for a gravitational stress-energy tensor, like any other stress-energy tensor $T_{\mu\nu}$, satisfy the energy-momentum balance equation (204):

$$\sum\frac{\partial}{\partial x_{\mu}{}'}\left(\sqrt{G}g_{m\mu}\gamma_{\mu\nu}\gamma_{\mu'\nu'}\frac{\partial\psi}{\partial x_\nu}\frac{\partial\psi}{\partial x_{\nu'}}\right)=\frac{1}{2}\sum\sqrt{G}\frac{\partial g_{\mu\mu'}}{\partial x_m}\gamma_{\mu\nu}\gamma_{\mu'\nu'}\frac{\partial\psi}{\partial x_\nu}\frac{\partial\psi}{\partial x_{\nu'}}\,. \tag{205}$$

Einstein began to simplify both sides of this equation. Using (in slightly modernized notation involving the Kronecker delta) $\sum g_{m\mu}\gamma_{\mu\nu}=\delta_{m\nu}$, he wrote the left-hand side as:

$$\sum\frac{\partial}{\partial x_{\mu}{}'}\left(\sqrt{G}\gamma_{\mu'\nu'}\frac{\partial\psi}{\partial x_m}\frac{\partial\psi}{\partial x_{\nu'}}\right)= \tag{206}$$

He used that

$$\frac{\partial g_{\mu\mu'}}{\partial x_m}\gamma_{\mu\nu}\gamma_{\mu'\nu'}=-g_{\mu\mu'}\frac{\partial\gamma_{\mu\nu}}{\partial x_m}\gamma_{\mu'\nu'}, \tag{207}$$

and that $\sum g_{\mu\mu'}\gamma_{\mu'\nu'}=\delta_{\mu\nu}$ to rewrite the right-hand side of equation (205) as

128 Einstein subsequently recognized that gravitational energy-momentum cannot be handled in the same way as the energy-momentum of matter. On p. 13R, he took a first step in this direction (see sec. , especially the discussion following equation (419)). On p. 19R, he had essentially arrived at the treatment of gravitational energy-momentum that he would use and defend in the ensuing years (see the discussion following equation (481) in sec. 5.4.2; see also pp. 20L, 21L, 24R–26R).

$$= -\frac{1}{2}\sum\sqrt{G}\frac{\partial\gamma_{\nu\nu'}}{\partial x_m}\frac{\partial\psi}{\partial x_\nu}\frac{\partial\psi}{\partial x_{\nu'}} . \tag{208}$$

At this point, Einstein abandoned the ill-conceived condition (205). He also abandoned his attempt to interpret the second term in equation (197) for the Beltrami invariant φ_2 as the contraction of $g_{i\kappa}$ with the gravitational stress-energy tensor (202) extracted from φ_1. Perhaps Einstein had come to realize these two expressions are not quite the same.[129] However, he retained the notion that the left-hand side of the field equations be the sum of two terms, each transforming as a tensor, one term being the core operator, the other term representing gravitational energy-momentum.[130]

At the bottom of p. 9L and the top of p. 9R, Einstein returned to the investigation of the covariance properties of the core operator through its relation with the Beltrami invariant φ_2 in equation (196).[131] The second term in expression (197) for φ_2, which Einstein had tried to interpret as a correction term to the core operator representing gravitational energy-momentum, now had to be dealt with in a new way.[132]

At the top of p. 9R, Einstein copied equation (196) for the second Beltrami invariant φ_2 (see the line connecting the two expressions for φ_2 on pp. 9L–R) leaving the exponent α undetermined rather than setting $\alpha = -1/2$: 9R

$$\varphi_2 = \sum\frac{\partial}{\partial x_\mu}\left(G^{\alpha+\frac{1}{2}}g_{i\kappa}\gamma_{\mu\nu}\frac{\partial\gamma_{i\kappa}}{\partial x_\nu}\right) . \tag{209}$$

In the top right corner of p. 9R, using equation (186) for $\partial G/\partial x_\nu$, Einstein calculated the derivative of $G^{\alpha+1/2}$

$$\frac{\partial G^{\alpha+\frac{1}{2}}}{\partial x_\mu} = \left(\alpha+\frac{1}{2}\right)G^{\alpha+\frac{1}{2}}\sum\frac{\partial g_{\rho\sigma}}{\partial x_\mu}\gamma_{\rho\sigma} . \tag{210}$$

With the help of this relation, φ_2 can be rewritten as:[133]

---

129 The former can be written as $\sum g_{lm}\gamma_{\mu\nu}\frac{\partial\gamma_{il}}{\partial x_\mu}\frac{\partial\gamma_{\kappa m}}{\partial x_\nu}$, the latter as $\sum g_{lm}\gamma_{\mu\nu}\frac{\partial\gamma_{i\kappa}}{\partial x_\mu}\frac{\partial\gamma_{lm}}{\partial x_\nu}$ (cf. footnote 126).

130 Exploiting this general feature, Einstein developed a strategy for finding field equations compatible with the conservation principle. An embryonic version of this strategy can be found on p. 13R. On pp. 26L–R, he used the mature version of this strategy to find the *Entwurf* field equations.

131 At the bottom of p. 9L, Einstein wrote $\varphi_2 = \sum$, drew and then deleted a line connecting this incomplete expression to equation (199) in the middle of the page. A completed version of the expression occurs at the top of p. 9R. The final pair of equations on p. 9L, $a'_\mu = p_{\mu\nu}a_\nu$ and $\alpha'_\mu = \pi_{\mu\nu}\alpha_\nu$, give the transformation laws for contravariant and covariant vectors, respectively (see equations (119) and (120)).

132 From a modern perspective the question arises why Einstein continued to concentrate on the covariance properties of the core operator itself rather than on the covariance properties of the sum of the core operator and the gravitational stress-energy (pseudo-)tensor. The answer to this question is that Einstein tacitly assumed (see footnote 124) that both terms of this sum would separately transform as a tensor.

$$\varphi_2 = \left(\alpha + \frac{1}{2}\right) G^{\alpha + \frac{1}{2}} \sum g_{i\kappa} \gamma_{\mu\nu} \frac{\partial \gamma_{i\kappa}}{\partial x_\nu} \gamma_{\rho\sigma} \frac{\partial g_{\rho\sigma}}{\partial x_\mu} + G^{\alpha + \frac{1}{2}} \sum \frac{\partial}{\partial x}\left(g_{i\kappa} \gamma_{\mu\nu} \frac{\partial \gamma_{i\kappa}}{\partial x_\nu} \right). \tag{211}$$

As is indicated by the grouping of the factors in the expression under the first summation sign, this summation can be rewritten as (cf. equation (186))

$$\sum g_{i\kappa} \gamma_{\mu\nu} \frac{\partial \gamma_{i\kappa}}{\partial x_\nu} \gamma_{\rho\sigma} \frac{\partial g_{\rho\sigma}}{\partial x_\mu} = -\frac{1}{G^2} \sum \gamma_{\mu\nu} \frac{\partial G}{\partial x_\nu} \frac{\partial G}{\partial x_\nu}. \tag{212}$$

As Einstein noted, this means that the first term on the right-hand side of equation (211) is a scalar: "Obvious because $\partial G / \partial x_\nu$ vector of the second kind [i.e., a covariant vector]" ("Selbstverst[ändlich] weil … Vektor zweiter Art").

The second term on the right-hand side of equation (211) can be written as $G^{\alpha + 1/2}$ times the sum of two terms (see equation (197)): the core operator (185) contracted with g_{ik} and the term

$$\gamma_{\mu\nu} \frac{\partial g_{i\kappa}}{\partial x_\mu} \frac{\partial \gamma_{i\kappa}}{\partial x_\nu}, \tag{213}$$

which Einstein had tried in vain on p. 9L to write as the contraction of the metric with a gravitational stress-energy tensor.

On the remainder of p. 9R Einstein investigated under which (infinitesimal) unimodular non-autonomous transformations expression (213) would be invariant. Under such a restricted class of transformations the core operator contracted with $g_{i\kappa}$ and multiplied by $G^{\alpha + 1/2}$ would also be a scalar since, by virtue of equations (211), (212) and (197), it is equal to

$$G^{\alpha + \frac{1}{2}} g_{i\kappa} (\gamma)_{i\kappa} = \varphi_2 - \left(\alpha + \frac{1}{2}\right) G^{\alpha - \frac{3}{2}} \sum \gamma_{\mu\nu} \frac{\partial G}{\partial x_\nu} \frac{\partial G}{\partial x_\nu} - G^{\alpha + \frac{1}{2}} \sum \gamma_{\mu\nu} \frac{\partial g_{i\kappa}}{\partial x_\mu} \frac{\partial \gamma_{i\kappa}}{\partial x_\nu}, \tag{214}$$

where we used Einstein's notation $(\gamma)_{i\kappa}$ of p. 7L for the core operator (see equation (130)). The first two terms on the right-hand side transform as scalars under arbitrary unimodular transformations. The left-hand side will thus transform as a scalar under the restricted class of unimodular transformations under which the third term on the right-hand side transforms as a scalar. Presumably, the core operator would then transform as a tensor under these transformations. Einstein thus set out to determine exactly how the class of unimodular transformations would have to be further restricted. As he wrote on the second line below equation (211): "Substitutions must be restricted more" ("Subst[itutionen] müssen mehr eingeschränkt werden").[134]

To find the condition for non-autonomous transformations leaving expression (213) invariant, Einstein used the usual two-step procedure.[135] He started with the expression in primed coordinates

133 The factor $(\alpha + 1/2) G^{\alpha + 1/2}$ in front of the summation sign was added later. In the second term $\partial / \partial x$ should be $\partial / \partial x_\mu$.

$$\sum \gamma'_{\mu\nu} \frac{\partial g'_{i\kappa}}{\partial x'_{\mu}} \frac{\partial \gamma'_{i\kappa}}{\partial x'_{\nu}} . \tag{215}$$

Using the transformation laws for the various ingredients of this expression (see equations (119) and (120) repeated at the bottom of p. 9L), he wrote expression (215) in terms of quantities in unprimed coordinates:

$$\sum \gamma'_{\mu\nu} \frac{\partial g'_{i\kappa}}{\partial x'_{\mu}} \frac{\partial \gamma'_{i\kappa}}{\partial x'_{\nu}} = \sum p_{\mu\alpha} p_{\nu\beta} \gamma_{\alpha\beta} \pi_{\mu l} \pi_{\nu m} \frac{\partial}{\partial x_l} (\pi_{i\delta} \pi_{\kappa\varepsilon} g_{\delta\varepsilon}) \frac{\partial}{\partial x_m} (p_{i\delta'} p_{\kappa\varepsilon'} \gamma_{\delta'\varepsilon'}) , \tag{216}$$

The right-hand side can be simplified by using that the matrices p and π are each other's inverse (see equations (123) and (124)):

$$\sum \gamma_{lm} \frac{\partial}{\partial x_l} (\pi_{i\delta} \pi_{\kappa\varepsilon} g_{\delta\varepsilon}) \frac{\partial}{\partial x_m} (p_{i\delta'} p_{\kappa\varepsilon'} \gamma_{\delta'\varepsilon'}) . \tag{217}$$

At this point, Einstein simplified the derivation by restricting himself to infinitesimal transformations. With the help of the Kronecker delta, the transformation matrices can then be written as

$$\begin{aligned} \pi_{i\delta} &= \delta_{i\delta} + \pi^x_{i\delta} , \\ p_{i\delta'} &= \delta_{i\delta} + p^x_{i\delta'} . \end{aligned} \tag{218}$$

For such infinitesimal transformations, expression (217) reduces to:

$$\sum \gamma_{lm} \frac{\partial g_{\delta\varepsilon}}{\partial x_l} \frac{\partial \gamma_{\delta\varepsilon}}{\partial x_m} + \text{terms of order } p^x \text{ and } \pi^x . \tag{219}$$

Einstein denoted these infinitesimal correction terms as "transformation infinitely small" ("Transformation unendlich klein"). Expression (219) shows that expression (213) transforms as a scalar, if the sum of all terms of order p^x and π^x vanish.[136] This then is the condition on the transformation matrices for infinitesimal non-autonomous transformations leaving expression (213) invariant.

---

134 The calculation on p. 9R thus provides the first example of a strategy that Einstein routinely availed himself of in investigating the covariance properties of candidate field equations extracted from the Riemann tensor (cf. the discussion in sec. 4.1). Rather than trying to find the class of non-autonomous transformations under which the candidate field equations themselves transform as a tensor, he tried to find the class of non-autonomous transformations under which the coordinate restriction with the help of which these field equations could be constructed out of an object of broad covariance transformed as a tensor. In this case that coordinate restriction is to transformations leaving expression (213) invariant.

135 See the introduction to sec. 4.3 for a discussion of how in general one finds the condition for non-autonomous transformations (i.e., transformations for which the transformation matrices depend on the metric and its derivatives) under which a given expression transforming as a tensor under (ordinary) linear transformations retains such transformation behavior under non-linear transformations.

136 This is a concrete example of the condition $\mathbf{C}(\partial, g, p, \pi) = 0$ in sec. 4.3 (see equation (114)).

This condition consists of four terms, which are obtained by differentiating one of the four transformation matrices in expression (217) and replacing the remaining three by Kronecker deltas:[137]

$$\sum \gamma_{lm} \frac{\partial \pi^x_{i\delta}}{\partial x_l} \frac{\partial \gamma_{i\kappa}}{\partial x_m} g_{\delta\kappa}\,, \tag{220}$$

$$\sum \gamma_{lm} \frac{\partial \pi^x_{\kappa\varepsilon}}{\partial x_l} \frac{\partial \gamma_{i\kappa}}{\partial x_m} g_{i\varepsilon}\,, \tag{221}$$

$$\sum \gamma_{lm} \frac{\partial p^x_{i\delta}}{\partial x_m} \frac{\partial g_{i\kappa}}{\partial x_l} \gamma_{\delta\kappa}\,, \tag{222}$$

$$\sum \gamma_{lm} \frac{\partial p^x_{\kappa\varepsilon}}{\partial x_m} \frac{\partial g_{i\kappa}}{\partial x_l} \gamma_{i\varepsilon}\,. \tag{223}$$

Einstein wrote the last two expressions next to the first two, separating the two pairs by a vertical line.

By rewriting expression (220) as[138]

$$\sum \gamma_{lm} \frac{\partial p_{\delta i}}{\partial x_l} \frac{\partial g_{\delta\kappa}}{\partial x_m} \gamma_{i\kappa} \tag{224}$$

(where the superscript "x" was dropped), Einstein showed that it was equal to expression (222) written next to it.[139] Similarly, expression (221) is equal to expression (223) written next to it. The condition for infinitesimal non-autonomous transformations under which expression (213) transforms as a scalar can thus be written as the vanishing of the sum of expressions (222) and (223):

$$\sum \gamma_{lm} \frac{\partial p^x_{i\delta}}{\partial x_m} \frac{\partial g_{i\kappa}}{\partial x_l} \gamma_{\delta\kappa} + \sum \gamma_{lm} \frac{\partial p^x_{\kappa\varepsilon}}{\partial x_m} \frac{\partial g_{i\kappa}}{\partial x_l} \gamma_{i\varepsilon} = 0\,. \tag{225}$$

An additional condition on the transformation coefficients was the requirement that the transformations be unimodular. Einstein thus turned to the determinant

137 Lacking the Kronecker delta, Einstein indicated this procedure as follows. For the term in which the first of the four coefficients $\pi_{i\delta}$, $\pi_{\kappa\varepsilon}$, $p_{i\delta'}$, and $p_{\kappa\varepsilon'}$ in expression (217) is differentiated he wrote underneath expression (219): "$1\ \kappa = \varepsilon\ \ i = \delta'\ \ \kappa = \varepsilon'$."

138 Equation (224) can be obtained from (220) with the help of (in modernized notation) $\gamma_{i\kappa,m} g_{\delta\kappa} = -\gamma_{i\kappa} g_{\delta\kappa,m}$ and $\pi^x_{i\delta} = -p^x_{\delta i}$. The latter relation follows from $p_{\delta\alpha}\pi_{i\alpha} = (\delta_{\delta\alpha} + p^x_{\delta\alpha})(\delta_{i\alpha} + \pi^x_{i\alpha}) = \delta_{\delta i} + p^x_{\delta i} + \pi^x_{i\delta} = \delta_{\delta i}$.

139 Expressions (222) and (224) turn into one another if the summation indices l and m and the summation indices δ and i are switched.

$$\begin{vmatrix} 1+p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & 1+p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & 1+p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & 1+p_{44} \end{vmatrix} \tag{226}$$

of the transformation matrix for the infinitesimal transformation in equation (218).[140] To first order in $p_{\mu\nu}$, the determinant (226) is equal to

$$1 + p_{11} + p_{22} + p_{33} + p_{44} \,. \tag{227}$$

So the condition that an infinitesimal transformation be unimodular is simply that the trace of $p_{\mu\nu}$ (and of $\pi_{\mu\nu}$) vanish:

$$\begin{aligned} \sum p^x_{\alpha\alpha} &= 0\,, \\ \sum \pi^x_{\alpha\alpha} &= 0\,. \end{aligned} \tag{228}$$

As with the conditions for non-autonomous transformations on pp. 7R and 8R, Einstein made no attempt to find transformation matrices depending on the metric that satisfy the conditions (225) and (228). So the calculation of p. 9R did not lead to the identification of any specific non-autonomous transformations under which the core operator would transform as a tensor.

4.5 *Exploring the Covariance of the Core Operator under Hertz Transformations (10L–12R, 41L–R)*

On the preceding pages, Einstein had investigated the covariance properties of the core operator (130),

$$\sum \frac{\partial}{\partial x_\mu}\left(\gamma_{\mu\nu}\frac{\partial \gamma_{i\kappa}}{\partial x_\nu}\right), \tag{229}$$

and of expression (213),

$$\sum \gamma_{\mu\nu}\frac{\partial g_{i\kappa}}{\partial x_\mu}\frac{\partial \gamma_{i\kappa}}{\partial x_\nu}, \tag{230}$$

by deriving the conditions for non-autonomous transformations under which these objects transform as tensors.[141] None of these investigations had been carried through

140 Following Einstein, we have dropped the subscript "x."

141 See the introduction of sec. 4.3 for discussion of the concept of non-autonomous transformation. On p. 7L, Einstein examined expression (229), deriving conditions (137) and (145) for non-autonomous transformations—finite and infinitesimal, respectively—under which the expression would transform as a tensor. On p. 9R, he examined expression (230), deriving conditions (225) and (228) for infinitesimal unimodular non-autonomous transformations under which this expression would transform as a scalar.

to the end. Errors were made and left uncorrected. Einstein had not even begun the task of finding non-autonomous transformations that would actually be solutions of such conditions. On pp. 10L–11L, however, he made a sustained effort to find infinitesimal unimodular non-autonomous transformations under which a much simpler object, the Hertz expression,

$$\sum \frac{\partial \gamma_{\mu\nu}}{\partial x_\nu}, \tag{231}$$

transforms as a vector. The task at hand was still to determine the covariance properties of the core operator. By focusing on the Hertz expression first, Einstein could split this task into two separate and more manageable tasks. The core operator (229) can be written as the sum of two terms, the first of which contains the Hertz expression:

$$\sum \frac{\partial \gamma_{\mu\nu}}{\partial x_\mu}\frac{\partial \gamma_{i\kappa}}{\partial x_\nu} + \sum \gamma_{\mu\nu}\frac{\partial^2 \gamma_{i\kappa}}{\partial x_\mu \partial x_\nu}. \tag{232}$$

10L–R On pp. 10L–10R, Einstein focused on the first term, deriving the condition for unimodular non-autonomous transformations under which the Hertz expression transforms as a vector. We shall call such transformations “Hertz transformations.” Einstein only dealt with infinitesimal Hertz transformations, which simplified his calculations considerably. The restriction to unimodular transformations in the calculations on these and the following pages indicates that he still wanted to connect the core operator to the second Beltrami invariant φ_2 (see equations (209)–(214)).

On p. 10L, in a first attempt to derive the condition for Hertz transformations, Einstein made a mistake he only discovered on p. 10R after he had already moved on to the second term in expression (232). By that time, he had convinced himself that rotation in Minkowski spacetime is a Hertz transformation. Not surprisingly, therefore, when Einstein had derived the correct condition for Hertz transformations, he carefully checked once more whether the class of Hertz transformation include the important special cases of rotation and uniform acceleration in Minkowski spacetime before returning to the investigation of the second term in expression (232).

11L On p. 11L, Einstein established that the matrices $p_{\mu\nu}$ and $\pi_{\mu\nu}$ for transformations to rotating frames in Minkowski spacetime with infinitesimally small angular velocity do indeed satisfy the conditions for infinitesimal Hertz transformations. He noticed that the transformation matrices for *finite* rotations satisfy these conditions for *infinitesimal* transformations as well. From this he seems to have drawn the erroneous conclusion that finite transformations to rotating coordinates in Minkowski spacetime are also Hertz transformations (see sec. 4.5.2). Einstein immediately recognized, however, that transformations to uniformly accelerating coordinate systems in Minkowski spacetime are not Hertz transformations, not even for infinitesimally small accelerations. He initially thought he could circumvent this problem by modifying the transformation (see sec. 4.5.3). So the class of Hertz transformations initially did seem to include rotation and uniform acceleration in Minkowski spacetime.

On p. 11R, Einstein turned to the second term in expression (232) for the core operator (see sec. 4.5.4). He imposed the "Hertz restriction," i.e., the condition that the Hertz expression vanish.[142] Under this restriction, expression (232) reduces to its second term. Einstein now checked whether this term transforms as a tensor under Hertz transformations. He discovered that it does not. At the bottom of p. 11R, he wrote: "Leads to difficulties" ("Führt auf Schwierigkeiten"). The class of Hertz transformations thus needs to be restricted further. If the metric is set equal to the Minkowski metric in its standard diagonal form, the condition expressing this further restriction reduces to the requirement that the matrices $p_{\mu\nu}$ and $\pi_{\mu\nu}$ of the infinitesimal transformations be anti-symmetric. It turns out that the requirement of anti-symmetry is all that is needed to satisfy the conditions defining the class of infinitesimal Hertz transformations as well. In other words, the core operator transforms as a tensor under all infinitesimal anti-symmetric transformations from an inertial pseudo-Cartesian coordinate system in Minkowski spacetime. 11R

At the top of p. 12L, Einstein therefore made an "attempt" ("Versuch") to find anti-symmetric transformations corresponding to infinitesimal rotation in Minkowski spacetime. He went back to p. 11L and noted how the transformation matrix for infinitesimal rotation needs to be changed to make it anti-symmetric. He realized that this is not feasible. The transformation matrix for infinitesimal uniform acceleration is not anti-symmetric either—be it in its original form, or in the modified form introduced on p. 11L. It seems that at this point Einstein deleted the modified form and accepted that the Hertz restriction rules out the important special case of uniform acceleration in Minkowski spacetime. 12L

The upshot then was that the strategy Einstein adopted on p. 10L to study the covariance of the core operator by splitting it into the two terms (see expression (232)) did not produce any physically interesting transformations, not even infinitesimal non-autonomous ones, under which the core operator would transform as a tensor. Einstein's first reaction was to change the definition of the core operator by inserting an extra factor of $\sqrt{G}$. Such extra factors, however, do not affect the argument on pp. 9R–12L. Rather than going through this argument again, Einstein, as he did on p. 11L, turned his attention to the important test cases of rotation and acceleration in Minkowski spacetime. On pp. 12L–R, he carefully examined rotation (see sec. 4.5.6). 12L–R,
On pp. 12R, 41L–R, he systematically studied (autonomous) infinitesimal unimodular transformations with a view to recovering the transformation to uniformly accelerating frames in Minkowski spacetime. His attempt, however, to find a unimodular transformation corresponding to acceleration failed (sec. 4.5.7). Since the case of uniform acceleration was drawn from his 1912 static theory, Einstein now reexamined an important insight connecting the 1912 theory based on one potential (the variable speed of light) and the metric theory based on ten potentials (the components of the metric tensor). This is the insight that the equation of motion of a test particle in a gravitational field can be obtained from a variational principle with the line element serv- 41L–R

142 See sec. 4.1 for the definition of a coordinate restriction.

ing as the Lagrangian. A consideration of constrained motion on a curved surface probably reassured him that this insight was sound (sec. 4.5.8).

4.5.1 *Deriving the Conditions for Infinitesimal Hertz Transformations (10L–R)*

10L At the top of p. 10L, Einstein wrote down the transformation law for the Hertz expression:

$$\sum \frac{\partial \gamma'_{\mu\nu}}{\partial x'_{\nu}} = \sum \pi_{\nu\sigma} \frac{\partial}{\partial x_{\sigma}} (p_{\mu\alpha} p_{\nu\beta} \gamma_{\alpha\beta}). \tag{233}$$

The structure of this quantity is considerably simpler than that of expression (213) on p. 9R. The condition for non-autonomous transformations under which the Hertz expression transforms as a vector will likewise be considerably simpler than condition (225) on p. 9R, for non-autonomous transformations under which expression (213) transforms as a scalar. Condition (225) is the one referred to in the header of p. 10L: "For comparison with this condition" ("Zum Vergleich mit dieser Bedingung"). Einstein applied his usual two-step procedure to find the condition characterizing what we called "Hertz transformations," i.e., unimodular non-autonomous transformations under which the Hertz expression transforms as a vector.[143]

Einstein rewrote the right-hand side of equation (233) as

$$\sum p_{\mu\alpha} \frac{\partial \gamma_{\alpha\sigma}}{\partial x_{\sigma}} + \sum \gamma_{\alpha\sigma} \frac{\partial p_{\mu\alpha}}{\partial x_{\sigma}}, \tag{234}$$

where he used that the matrices $\pi_{\nu\sigma}$ and $p_{\nu\beta}$ —connected by the line in equation (233)—are the inverse of one another and assumed that $\sum \pi_{\nu\sigma}(\partial p_{\nu\beta}/\partial x_{\sigma})$ vanishes. This assumption holds for infinitesimal unimodular transformations (see note 148 below) but not in general, as Einstein soon came to realize.[144]

Einstein was now ready for the first step of his two-step procedure. For non-autonomous transformations under which the Hertz expression transforms as a vector, the transformation of the Hertz expression from primed to unprimed coordinates is given by:

$$\frac{\partial \gamma_{\kappa\lambda}}{\partial x_{\lambda}} = \pi_{\mu\kappa} \frac{\partial \gamma'_{\mu\nu}}{\partial x'_{\nu}}. \tag{235}$$

Einstein omitted this step and immediately wrote down the second step, expressing the right-hand side in terms of unprimed quantities. Using the abbreviation α_{κ} for the Hertz expression,[145] and using equations (233)–(234), he wrote:

---

143 On p. 10L Einstein only gives the condition for infinitesimal Hertz transformations. See footnote 146 below for a self-contained derivation of the condition for finite transformations. See the introduction to sec. 4.3 for general discussion of the two-step procedure.

144 Einstein had made use of the relation $\sum \pi_{\nu\sigma}(\partial p_{\nu\beta}/\partial x_{\sigma}) = 0$ in situations in which this was not warranted before (see equation (128) and footnote 98).

$$\alpha_\kappa = \sum \pi_{\mu\kappa} p_{\mu\alpha} \frac{\partial \gamma_{\alpha\sigma}}{\partial x_\sigma} + \sum \pi_{\mu\kappa} \frac{\partial p_{\mu\alpha}}{\partial x_\sigma} \gamma_{\alpha\sigma} . \tag{236}$$

The first term on the right-hand side is equal to α_κ . Einstein thus concluded that the condition for non-autonomous transformations under which the Hertz expression transforms as a tensor is:

$$\sum \pi_{\mu\kappa} \frac{\partial p_{\mu\alpha}}{\partial x_\sigma} \gamma_{\alpha\sigma} = 0 . \tag{237}$$

For infinitesimal unimodular transformations this is correct (see note 148 below), but in general, there will be additional terms.[146]

To simplify condition (237), Einstein restricted himself to infinitesimal transformations,

$$\begin{aligned} p_{\mu\nu} &= \delta_{\mu\nu} + p^x_{\mu\nu} \\ \pi_{\mu\nu} &= \delta_{\mu\nu} + \pi^x_{\mu\nu} \end{aligned} \tag{238}$$

with $p^x_{\mu\nu} = -\pi^x_{\nu\mu}$ (see p. 9R and footnote 138). As he wrote underneath the second term on the right-hand side of equation (236): “for infinitesimal transformations” (“für infinitesimale Transformationen”)

$$\sum \frac{\partial p^x_{\kappa\alpha}}{\partial x_\sigma} \gamma_{\alpha\sigma} = 0 . \tag{239}$$

This equation, he continued, “Is a system of four conditions for the p^x if this should always vanish. Furthermore, the determinant should always be equal to 1.

145 On p. 6R, Einstein had used the notation α_κ for a covariant vector (see equation (88)).

146 The condition for *finite* Hertz transformations will be important for understanding Einstein’s argument on p. 11L (see sec. 4.5.2). The derivation of this condition is fully analogous to the derivation of condition (237) for infinitesimal Hertz transformations. For finite transformations, equation (236) needs to be replaced by:

$$\alpha_\kappa = \pi_{\mu\kappa} \alpha'_\mu = \pi_{\mu\kappa} \left(\frac{\partial \gamma'_{\mu\nu}}{\partial x'_\nu} \right) = \pi_{\mu\kappa} \pi_{\nu\sigma} \frac{\partial}{\partial x_\sigma} (p_{\mu\alpha} p_{\nu\beta} \gamma_{\alpha\beta})$$

This gives a sum of three terms

$$\alpha_\kappa = \pi_{\mu\kappa} \pi_{\nu\sigma} p_{\mu\alpha} p_{\nu\beta} \frac{\partial \gamma_{\alpha\beta}}{\partial x_\sigma} + \pi_{\mu\kappa} \pi_{\nu\sigma} \frac{\partial p_{\mu\alpha}}{\partial x_\sigma} p_{\nu\beta} \gamma_{\alpha\beta} + \pi_{\mu\kappa} \pi_{\nu\sigma} p_{\mu\alpha} \frac{\partial p_{\nu\beta}}{\partial x_\sigma} \gamma_{\alpha\beta} ,$$

which can be rewritten as

$$\alpha_\kappa = \alpha_\kappa + \pi_{\mu\kappa} \frac{\partial p_{\mu\alpha}}{\partial x_\sigma} \gamma_{\alpha\sigma} - \frac{\partial \pi_{\nu\sigma}}{\partial x_\sigma} p_{\nu\beta} \gamma_{\kappa\beta} ,$$

where in the third term on the right-hand side the relation $\pi_{\nu\sigma}(\partial p_{\nu\beta}/\partial x_\sigma) = -(\partial \pi_{\nu\sigma}/\partial x_\sigma) p_{\nu\beta}$ was used, which follows from $\partial/\partial x_\sigma(\pi_{\nu\sigma} p_{\nu\beta})$. Hence the condition for *finite* Hertz transformations is:

$$\pi_{\mu\kappa} \frac{\partial p_{\mu\alpha}}{\partial x_\sigma} \gamma_{\alpha\sigma} - \frac{\partial \pi_{\nu\sigma}}{\partial x_\sigma} p_{\nu\beta} \gamma_{\kappa\beta} = 0 .$$

$\sum p^x_{\alpha\alpha} = 0$.” (“Ist für die p^x ein System von 4 Bedingungen, wenn dies stets verschwinden soll. Ferner soll Determinante stets gleich 1 sein. $\sum p^x_{\alpha\alpha} = 0$.”).

As Einstein notes here, for infinitesimal unimodular transformations

$$\det(p) = \det(\pi) = 1 . \tag{240}$$

It follows that the trace of the matrices $p^x_{\mu\nu}$ and $\pi^x_{\mu\nu}$ vanishes:[147,148]

$$\sum p^x_{\alpha\alpha} = \sum \pi^x_{\alpha\alpha} = 0 . \tag{241}$$

Together, equations (239) and (241) thus determine the class of infinitesimal Hertz transformations (i.e., infinitesimal unimodular non-autonomous transformations under which the Hertz expression transforms as a vector).

Raising the question, “Is it possible to have both?” (“Ist beides möglich?”), Einstein now set out to solve equations (239) and (241) for the transformation matrix $p^x_{\mu\nu}$ in the special case that $\gamma_{\mu\nu}$ in equation (239) is a constant diagonal metric (be it Euclidean or Minkowskian). He first considered the two-dimensional and then the three-dimensional case.

Einstein introduced the coordinate transformations

$$\begin{aligned} x' &= X(x, y), \\ y' &= Y(x, y). \end{aligned} \tag{242}$$

The corresponding differentials can be written as:

$$\begin{aligned} dx' &= \frac{\partial X}{\partial x}dx + \frac{\partial X}{\partial y}dy, \\ dy' &= \frac{\partial Y}{\partial x}dx + \frac{\partial Y}{\partial y}dy. \end{aligned} . \tag{243}$$

From this one can read off the matrix

$$p_{\mu\nu} = \begin{pmatrix} \frac{\partial X}{\partial x} & \frac{\partial X}{\partial y} \\ \frac{\partial Y}{\partial x} & \frac{\partial Y}{\partial y} \end{pmatrix} . \tag{244}$$

Condition (241) then reduces to[149]

$$\frac{\partial X}{\partial x} + \frac{\partial Y}{\partial y} = 2 . \tag{245}$$

To the right of equation (245), Einstein specified that the metric be diagonal[150]

147 Cf. equation (227). The condition is illustrated by a simple example written underneath it:

$$\begin{matrix} 1+\varepsilon_1 & & \\ & 1+\varepsilon_2 & \end{matrix} \quad 1 + \varepsilon_1 + \varepsilon_2 = 1$$

$$\begin{matrix} \gamma_{11} = 1 & \gamma_{22} = 1 \\ \multicolumn{2}{c}{\gamma_{12} = 0} \end{matrix} \quad . \tag{246}$$

Inserting these values into condition (239), one finds:

$$\begin{aligned} \frac{\partial p_{1\alpha}^{x}}{\partial x_{\sigma}}\gamma_{\alpha\sigma} &= \frac{\partial p_{11}^{x}}{\partial x_{1}}\gamma_{11} + \frac{\partial p_{12}^{x}}{\partial x_{2}}\gamma_{22} = \frac{\partial p_{11}^{x}}{\partial x_{1}} + \frac{\partial p_{12}^{x}}{\partial x_{2}} = 0, \\ \frac{\partial p_{2\alpha}^{x}}{\partial x_{\sigma}}\gamma_{\alpha\sigma} &= \frac{\partial p_{21}^{x}}{\partial x_{1}}\gamma_{11} + \frac{\partial p_{22}^{x}}{\partial x_{2}}\gamma_{22} = \frac{\partial p_{21}^{x}}{\partial x_{1}} + \frac{\partial p_{22}^{x}}{\partial x_{2}} = 0. \end{aligned} \tag{247}$$

Inserting equation (244) for $p_{\mu\nu}$ into equation (247), one finds:

$$\begin{aligned} \frac{\partial^2 X}{\partial x^2} + \frac{\partial^2 X}{\partial y^2} &= 0, \\ \frac{\partial^2 Y}{\partial x^2} + \frac{\partial^2 Y}{\partial y^2} &= 0. \end{aligned} \tag{248}$$

Einstein omitted the terms with p_{12}^{x} and p_{21}^{x} in equation (247) at this point and used the erroneous set of equations

$$\frac{\partial^2 X}{\partial x^2} = 0, \quad \frac{\partial^2 Y}{\partial y^2} = 0 . \tag{249}$$

---

148 It turns out that for infinitesimal unimodular transformations $\sum \pi_{\nu\sigma}(\partial p_{\nu\beta}/\partial x_{\sigma}) = 0$ and that equation (234) is correct. This, in turn, means that conditions (239)–(241) correctly specify the class of infinitesimal Hertz transformations. The aborted calculation in the top-right corner of p. 10L suggests that Einstein realized this. He began by writing down the term he had omitted in going from equation (233) to equation (234):

$$\sum \pi_{\nu\sigma} p_{\mu\alpha} \frac{\partial p_{\nu\beta}}{\partial x_{\sigma}} \gamma_{\alpha\beta} .$$

Underneath this expression, he wrote:

$$\sum \frac{\partial \pi_{\nu\sigma}}{\partial x_{\sigma}} .$$

One can indeed transfer the derivative operator $\partial/\partial x_{\sigma}$ in expression above from $p_{\nu\beta}$ to $\pi_{\nu\sigma}$. This follows from:

$$0 = \frac{\partial}{\partial x_{\sigma}}(\pi_{\nu\sigma} p_{\nu\beta}) = \pi_{\nu\sigma}\frac{\partial p_{\nu\beta}}{\partial x_{\sigma}} + \frac{\partial \pi_{\nu\sigma}}{\partial x_{\sigma}} p_{\nu\beta} .$$

Condition (241) for infinitesimal unimodular transformations implies that $\partial \pi_{\nu\sigma}/\partial x_{\sigma}$ vanishes:

$$\frac{\partial \pi_{\nu\sigma}}{\partial x_{\sigma}} = \frac{\partial \pi_{\nu\sigma}^{x}}{\partial x_{\sigma}} = -\frac{\partial p_{\sigma\nu}^{x}}{\partial x_{\sigma}} = -\frac{\partial p_{\sigma\sigma}^{x}}{\partial x_{\nu}} = 0 .$$

where we used that $\pi_{\nu\sigma}^{x} = p_{\sigma\nu}^{x}$ (see footnote 138) and that

$$\frac{\partial p_{\sigma\nu}^{x}}{\partial x_{\sigma}} = \frac{\partial}{\partial x_{\sigma}}\left(\frac{\partial x'_{\sigma}}{\partial x_{\nu}}\right) = \frac{\partial}{\partial x_{\nu}}\left(\frac{\partial x'_{\sigma}}{\partial x_{\sigma}}\right) = \frac{\partial p_{\sigma\sigma}^{x}}{\partial x_{\nu}} .$$

This concludes the proof that the expression in the top-right corner of p. 10L vanishes for infinitesimal unimodular transformations and that the transition from equation (233) to equation (234) is justified in this case.

149 Einstein originally wrote X' and Y' and then deleted the primes.

Einstein eventually discovered his error. The calculation on pp. 10L–R based on equations (249) is deleted and a fresh start is made at the bottom of p. 10R, based on the correct set of equations (248) (see equations (290)–(291) below). But first we shall discuss the deleted calculations on pp. 10L–R

Integrating equations (249), Einstein arrived at

$$\frac{\partial X}{\partial x} = \psi(y) \qquad \frac{\partial Y}{\partial y} = \chi(x) \tag{250}$$

with arbitrary functions ψ and χ. The unimodularity condition (245) requires that:

$$\psi(y) + \chi(x) = 2\,. \tag{251}$$

Einstein concluded that ψ and χ would "both [be] constant" ("beide konstant."). So the only infinitesimal Hertz transformations in the two-dimensional case are linear transformations.

In an attempt to find non-linear transformations, Einstein turned to the next simplest case, a Euclidean space of "three dimensions" ("drei Dimensionen") with a constant diagonal metric. Starting point of the calculation is the coordinate transformation (cf. equation (242)):

$$\begin{aligned} x' &= X(x, y, z), \\ y' &= Y(x, y, z), \\ z' &= Z(x, y, z). \end{aligned} \tag{252}$$

The matrix $p_{\mu\nu}$ is thus given by (cf. equation (244)):

$$p_{\mu\nu} = \begin{pmatrix} \frac{\partial X}{\partial x} & \frac{\partial X}{\partial y} & \frac{\partial X}{\partial z} \\ \frac{\partial Y}{\partial x} & \frac{\partial Y}{\partial y} & \frac{\partial Y}{\partial z} \\ \frac{\partial Z}{\partial x} & \frac{\partial Z}{\partial y} & \frac{\partial Z}{\partial z} \end{pmatrix}. \tag{253}$$

The analogue of equation (245), expressing unimodularity, is

$$\frac{\partial X}{\partial x} + \frac{\partial Y}{\partial y} + \frac{\partial Z}{\partial z} = 3\,, \tag{254}$$

while the analogues of conditions (249)—again with omission of non-diagonal terms of $p^x_{\mu\nu}$ — are

$$\frac{\partial^2 X}{\partial x^2} = 0, \qquad \frac{\partial^2 Y}{\partial y^2} = 0, \qquad \frac{\partial^2 Z}{\partial z^2} = 0\,. \tag{255}$$

150 He first wrote down the Euclidean metric $\mathrm{diag}(1, 1)$ and later changed it to the Minkowski metric $\mathrm{diag}(-1, 1)$.

Integrating these equations, Einstein wrote (cf. equation (250)):

$$\frac{\partial X}{\partial x} = \psi_1(y, z)\,, \tag{256}$$

$$\frac{\partial Y}{\partial y} = \psi_2(x, z)\,, \tag{257}$$

$$\frac{\partial Z}{\partial z} = \psi_3(x, y)\,. \tag{258}$$

Inserting these expressions into equation (254) and introducing

$$\psi_i^x \equiv \psi_i - 1\ \ , \tag{259}$$

he obtained:

$$\psi_1^x(y, z) + \psi_2^x(x, z) + \psi_3^x(x, y) = 0\,. \tag{260}$$

Taking the derivative of equation (260) with respect to z and dropping the superscript x, Einstein found, at the bottom of p. 10L:

$$\frac{\partial\psi_1}{\partial z} + \frac{\partial\psi_2}{\partial z} = 0\,. \tag{261}$$

At the top of p. 10R, Einstein integrated this last equation, writing the result as 10R

$$\psi_1 + \psi_2 = \chi_3(x, y)\,. \tag{262}$$

Taking the derivative of equation (260) with respect to x and y, one similarly finds

$$\psi_2 + \psi_3 = \chi_1(y, z)\,, \tag{263}$$

and

$$\psi_1 + \psi_3 = \chi_2(x, z)\,, \tag{264}$$

respectively. Einstein explicitly wrote down equation (263), but not equation (264).

Since both $\psi_1(y, z)$ and $\psi_2(z, x)$ depend on z, while their sum, according to (262), does not, Einstein could write them in the form

$$\psi_1(y, z) = \psi_1(y) + \zeta\,, \tag{265}$$

$$\psi_2(z, x) = \psi_2(x) - \zeta\,, \tag{266}$$

where ζ is some function of z. He used equation (260) to write ψ_3 as:

$$\psi_3(x, y) = -\psi_1(y) - \psi_2(x)\,. \tag{267}$$

Now insert equations (265)–(267) into the equations (256)–(258), keeping in mind that *all* ψ_is in equations (265)–(267) are actually ψ_i^xs related to the ψ_is in equations (256)–(258) through $\psi_i = 1 + \psi_i^x$.

$$\frac{\partial X}{\partial x} = 1 + \psi_1^x(y) + \zeta\,, \tag{268}$$

$$\frac{\partial Y}{\partial y} = 1 + \psi_2^x(x) - \zeta\,, \tag{269}$$

$$\frac{\partial Z}{\partial z} = 1 - \psi_1^x(y) - \psi_2^x(x)\,. \tag{270}$$

Integrating these equations, one finds[151]

$$X = x(1 + \psi_1^x(y) + \zeta) + \omega_1(y, z)\,, \tag{271}$$

$$Y = y(1 + \psi_2^x(x) - \zeta) + \omega_2(x, z)\,, \tag{272}$$

$$Z = z(1 - \psi_1^x(y) - \psi_2^x(x)) + \omega_3(x, y)\,. \tag{273}$$

Einstein now considered a special case. Writing “Specified” (“Spezialisiert”), he set $\psi_1^x(y)$, $\psi_2^x(x)$, and the quantities ω_i to zero and $\zeta = \alpha z$, with $\alpha \ll 1$. Equations (271)–(273) then reduce to:[152]

$$\begin{aligned} X &= x + \alpha xz, \\ Y &= y - \alpha yz, \\ Z &= z. \end{aligned} \tag{274}$$

Next to this transformation, Einstein wrote: “Is torsion and, in the case $z = t$, uniform rotation. Torsion very special case.” (“Ist Torsion & im Falle $z = t$ gleichformige Drehung. Torsion ganz spezieller Fall.”). The transformation (274) is indeed a very special case of the much more general transformation (271)–(273); it does not, however, correspond to torsion, nor, with $z = t$, to rotation.

The transformation setting an inertial pseudo-Cartesian coordinate system in Minkowski spacetime rotating with angular velocity ω is given by:

$$\begin{aligned} x' &= x\cos\omega t + y\sin\omega t, \\ y' &= -x\sin\omega t + y\cos\omega t, \\ t' &= t. \end{aligned} \tag{275}$$

---

151 Instead of equations (271)–(273), Einstein wrote down the equations:

$$\begin{aligned} X &= x(\psi_1(y) + \zeta) = x(-\eta + \zeta) + \omega_1(y, z), \\ Y &= y(\psi_2(x) - \zeta) = y(-\zeta + \xi) + \omega_2(z, x), \\ Z &= -z(\psi_1(y) - \psi_2(z)) = z(-\xi + \eta) + \omega_3(x, y), \end{aligned}$$

where $\xi \equiv \psi_2(x)$ and $\eta \equiv -\psi_1(y)$. These equations contain a number of mistakes. In the third line, $\psi_1(y) - \psi_2(z)$ should be $\psi_1(y) + \psi_2(x)$. In all three lines, ψ_i^x and ψ_i are conflated and the integration constants ω_i are missing after the first equality sign. Einstein also defined and then deleted the quantities $\delta x \equiv \psi_1$ and $\delta y \equiv \psi_2$ (cf. equation (373) at the bottom of p. 12R where δx is essentially defined as dX.).

152 As a consequence of Einstein’s conflation of ψ_i^x and ψ_i (see the preceding note), the notebook has $X = \alpha xz$, $Y = -\alpha yz$ and $Z = 0$ instead of equations (274).

For infinitesimal ω, this transformation reduces to:

$$\begin{aligned} x' &= x + \omega y t, \\ y' &= y - \omega x t, \\ t' &= t, \end{aligned} \tag{276}$$

which is clearly not the same as transformation (274) with $z = t$. Einstein did not realize this until later.

He did, however, realize at this point that he had made an error in evaluating condition (239) for the special case of a constant diagonal metric.[153] Under the header, "Conditions of integrability" ("Integrabilitätsbedingungen"), he partly corrected this error:[154]

$$\frac{\partial p^x_{xx}}{\partial x} + \frac{\partial p^x_{yx}}{\partial y} + \cdot = \frac{\partial}{\partial x}\frac{\partial X}{\partial x} + \frac{\partial}{\partial y}\frac{\partial Y}{\partial x} + \frac{\partial}{\partial z}\frac{\partial Z}{\partial x} = 0\,. \tag{277}$$

This condition, however, is still satisfied by the solutions of the erroneous conditions (255) that Einstein had used up to this point.[155] A line drawn from the header above equation (277) to these solutions (see equations (271)–(273) and footnote 151), suggests that Einstein actually checked this. This would explain why he initially simply proceeded with the next part of his investigation of the transformation properties of the core operator.

On the next line he wrote the contraction of the core operator (130) with some arbitrary covariant tensor $T_{i\kappa}$:[156]

$$\sum T_{i\kappa}\frac{\partial}{\partial x_\mu}\left(\gamma_{\mu\nu}\frac{\partial \gamma_{i\kappa}}{\partial x_\nu}\right) = \text{Skalar}\,. \tag{278}$$

The core operator will transform as a tensor under all transformations under which its contraction with $T_{i\kappa}$ transforms as a scalar. When the Hertz restriction (i.e., $\sum \partial\gamma_{\mu\nu}/\partial x_\mu = 0$) is imposed, the contraction (278) reduces to:

153 See equation (248) for the correct form of the condition in the two-dimensional case and equation (255) for the incorrect form in the three dimensional case used by Einstein in his derivation of the general transformation (271)–(273).

154 The crucial residual error in the equation below is that it has $\partial p^x_{yx}/\partial y$ instead of $\partial p^x_{xy}/\partial y$. Inserting $\gamma_{\mu\nu} = \mathrm{diag}(1, 1, 1)$ into condition (239), $\sum(\partial p^x_{\kappa\alpha}/\partial x_\sigma)\gamma_{\alpha\sigma} = 0$, one finds $\frac{\partial p^x_{xx}}{\partial x} + \frac{\partial p^x_{xy}}{\partial y} + \frac{\partial p^x_{xz}}{\partial z} = \frac{\partial^2 X}{\partial x^2} + \frac{\partial^2 X}{\partial y^2} + \frac{\partial^2 X}{\partial z^2}$ for the $\kappa = 1$ component instead of equation (277) in the notebook.

155 Inserting equations (271)–(273) into the three terms on the right-hand side of equation (277), one finds zero, $\partial\psi^x_2(x)/\partial x$, and $-\partial\psi^x_2(x)/\partial x$, respectively.

156 So far Einstein had used Latin letters for contravariant objects (see, e.g., equation (146) for a_μ and $T_{\mu\nu}$ on p. 7R)

$$\sum T_{i\kappa}\gamma_{\mu\nu}\frac{\partial^2\gamma_{i\kappa}}{\partial x_\mu \partial x_\nu}. \tag{279}$$

The Hertz restriction is invariant under Hertz transformations. Hence, if Einstein could show that expression (279) transforms as a scalar under infinitesimal Hertz transformations, he would have shown that the core operator transforms as a tensor under such transformations. This is precisely the problem that Einstein takes up at the top of p. 11R, where he raises the question whether expression (279) is a scalar. On p. 10R, however, he did not proceed beyond equation (278).

Expressions (277) and (278) were deleted and Einstein returned to condition (239) for infinitesimal Hertz transformations. It was probably at this point that in the upper-right corner of p. 10R, he wrote down the $\kappa = 1$ component of condition (239) for the special case of a diagonal Euclidean metric:

$$\frac{\partial p_{11}}{\partial x_1}+\frac{\partial p_{12}}{\partial x_2}+\frac{\partial p_{13}}{\partial x_3} = 0. \tag{280}$$

Inserting equation (253) for $p_{\mu\nu}$ into equation (280), one finds:

$$\frac{\partial^2 X}{\partial x^2}+\frac{\partial^2 X}{\partial y^2}+\frac{\partial^2 X}{\partial z^2} = 0. \tag{281}$$

For the $\kappa = 2$ and $\kappa = 3$ components, one similarly finds:

$$\frac{\partial^2 Y}{\partial x^2}+\frac{\partial^2 Y}{\partial y^2}+\frac{\partial^2 Y}{\partial z^2} = 0, \tag{282}$$

$$\frac{\partial^2 Z}{\partial x^2}+\frac{\partial^2 Z}{\partial y^2}+\frac{\partial^2 Z}{\partial z^2} = 0. \tag{283}$$

Einstein, however, still continued to use the transformations (271)–(273) he found as a solution of conditions (255), an erroneous version of conditions (281)–(283) and similar conditions for Y and Z. Inserting equation (271) for X into equation (281), he found:

$$\Delta X = 0 = \frac{\partial^2\omega_1}{\partial y^2}+\frac{\partial^2\omega_1}{\partial z^2}. \tag{284}$$

The solution of this harmonic differential equation is:

$$\omega_1 = \alpha+\beta y+\gamma z+\delta yz+\varepsilon(y^2-z^2). \tag{285}$$

He added one more line,[157]

$$\delta x = \text{konst.}+\alpha_1 z\langle\delta y\rangle+\alpha_2 y\langle\delta z\rangle+\quad, \tag{286}$$

before he realized that the replacement of conditions (255) by conditions (281)–(283) invalidated much of the subsequent calculation on pp. 10L–R. He deleted this calcu-

157 Cf. equation (373) at the bottom of p. 12R where δx is essentially defined as dX.

lation—equations (245)–(286), with the exception of equation (280) at the top of p. 10R—and made a fresh start.

Einstein went back to the two-dimensional case examined on p. 10L. The unimodularity condition (241) for this case can still be written as (see equation (245)):

$$\frac{\partial X}{\partial x} + \frac{\partial Y}{\partial y} = 2 , \tag{287}$$

He then turned to condition (239) for non-autonomous transformations with $\gamma_{\mu\nu} = \delta_{\mu\nu}$ (see equations (247)–(248)). For the first component of condition (239), he now correctly wrote:

$$\frac{\partial p_{11}}{\partial x_1} + \frac{\partial p_{12}}{\partial x_2} = 0 . \tag{288}$$

Substituting

$$p_{11} = \frac{\partial X}{\partial x} \qquad p_{12} = \frac{\partial X}{\partial y} \tag{289}$$

(cf. equation (244)) into in equation (288), he found:

$$\frac{\partial^2 X}{\partial x^2} + \frac{\partial^2 X}{\partial y^2} = 0 . \tag{290}$$

For the second component of condition (239), he similarly wrote

$$\frac{\partial^2 Y}{\partial x^2} + \frac{\partial^2 Y}{\partial y^2} = 0 . \tag{291}$$

Equations (290) and (291) are easily solved:

$$X = \alpha_1 xy + \alpha_2 (x^2 - y^2) + x , \tag{292}$$

$$Y = \beta_1 xy + \beta_2 (x^2 - y^2) + y . \tag{293}$$

Einstein started to calculate $p_{11} = \partial X / \partial x$ for these coordinate transformations, but proceeded no further than

$$p_{11} = \alpha_1 y + 2 . \tag{294}$$

He then substituted equations (292)–(293) into the unimodularity condition (287) and found

$$\alpha_1 y + 2\alpha_2 x + \beta_1 x - 2\beta_2 y = 0 . \tag{295}$$

The constants of integration α_1, α_2, β_1, and β_2 thus have to satisfy

$$\beta_1 = -2\alpha_2 , \tag{296}$$

$$\alpha_1 = 2\beta_2 . \tag{297}$$

Equations (292)–(293) with conditions (296)–(297) give the general form for infinitesimal Hertz transformations for the special case of the two-dimensional Euclidean metric $\gamma_{\mu\nu} = \delta_{\mu\nu}$.

11L

4.5.2 *Checking Whether Rotation in Minkowski spacetime Is a Hertz Transformation (11L)*

On p. 11L Einstein checked whether transformations to uniformly rotating and uniformly accelerating frames in Minkowski spacetime are included in the class of (infinitesimal) Hertz transformations that he had studied on pp. 10L–R. At the top of the page, under the header "Rotation" ("Drehung"), he dealt with the former; under a horizontal line in the middle of the page, under the header "Acceleration" ("Beschleunigung"), he dealt with the latter. Einstein concluded that both infinitesimal and finite rotations in Minkowski spacetime are Hertz transformations. This is true for infinitesimal but not for finite rotations. Einstein also found, however, that uniform accelerations in Minkowski spacetime, whether infinitesimal or finite, are not Hertz transformations (see sec. 4.5.3).

The condition for some unimodular non-autonomous transformation to be a Hertz transformation (i.e., a unimodular transformation under which the Hertz expression, $\partial\gamma_{\mu\nu}/\partial x_\nu$, transforms as a vector) is:

$$\pi_{\mu\kappa}\frac{\partial p_{\mu\alpha}}{\partial x_\sigma}\gamma_{\alpha\sigma} - \frac{\partial \pi_{\nu\sigma}}{\partial x_\sigma}p_{\nu\beta}\gamma_{\kappa\beta} = 0 \tag{298}$$

(see footnote 146). For infinitesimal transformations the second term vanishes and this condition reduces to:

$$\frac{\partial p_{\kappa\alpha}}{\partial x_\sigma}\gamma_{\alpha\sigma} = 0 \tag{299}$$

(see equation (239)). For infinitesimal transformation the condition for unimodularity reduces to the requirement that the trace of $p^x_{\mu\nu} \equiv p_{\mu\nu} - \delta_{\mu\nu}$ and $\pi^x_{\mu\nu} \equiv \pi_{\mu\nu} - \delta_{\mu\nu}$ vanish:

$$p^x_{\alpha\alpha} = \pi^x_{\alpha\alpha} = 0\,. \tag{300}$$

(see equation (241)).

At the top of p. 11L, next to a drawing indicating rotation, Einstein wrote down the transformation from an inertial pseudo-Cartesian coordinate system in Minkowski spacetime to a coordinate system uniformly rotating with angular velocity ω:[158]

$$\begin{aligned} x' &= x\cos\omega t + y\sin\omega t \\ y' &= -x\sin\omega t + y\cos\omega t \\ t' &= t\,. \end{aligned} \tag{301}$$

158 Anticipating the next step in the calculation, Einstein wrote $dt' = dt$ instead of $t' = t$.

The differentials of the coordinates transform as:

$$\begin{aligned} dx' &= \cos\omega t\ dx + \sin\omega t\ dy + (-x\sin\omega t + y\cos\omega t)\omega dt \\ dy' &= -\sin\omega t\ dx + \cos\omega t\ dy + (-x\cos\omega t - y\sin\omega t)\omega dt \\ dt' &= \quad 0 \quad dx + \quad 0 \quad dy \qquad\qquad + \qquad\qquad dt. \end{aligned} \tag{302}$$

For an infinitesimal rotation, equation (302) reduces to:

$$\begin{aligned} dx' &= \qquad dx + \omega t\ dy + y\omega\ dt \\ dy' &= -\omega t\ dx + \qquad dy - x\omega\ dt \\ dt' &= \qquad\qquad\qquad\qquad dt, \end{aligned} \tag{303}$$

from which one can read off the components of the transformation matrix $p_{\alpha\beta}$, the "table of p" ("Tabelle der p"):[159]

$$p_{\alpha\beta} = \begin{pmatrix} 1 & \omega t & y\omega \\ -\omega t & 1 & -x\omega \\ 0 & 0 & 1 \end{pmatrix}. \tag{304}$$

For this matrix, the trace $p^x_{\alpha\alpha}$ vanishes, and so does $\pi^x_{\alpha\alpha}$.[160] Hence, condition (300) is satisfied. Inserting the diagonal Minkowski metric for $\gamma_{\alpha\beta}$ in equation (299), one readily verifies that $p_{\alpha\beta}$ in equation (304) satisfies conditions (299) as well:

$$\frac{\partial p_{\kappa 1}}{\partial x} + \frac{\partial p_{\kappa 2}}{\partial y} - \frac{\partial p_{\kappa 3}}{\partial t} = 0 \tag{305}$$

for $\kappa = 1, 2, 3$. It follows that infinitesimal rotations are indeed infinitesimal Hertz transformations. Next to his "table of p" Einstein accordingly wrote: "Correct" ("Stimmt").

In what looks like a later addition to the page, Einstein checked whether finite rotations are Hertz transformations too. In the lower-left part of p. 11L, Einstein wrote down the "table of the p" ("Tafel der p") for a finite rotation, which can be read off from equation (302):[161]

$$p_{\mu\nu} = \begin{pmatrix} \cos\omega t & \sin\omega t & (-x\sin\omega t + y\cos\omega t)\omega \\ -\sin\omega t & \cos\omega t & (-x\cos\omega t - y\sin\omega t)\omega \\ 0 & 0 & 1 \end{pmatrix}. \tag{306}$$

Inverting this matrix, one finds the corresponding "table of the π" ("Tafel der π"):[162]

159 The 13-component originally had an additional term $-x\omega t$. This term should be $-x\omega^2 t$ and can therefore be neglected. The 23-component likewise has a deleted term $-y\omega t$, which should be $-y\omega^2 t$. The expressions $-y\omega$ and $x\omega$ underneath the 31- and 32-components were added later (see equation (343) in sec. 4.5.5).

160 It follows from $p_{\mu\alpha}\pi_{\nu\alpha} = \delta_{\mu\nu}$ that $\pi^x_{\mu\nu} = -p^x_{\nu\mu}$ (see footnote 138).

161 Einstein omitted a factor ω in p_{13} and p_{23}.

$$\pi_{\mu\nu} = \begin{pmatrix} \cos\omega t & \sin\omega t & 0 \\ -\sin\omega t & \cos\omega t & 0 \\ -y\omega & x\omega & 1 \end{pmatrix}. \tag{307}$$

Einstein noted that the second term in condition (298) for finite Hertz transformations vanishes for this transformation matrix. He noted that

$$\sum \frac{\partial \pi_{\mu\nu}}{\partial x_\nu} = 0 \tag{308}$$

is "always fulfilled" ("immer erfüllt"). Since the transformation to rotating coordinates is also unimodular— $\det(p_{ik}) = 1$ for the matrix in equation (306)—Einstein presumably concluded that finite rotations, like infinitesimal ones, are Hertz transformations. This conclusion, however, is not warranted.

The problem is that the *first* term of equation (298) does not vanish for finite rotations. Substituting the diagonal Minkowski metric for $\gamma_{\alpha\sigma}$ in the first term of equation (298), one finds

$$\pi_{\mu\kappa}\left(\frac{\partial p_{\mu 1}}{\partial x} + \frac{\partial p_{\mu 2}}{\partial y} - \frac{\partial p_{\mu 3}}{\partial t}\right). \tag{309}$$

Since $\partial p_{\mu 1}/\partial x = \partial p_{\mu 2}/\partial y = \partial p_{33}/\partial t = 0$, this expression reduces to

$$-\pi_{1\kappa}\frac{\partial p_{13}}{\partial t} - \pi_{2\kappa}\frac{\partial p_{23}}{\partial t}. \tag{310}$$

This expression does not vanish for the coefficients $p_{\mu\nu}$ in equation (306) for a finite rotation. For $\kappa = 1$, for instance, it is equal to $x\omega^2$.[163] It follows that condition (298)

---

162 The notebook has $\pi_{31} = y$ and $\pi_{32} = x$. The inversion is done with the formula $\pi_{ik} = (-1)^{i+k}(\Delta_{ik}/p)$, where Δ_{ik} is the co-factor of p_{ik} and $p \equiv \det(p_{ik}) = 1$. The expression for π_{31} is found as follows:

$$\begin{aligned} \pi_{31} &= \begin{vmatrix} \sin\omega t & (-x\sin\omega t + y\cos\omega t)\omega \\ \cos\omega t & (-x\cos\omega t - y\sin\omega t)\omega \end{vmatrix} \\ &= \omega\{\sin\omega t(-x\cos\omega t - y\sin\omega t) - \cos\omega t(-x\sin\omega t + y\cos\omega t)\} \\ &= -\omega y \end{aligned}$$

A completely analogous calculation gives $\pi_{32} = \omega x$. Inserting equations (306) and (307) into $p_{\mu\alpha}\pi_{\nu\alpha}$, one readily verifies that this gives $\delta_{\mu\nu}$.

163 Using equation (306), one finds that

$$\frac{\partial p_{13}}{\partial t} = \omega^2(-x\cos\omega t - y\sin\omega t), \quad \frac{\partial p_{23}}{\partial t} = \omega^2(x\sin\omega t - y\cos\omega t)$$

Inserting these expressions into the $\kappa = 1$ component of expression (310) and using equation (307), one finds

$$-\pi_{11}\frac{\partial p_{13}}{\partial t} - \pi_{21}\frac{\partial p_{23}}{\partial t} = -\omega^2\{\cos\omega t(-x\cos\omega t - y\sin\omega t) - \sin\omega t(x\sin\omega t - y\cos\omega t)\},$$

which is equal to $\omega^2 x$.

is not satisfied in the case of finite rotations in Minkowski spacetime, which means that these transformations cannot be Hertz transformations.[164]

4.5.3 *Checking Whether Acceleration in Minkowski Spacetime is a Hertz Transformation (11L)* 11L

On the lower half of p. 11L, under the heading "Acceleration" ("Beschleunigung"), Einstein checked whether a transformation to a uniformly accelerated frame in Minkowski spacetime is a Hertz transformation. He started from the transformation equations that he had found for this case in the course of the work on his theory for static gravitational fields:[165]

$$\begin{aligned} \xi &= x + \frac{c}{2}\frac{dc}{dx}t^2 \\ \tau &= ct, \end{aligned} \tag{311}$$

where c is the variable speed of light that served as the gravitational potential in Einstein's static theory. Einstein assumed c to be of the form[166]

$$c = e^{ax}. \tag{312}$$

In the notebook, Einstein used (x, t) instead of (ξ, τ) and (x', t') instead of (x, t). The transformation (311) then becomes:

$$\begin{aligned} x &= x' + \frac{c(x')}{2}\frac{dc(x')}{dx'}t'^2 \\ t &= ct'. \end{aligned} \tag{313}$$

With the ansatz (312)—in terms of x' rather than x—the transformation (313) turns into:

$$\begin{aligned} x &= x' + \frac{a}{2}e^{2ax'}t'^2 \\ t &= e^{ax'}t'. \end{aligned} \tag{314}$$

Inverting this transformation while neglecting terms quadratic in a and smaller, Einstein found

---

164 A more direct way to arrive at this conclusion is to note that the Hertz expression, $\partial\gamma_{\mu\nu}/\partial x_\nu$, vanishes for the standard diagonal Minkowski metric but not for the Minkowski metric in rotating coordinates. The contravariant form of the latter is given by:

$$\gamma_{\mu\nu} = \begin{pmatrix} -1+\omega^2y^2 & -\omega^2xy & 0 & \omega y \\ -\omega^2xy & -1+\omega^2x^2 & 0 & -\omega x \\ 0 & 0 & -1 & 0 \\ \omega y & -\omega x & 0 & 1 \end{pmatrix}$$

For this metric, the Hertz expression is: $\partial\gamma_{\mu\nu}/\partial x_\nu = (-\omega^2x, -\omega^2y, 0, 0) \neq 0$.

165 See (Einstein 1912b, 456).

166 Initially, Einstein wrote $c = c_0e^{ax}$ but then deleted the factor c_0.

$$\begin{aligned} x' &= x - \frac{a}{2}t^2 \\ t' &= t(1-ax). \end{aligned} \tag{315}$$

As he had done for rotation, Einstein considered the infinitesimal transformation

$$\begin{aligned} dx' &= dx - atdt \\ dt' &= -atdx + (1-ax)dt. \end{aligned} \tag{316}$$

From the transformation matrix

$$p_{ij} = \begin{pmatrix} 1 & -at \\ -at & 1-ax \end{pmatrix}, \tag{317}$$

one immediately sees that its elements do not satisfy the two conditions (239) and (241) for infinitesimal Hertz transformations (see also equations (299) and (300)). Inserting $\gamma_{\mu\nu} = \delta_{\mu\nu}$ in equation (239), we find that the $\kappa = 1$ component is:

$$\frac{\partial p^x_{11}}{\partial x_1} + \frac{\partial p^x_{12}}{\partial x_2} = -a \neq 0 \quad , \tag{318}$$

Condition (241) is not satisfied either:

$$\sum p^x_{\alpha\alpha} = -ax \neq 0 \,. \tag{319}$$

Both problems could be fixed by setting:

$$p_{11} = 1 + ax \,. \tag{320}$$

It looks as if Einstein considered this modification. He changed the first line of equation (316) to

$$dx' = (1 + \langle 2 \rangle x)dx - atdt \,, \tag{321}$$

and added the remark: "is also correct for a suitable shift of scale" ("stimmt auch bei geeigneter Massstabverschiebung"). He subsequently deleted this remark.

For the time being, however, Einstein seems to have been satisfied that he could ensure in this fashion that transformations to uniformly accelerating frames in Minkowski spacetime would be included in the class of Hertz transformations. The modification (320) of the transformation matrix (316), however, is not allowed. The form of the transformation (311), of which transformation (316) is a special case, was derived from the equivalence of the propagation of light in the two systems (ξ, τ) and (x, t)[167]

$$d\xi^2 - d\tau^2 = dx^2 - c^2dt^2 \,. \tag{322}$$

With the adjustment (320) this fundamental equation would no longer be valid.

167 See (Einstein 1912a, sec. 1).

4.5.4 *Trying to Find Hertz Transformations under which the Core Operator Transforms as a Tensor (11R)* 11R

Having satisfied himself for the time being that the class of Hertz transformations includes transformations to rotating and uniformly accelerating frames in Minkowski spacetime, Einstein once again turned his attention to the core operator. Picking up on an idea that had made its first appearance on p. 10R (see equations (278)–(279)), Einstein examined under which transformations the contraction of the core operator and some arbitrary covariant second-rank tensor would transform as a scalar.

On p. 10R, Einstein had written this contraction as (see equation (278)):

$$\sum T_{i\kappa} \frac{\partial}{\partial x_\mu}\left(\gamma_{\mu\nu} \frac{\partial \gamma_{i\kappa}}{\partial x_\nu}\right) = \text{Skalar} . \tag{323}$$

At the top of p. 11R Einstein wrote down a very similar expression (down to the labeling of the indices),

$$\sum T^x_{i\kappa} \gamma_{\mu\nu} \frac{\partial}{\partial x_\mu} \frac{\partial}{\partial x_\nu}(\gamma_{i\kappa}) , \tag{324}$$

and asked: “Is this a scalar?” (“Ist dies ein Skalar?”). If $T_{i\kappa}$ in equation (323) is replaced by $T^x_{i\kappa}$,[168] and, more importantly, if the Hertz restriction,

$$\sum \frac{\partial \gamma_{\mu\nu}}{\partial x_\nu} = 0 , \tag{325}$$

is imposed, the expressions on 10R and 11R are equivalent. So if expression (324) transforms as a scalar under Hertz transformations, so will expression (323). Einstein set out to find what further coordinate restrictions over and above the Hertz restriction would be needed for expression (324) to transform as a scalar.

To this end he once again used his two step procedure.[169] He wrote expression (324) in primed coordinates, and expressed its various components in unprimed coordinates:

$$S = \sum \pi_{i\alpha_1} \pi_{\kappa\alpha_2} T^x_{\alpha_1\alpha_2} p_{\mu\beta_1} p_{\nu\beta_2} \gamma_{\beta_1\beta_2} \pi_{\mu\sigma} \frac{\partial}{\partial x_\sigma}\left[\pi_{\nu\tau} \frac{\partial}{\partial x_\tau}(p_{i\varepsilon_1} p_{\kappa\varepsilon_2} \gamma_{\varepsilon_1\varepsilon_2})\right] . \tag{326}$$

Einstein connected the factors $p_{\mu\beta_1}$ and $\pi_{\mu\sigma}$ by a solid V-shaped line to indicate that they combine to form (in modern notation) $\delta_{\beta_1\sigma}$, Underneath this line, he wrote $\beta_1 = \sigma$. He likewise connected $p_{\nu\beta_2}$ and $\partial / \partial x_\sigma$ by a dashed V-shaped line and directly underneath wrote

---

168 It is not clear why Einstein added an ‘x’ added to $T_{i\kappa}$ at this point. This superscript was Einstein standard notation for first-order deviations from constant values. The argument on p. 11R works with both $T_{i\kappa}$ and $T^x_{i\kappa}$.

169 See the introduction to sec. 4.3 for a general discussion of this procedure and equations (215)–(217) for a concrete example involving a scalar.

$$\sum p_{\nu\beta_2}\frac{\partial\pi_{\nu\tau}}{\partial x_\sigma}\gamma_{\beta_1\beta_2} = -\sum\pi_{\nu\tau}\frac{\partial p_{\nu\beta_2}}{\partial x_\sigma}\underset{\sigma}{\gamma_{\beta_1\beta_2}} = 0. \tag{327}$$

In the first step he used that

$$0 = \frac{\partial}{\partial x_\sigma}(p_{\nu\beta_2}\pi_{\nu\tau}) = p_{\nu\beta_2}\frac{\partial\pi_{\nu\tau}}{\partial x_\sigma} + \pi_{\nu\tau}\frac{\partial p_{\nu\beta_2}}{\partial x_\sigma}. \tag{328}$$

He replaced β_1 in $\gamma_{\beta_1\beta_2}$ by σ, using that $p_{\mu\beta_1}\pi_{\mu\sigma} = \delta_{\beta_1\sigma}$. Expression (327) has exactly the form of the last term in equation (236) at the top of p. 10L. At that point, Einstein had drawn the conclusion that the vanishing of this expression (see equation (237)) was the condition for non-autonomous transformation under which the Hertz expression transforms as a tensor. Equation (237) does indeed express the Hertz restriction for infinitesimal transformations, but not for finite ones (see the discussion following equation (237) and footnote 148). In the second step in equation (327), Einstein nonetheless used equation (237).

On the basis of equation (327), Einstein could move $\pi_{\nu\tau}$ outside the scope of the differential operator $\partial/\partial x_\sigma$ in equation (326). As he wrote: "hence $\pi_{\nu\tau}$ can be taken outside" ("also $\pi_{\nu\tau}$ heraus setzbar"). The factor $\pi_{\nu\tau}$ combines with $p_{\nu\beta_2}$ to give $\delta_{\tau\beta_2}$. In a separate box Einstein summarized the simplifications $p_{\mu\beta_1}\pi_{\mu\sigma} = \delta_{\beta_1\sigma}$ and $\pi_{\nu\tau}p_{\nu\beta_2} = \delta_{\tau\beta_2}$ in equation (326):

$$\begin{aligned}\beta_1 &= \sigma\\ \beta_2 &= \tau.\end{aligned} \tag{329}$$

With these simplifications, equation (326) becomes:

$$S' = \sum\pi_{i\alpha_1}\pi_{\kappa\alpha_2}T^x_{\alpha_1\alpha_2}\gamma_{\sigma\tau}\frac{\partial}{\partial x_\sigma}\frac{\partial}{\partial x_\tau}(p_{i\varepsilon_1}p_{\kappa\varepsilon_2}\gamma_{\varepsilon_1\varepsilon_2}). \tag{330}$$

Underneath this equation Einstein wrote: "Let us restrict ourselves to an infinitesimal substitution" ("Beschränken wir uns auf infinitesimale Substitution"). Whether Einstein realized it or not, this immediately takes care of the problem that to arrive at equation (330) he had used the Hertz restriction in a form that only holds for infinitesimal transformations.

Einstein's task now was to identify all those terms in equation (330) that would have to vanish for the right-hand side to reduce to expression (324) that he had started from. These are all terms in which the elements $p_{\mu\nu}$ of the transformation matrix are differentiated at least once.[170] The condition determining under which subclass of infinitesimal Hertz transformations expression (324) transforms as a scalar is obtained by setting the sum of all these terms equal to zero. If in a product of several matrix elements $p_{\mu\nu}$ of an infinitesimal transformation one is differentiated, all others can be replaced by Kronecker deltas. Equation (330) can thus be rewritten as:

170 Einstein had made this observation twice on p. 7R in comments on equations (139) and (145).

$$\begin{aligned} S' &= \delta_{i\alpha_1}\delta_{\kappa\alpha_2}T^x_{\alpha_1\alpha_2}\gamma_{\sigma\tau}\frac{\partial}{\partial x_\sigma}\frac{\partial}{\partial x_\tau}(p_{i\varepsilon_1}\delta_{\kappa\varepsilon_2}\gamma_{\varepsilon_1\varepsilon_2} + \delta_{i\varepsilon_1}p_{\kappa\varepsilon_2}\gamma_{\varepsilon_1\varepsilon_2}) \\ &= T^x_{i\kappa}\gamma_{\sigma\tau}\frac{\partial}{\partial x_\sigma}\frac{\partial}{\partial x_\tau}(p_{i\varepsilon_1}\gamma_{\varepsilon_1\kappa} + p_{\kappa\varepsilon_2}\gamma_{i\varepsilon_2}) \end{aligned} \tag{331}$$

As Einstein put it: "If one of the p's is differentiated at all, for instance $p_{i\varepsilon_1}$, then one has to set $\kappa = \varepsilon_2$, $i = \alpha_1$, $\kappa = \alpha_2$" ("Dann muss, falls überhaupt eines der p diff wird, z. B. $p_{i\varepsilon_1}$ $\kappa = \varepsilon_2$, $i = \alpha_1$, $\kappa = \alpha_2$ gesetzt werden"). The second step in equation (331) is indeed to set $\alpha_1 = i$, $\alpha_2 = \kappa$ and to set $\varepsilon_2 = \kappa$ if $p_{i\varepsilon_1}$ is differentiated and $\varepsilon_1 = i$ if $p_{\kappa\varepsilon_2}$ is. Einstein could thus rewrite equation (330) as:[171]

$$S' = \sum T^x_{i\kappa}\frac{\partial}{\partial x_\sigma}\frac{\partial}{\partial x_\tau}\gamma_{\sigma\tau}(p_{i\varepsilon_1}\gamma_{\varepsilon_1\kappa} + p_{\kappa\varepsilon_2}\gamma_{\varepsilon_2 i}) + S, \tag{332}$$

adding: "where p is to be differentiated at least once" ("wobei p mindestens einmal zu differenzieren ist"). The last term, S, is obtained if the two differential operators both act on $\gamma_{\varepsilon_1\kappa}$ or $\gamma_{\varepsilon_2 i}$ instead of at least one of them acting on $p_{i\varepsilon_1}$ or $p_{\kappa\varepsilon_2}$. For S to be a scalar, the sum of all terms containing derivatives of $p_{\mu\nu}$ in the first term on the right-hand side of equation (332) should vanish.

Einstein first considered the expression[172]

$$T_{i\kappa}\frac{\partial}{\partial x_\sigma}\left(\gamma_{\sigma\tau}\frac{\partial p_{i\varepsilon_1}}{\partial x_\tau}\gamma_{\varepsilon_1\kappa}\right). \tag{333}$$

Since $\partial p_{i\varepsilon_1}/\partial x_\tau = \partial p_{i\tau}/\partial x_{\varepsilon_1}$ (see footnote 148) this can be rewritten as:

$$T_{i\kappa}\frac{\partial}{\partial x_\sigma}\left(\gamma_{\sigma\tau}\frac{\partial p_{i\tau}}{\partial x_{\varepsilon_1}}\gamma_{\varepsilon_1\kappa}\right). \tag{334}$$

Because of the Hertz restriction, $\partial\gamma_{\mu\nu}/\partial x_\nu = 0$, $\gamma_{\sigma\tau}$ can be taken outside the scope of $\partial/\partial x_\sigma$ and $\gamma_{\varepsilon_1\kappa}$ can be taken inside the scope of $\partial/\partial x_{\varepsilon_1}$. Einstein thus arrived at:

$$\sum T_{i\kappa}\gamma_{\sigma\tau}\frac{\partial}{\partial x_\sigma}\frac{\partial}{\partial x_{\varepsilon_1}}(p_{i\tau}\gamma_{\varepsilon_1\kappa}). \tag{335}$$

He repeated this result on the next line,

$$T_{i\kappa}\frac{\partial}{\partial x_\sigma}\gamma_{\sigma\tau}\frac{\partial^2}{\partial x_{\varepsilon_1}\partial\sigma}(p_{i\tau}\gamma_{\varepsilon_1\kappa}, \tag{336}$$

171 The second term in parentheses in equation (332) is written underneath the first. Both S and $\gamma_{\sigma\tau}$ in equation (332) are interlineated. Note that $\gamma_{\sigma\tau}$ is placed within the scope of the differential operators $\partial/\partial x_\sigma$ and $\partial/\partial x_\tau$, whereas in equations (330) and (331) it was not. Because of the Hertz restriction, however, $\partial\gamma_{\sigma\tau}/\partial x_\sigma = \partial\gamma_{\sigma\tau}/\partial x_\tau = 0$, so this makes no difference.

172 Three lines farther down, Einstein added: "Already sum over ε_1" ("Schon Summe über ε_1"), drawing a line from expression (333) to this comment. The comment refers, perhaps, to the similarity between the sum over ε_1 in $(\partial p_{i\varepsilon_1}/\partial x_\tau)\gamma_{\varepsilon_1\kappa}$ in expression (333) and the sum over α in equation (239), one of the conditions for infinitesimal Hertz transformations.

but did not pursue the calculation any further. He went back to equation (332) and, contrary to what he did in expression (333), treated the two derivative operators $\partial/\partial x_\sigma$ and $\partial/\partial x_\tau$ on an equal footing. For the term with $p_{i\varepsilon_1}\gamma_{\varepsilon_1\kappa}$ in equation (332) he wrote:

$$\sum T_{i\kappa}\gamma_{\sigma\tau}\left(\frac{\partial p_{i\varepsilon_1}}{\partial x_\sigma}\frac{\partial\gamma_{\varepsilon_1\kappa}}{\partial x_\tau}+\frac{\partial p_{i\varepsilon_1}}{\partial x_\tau}\frac{\partial\gamma_{\varepsilon_1\kappa}}{\partial x_\sigma}+\frac{\partial^2 p_{i\varepsilon_1}}{\partial x_\sigma\partial x_\tau}\gamma_{\varepsilon_1\kappa}\right). \tag{337}$$

He indicated that the term with $p_{\kappa\varepsilon_2}\gamma_{\varepsilon_2 i}$ gives a similar contribution by noting: "+ the same with i & κ exchanged" ("+ dasselbe mit vert[auschten] i & κ").

Here the calculation seems to break off abruptly with Einstein concluding: "Leads to difficulties" ("Führt auf Schwierigkeiten."). However, the considerations at the top of p. 12L (and some additions to p. 11L resulting from them) can be seen as a natural continuation of the search on p. 11R for Hertz transformations under which the core operator transforms as a tensor. Einstein may therefore only have added this final remark on p. 11R after running into difficulties on pp. 12L and 11L.

4.5.5 *Checking Whether Rotation in Minkowski Spacetime Is a Hertz Transformation Under Which the Core Operator Transforms as a Tensor (12L, 11L)*

At the bottom of p. 11R, Einstein had derived a condition determining under which subclass of infinitesimal Hertz transformations expression (323)—the contraction of the core operator and an arbitrary second-rank covariant tensor—transforms as a scalar. He had found that, given the metric field, the matrices $p_{\mu\nu}$ for such transformations must satisfy the condition that the sum of expression (337) and a similar expression obtained by switching the indices i and κ vanish. For the special case of a flat diagonal metric, $\gamma_{\mu\nu} = \delta_{\mu\nu}$, this condition reduces to:

$$\gamma_{\sigma\tau}\left(\frac{\partial^2 p_{i\kappa}}{\partial x_\sigma\partial x_\tau}+\frac{\partial^2 p_{\kappa i}}{\partial x_\sigma\partial x_\tau}\right) = 0, \tag{338}$$

which is satisfied if $p^x_{\mu\nu} \equiv (p_{\mu\nu}-\delta_{\mu\nu})$ is anti-symmetric, i.e., $p^x_{\mu\nu} = -p^x_{\mu\nu}$.[173]

12L At the top of p. 12L, under the heading: "Attempt. Infinitesimal transformation is anti-symmetric. Rotation modified" ("Versuch. Infinitesimale Transformation ist schief symmetrisch. Drehung modifiziert"), Einstein turns to the investigation of anti-symmetric infinitesimal transformations. This quickly aborted attempt can thus be seen as a natural continuation of the considerations on p. 11R.

Einstein wrote down the transformation law for the differentials dx_μ under an infinitesimal coordinate transformation:

$$dx'_\nu = dx + p^x_{\nu\kappa}dx_k. \tag{339}$$

173 In that case, conditions (239) and (241) for infinitesimal Hertz transformations found on p. 10L are automatically satisfied as well, the latter because $p^x_{\kappa\sigma}$ is traceless, the former because (cf. footnote 148): $\partial p^x_{\kappa\sigma}/\partial x_\sigma = -\partial p^x_{\sigma\kappa}/\partial x_\sigma = -\partial p^x_{\sigma\sigma}/\partial x_\kappa$, which once again vanishes because $p^x_{\kappa\sigma}$ is traceless.

and noted the condition of anti-symmetry

$$p^x_{\nu\kappa} = -p^x_{\kappa\nu}\,. \tag{340}$$

As on p. 10L, he expressed the coefficients of the transformation in terms of the functions X_μ describing the coordinate transformation (cf. equations (242)–(244)),

$$p_{\nu\kappa} = \frac{\partial X_\nu}{\partial x_\kappa}\,, \tag{341}$$

with the help of which he rewrote condition (340) as

$$\frac{\partial X_\nu}{\partial x_\kappa} = -\frac{\partial X_\kappa}{\partial x_\nu}\,. \tag{342}$$

The comment "Rotation modified" ("Drehung modifiziert") at the top of p. 12L indicates that Einstein was interested in the special case of rotation in Minkowski spacetime at this point. If the matrix $p^x_{\mu\nu}$ for this transformation were anti-symmetric—which, of course, it is not—an infinitesimal rotation in Minkowski spacetime would be an example of a non-autonomous transformation under which expression (323) transforms as a scalar. It would then, presumably, also be a transformation under which the core operator transforms as a tensor. Einstein thus explored whether the matrix $p^x_{\mu\nu}$ for rotation can meaningfully be made anti-symmetric.

If this was indeed the point of modifying the matrix $p^x_{\mu\nu}$ for rotation, Einstein had already achieved his goal without such modification. What he appears to have overlooked is that infinitesimal rotations in Minkowski spacetime already are infinitesimal Hertz transformations under which expression (323) transforms as a scalar. That they are Hertz transformations was shown in p. 11L (see sec. 4.5.2). Moreover, the matrix $p^x_{\mu\nu}$ for such transformations satisfies condition (338). After all, $p^x_{\mu\nu}$ is linear in x_μ and condition (338) involves only second-order derivatives of $p^x_{\mu\nu}$. It is true that condition (340) is not satisfied, since $p^x_{\mu\nu} \neq p^x_{\nu\mu}$, but that condition, although sufficient, is not necessary to meet condition (338). Hence, Einstein did not need to modify $p^x_{\mu\nu}$ for rotation at all.

Einstein seems to have missed this and returned to p. 11L to see whether the matrix 11L
(304) for $p^x_{\mu\nu}$ for infinitesimal rotation in Minkowski spacetime could be made anti-symmetric. As he indicated underneath the matrix on p. 11L, he replaced $p^x_{31} = p^x_{32} = 0$ by $-\omega y$ and ωx, respectively:

$$\begin{matrix} 1 & \omega t & +y\omega \\ -\omega t & 1 & -x\omega \\ 0 & 0 & 1 \\ -\omega y & +\omega x\,. & \end{matrix} \tag{343}$$

With this anti-symmetrized matrix the differential of the time coordinate transforms as:

$$dt' = -y\omega dx + x\omega dy + dt\,. \tag{344}$$

Einstein wrote this equation in a separate box in the left margin of p. 11L. For this equation to be a coordinate transformation, dt' has to be an exact differential, i.e., it must be possible to write it as:

$$dt' = \frac{\partial\varphi}{\partial x}dx + \frac{\partial\varphi}{\partial y}dy + \frac{\partial\varphi}{\partial t}dt\,. \tag{345}$$

Comparison of equations (345) and (344) gives $\partial\varphi/\partial x = -y\omega$ and $\partial\varphi/\partial y = x\omega$. This implies that

$$\frac{\partial^2\varphi}{\partial x\partial y} \neq \frac{\partial^2\varphi}{\partial y\partial x}\,. \tag{346}$$

Hence dt' in equation (345) is not an exact differential. Einstein seems to have gone through this same argument himself, although the only trace of this in the notebook are the terms

$$\frac{\partial\varphi}{\partial x} \quad \frac{\partial\varphi}{\partial y} \quad \frac{\partial\varphi}{\partial z}\,, \tag{347}$$

written underneath equation (344) in the same separate box. In any event, he concluded that a transformation characterized by equation (344), which would yield an antisymmetric matrix $p^x_{\mu\nu}$, is "impossible" ("unmöglich"). This is the last word in the separate box on p. 11L, and it signals the end of this whole line of reasoning, which started on p. 10L and ended with the first horizontal line on p. 12L. Einstein seems to have reached the conclusion that the core operator does not transform as a tensor under infinitesimal rotations in Minkowski spacetime.

Initially, Einstein, it seems, considered changing the form of the core operator. Following the first horizontal line on p. 12L, Einstein changed the core operator (130) on p. 7L to:

$$\frac{1}{\sqrt{G}}\sum\frac{\partial}{\partial x_\mu}\left\{\sqrt{G}\gamma_{\mu\nu}\frac{\partial\gamma_{i\kappa}}{\partial x_\nu}\right\}. \tag{348}$$

The extra factors of $\sqrt{G}$ make this expression resemble the second Beltrami invariant more closely (cf. equation (82) on p. 6L). Einstein did not even begin the search for non-autonomous transformations under which this modified core operator transforms as a tensor. Instead he drew another horizontal line and turned to a closer examination of the important special case of rotation in Minkowski spacetime that had spelled trouble for the original form of the core operator.

4.5.6 *Deriving the Exact Form of the Rotation Metric (12L–R)*

12L In the middle of p. 12L, under the heading, "The Rotational Field in First Approximation" ("Drehungsfeld in erster Annäherung"), Einstein wrote down the line element

$$g_{11}dx^2 + \ldots + g_{44}dt^2 = ds^2\,, \tag{349}$$

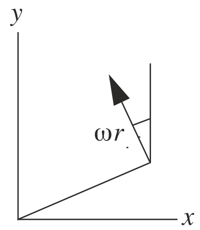
and defined H, the "Lagrangian function" ("Lagrange'sche Funktion") for a point particle in a metric field, in terms of a potential term Φ and a kinetic term L:

$$\Phi - L = H = \frac{ds}{dt}. \tag{350}$$

The accompanying diagram (see below) and the subsequent calculations make it clear that Einstein considered the motion of the particle in a coordinate system $x_\mu = (x, y, z, t)$ rotating counterclockwise at constant angular velocity ω around the z-axis, coinciding with the z'-axis of an inertial coordinate system $x'_\mu = (x', y', z', t')$. The line element in the inertial coordinate system is given by:

$$ds^2 = \eta_{\mu\nu}dx'_\mu dx'_\nu = (1 - \mathbf{v}'^2)dt'^2 \tag{351}$$

where $\eta_{\mu\nu} \equiv \mathrm{diag}(-1, -1, -1, 1)$ (coordinates are chosen such that $c = 1$) and $v'_i = \dot{x}'_i = dx'_i/dt'$.



The relation between velocity in the inertial frame and velocity in the rotating frame is given by:

$$\mathbf{v}' = \mathbf{v} + \vec{\omega} \times \mathbf{x} = \begin{pmatrix} \dot{x} - \omega y \\ \dot{y} + \omega x \\ \dot{z} \end{pmatrix} = \begin{pmatrix} \dot{x} - \omega r \sin\omega t \\ \dot{y} + \omega r \cos\omega t \\ \dot{z} \end{pmatrix}, \tag{352}$$

where $\vec{\omega} \equiv (0, 0, \omega)$ and $r^2 \equiv x^2 + y^2$.

With the help of equation (351) and the transformation equation $dt' = dt$, the Lagrangian (350) can be written as

$$H = \sqrt{\frac{ds^2}{dt^2}} = \sqrt{1 - \mathbf{v}'^2} \approx 1 - \frac{\mathbf{v}'^2}{2}. \tag{353}$$

Comparison with $H = \Phi - L$ leads to the identification $2L = \mathbf{v}'^2$. Einstein presumably arrived at this equation simply on the basis of the interpretation of L as the kinetic energy. Using equation (352), he found:

$$2L = (\dot{x} - \omega r \sin\omega t)^2 + (\dot{y} + \omega r \cos\omega t)^2 + \dot{z}^2, \tag{354}$$

which he expanded to:

$$2L = \dot{x}^2 + \dot{y}^2 + \dot{z}^2 - 2\omega r \sin\omega t \dot{x} + 2\omega r \cos\omega t \dot{y} + \omega^2 r^2. \tag{355}$$

Substituting $x = r\cos\omega t$ and $y = r\sin\omega t$ (see equation (352)) and introducing the potential energy $\Phi = A$,[174] Einstein wrote the Lagrangian H as

$$\Phi - L = A - \frac{\omega^2 r^2}{2} - \frac{\dot{x}^2 + \dot{y}^2 + \dot{z}^2}{2} + \omega y\dot{x} - \omega x\dot{y}\,. \tag{356}$$

Given equation (350), Einstein could find ds^2/dt^2 by squaring equation (356). Writing "ds^2/dt [dt should be dt^2] calculated up to and including ω^2 & $\dot{x}^2$" ("ds^2/dt berechnet bis und mit ω^2 u. $\dot{x}^2$"), he arrived at:

$$\begin{aligned}\frac{ds^2}{dt^2} &= A^2 + \omega^2 y^2\dot{x}^2 + \omega^2 x^2\dot{y}^2 - A\omega^2 r^2 \\ &\quad - \left(A - \frac{\omega^2 r^2}{2}\right)(\dot{x}^2 + \dot{y}^2 + \dot{z}^2) + 2A\omega y\dot{x} - 2A\omega x\dot{y}\end{aligned} \tag{357}$$

A simpler way of finding an expression for ds^2/dt^2 is to use equations (351) and (352):

$$\frac{ds^2}{dt^2} = 1 - \mathbf{v}'^2 = 1 - \omega^2 r^2 - \dot{x}^2 - \dot{y}^2 - \dot{z}^2 + 2\omega y\dot{x} - 2\omega x\dot{y}\,. \tag{358}$$

If we neglect terms containing *both* ω^2 and $\dot{x}_i^2$ in eq. (357) and insert $A = 1$, we recover equation (358).

12R At the top of p. 12R, Einstein used equation (357) in combination with $ds^2 = g_{\mu\nu}dx_\mu dx_\nu$, the expression for the line element in rotating coordinates, to identify the components of $g_{\mu\nu}$, the Minkowski metric in rotating coordinates:

$$\begin{array}{llll} g_{11} = -\left(A - \cancel{\frac{\omega^2 r^2}{2}}\right) + \omega^2 y^2 & g_{12} = 0 & g_{13} = 0 & g_{14} = 2A\omega y \\ g_{21} = 0 & g_{22} = -\left(A - \cancel{\frac{\omega^2 r^2}{2}}\right) + \omega^2 x^2 & g_{23} = 0 & g_{24} = -2A\omega x \\ g_{31} = 0 & g_{32} = 0 & g_{33} = -\left(A - \cancel{\frac{\omega^2 r^2}{2}}\right) & g_{34} = 0 \\ g_{41} = 2A\omega y & g_{42} = -2A\omega x & g_{43} = 0 & g_{44} = A^2 - A^2\omega^2 r^2 \end{array} \tag{359}$$

This matrix contains several errors. First, the ω^2-terms in g_{11} and g_{22} come from terms in equation (357) containing both ω^2 and $\dot{x}_i^2$, which are negligible. This mistake was partly corrected.[175] Second, the factors of 2 in g_{14}, g_{24}, g_{41}, and g_{42} should be 1. The term $2A\omega y dx dt$ in (dt^2 times) equation (357), for instance, should be set equal to $g_{14}dx_1 dx_4 + g_{41}dx_4 dx_1$, leading to the identification $g_{14} = g_{41} = A\omega y$.[176] Immediately below the matrix (359), Einstein noted that $A = 1$ (cf. note 174)

174 As follows directly from equation (353) and as Einstein subsequently realized (see p. 12R), A cannot be chosen freely but has to be equal to 1.

175 The term $\omega^2 r^2/2$ in the expressions for g_{11}, g_{22}, and g_{33} may have been deleted in the course of Einstein's evaluation of the determinant of this metric.

Next, Einstein computed the determinant G of (359), again retaining only terms up to order ω^2 or $\dot{x}_i^2$. Since so many of the elements of the matrix (359) are zero, there are only a few contributions to its determinant:

$$G = g_{11}g_{22}g_{33}g_{44} - g_{11}g_{33}(g_{42}^2 + g_{41}^2)\,. \tag{360}$$

Inserting the values given in the matrix (359) into this expression, setting $A = 1$, and neglecting terms smaller than of order ω^2 or $\dot{x}_i^2$, we find a result very similar to the following expression in the notebook at this point:[177]

$$G = \left[\left(-1 + \frac{\omega^2 r^2}{2} + \omega^2 y^2\right)\left\{\left(1 - \frac{\omega^2 r^2}{2} - \omega^2 x^2 - \frac{\omega^2 r^2}{2} - \omega^2 r^2\right) + 4\omega^2 x^2\right\}\right. \\ \left. + 4\omega^2 y^2.\right. \tag{361}$$

Einstein rewrote the right-hand side of this equation as:

$$= -1 + \frac{\omega^2 r^2}{2} + \omega^2 y^2 + \frac{\omega^2 r^2}{2} + \omega^2 x^2 + \frac{\omega^2 r^2}{2} + \omega^2 r^2 - 4\omega^2 x^2 + 4\omega^2 y^2 \tag{362}$$

and then as:

$$= -1 + 2.5\omega^2 r^2 - \langle \omega^2 r^2 \rangle - 3\omega^2 x^2 + 5\omega^2 y^2\,. \tag{363}$$

These last two equations inherit the errors made in equation (361) (see note 177). Einstein may have realized that these equations contained some errors. He subsequently deleted all three equations (361)–(363). However, he retained the main result of his calculation on the lower half of p. 12L and the upper half of p. 12R, expression (359) for the Minkowski metric in rotating coordinates, which still contains several errors.

4.5.7 *Trying to Find Infinitesimal Unimodular Transformations Corresponding to Uniform Acceleration (12R, 41L–R)*

In the middle of p. 12R, under the heading, "Substitutions with Determinant 1. Infin- 12R
itesimal in 2 Variables" ("Substitutionen mit Determinante 1. Infinitesimal in 2 Variablen"), Einstein wrote down the transformation law for coordinate differentials,

$$\begin{aligned} dx' &= dx + (p_{11}^x dx + p_{12}^x dy), \\ dy' &= dy + (p_{21}^x dx + p_{22}^x dy), \end{aligned} \tag{364}$$

under the transformation

176 Einstein made the same mistake on p. 42R (see footnote 308) and in the Einstein-Besso manuscript (CPAE 4, Doc. 14, pp. [41–42]). Largely due to this error, Einstein convinced himself at that point that the rotation metric is a solution of the field equations of his *Entwurf* theory (see Janssen 1999, 145–146, and "What Did Einstein Know …" sec. 3 (in this volume).

177 Expression (361) contains a number of errors. First, the last three minus signs in the second expression in ordinary brackets were all corrected from plus signs. The first two should indeed be minus signs but the third should be a plus sign. Secondly, the terms $4\omega^2 x^2$ and $4\omega^2 y^2$, coming from the second term in equation (360), should both be inside the curly brackets with a minus sign.

$$\begin{aligned} x' &= x + X(x, y), \\ y' &= y + Y(x, y), \end{aligned} \tag{365}$$

with (cf., e.g., p. 10L and equations (242)–(244)):

$$p^x_{ij} = \begin{pmatrix} \frac{\partial X}{\partial x} & \frac{\partial X}{\partial y} \\ \frac{\partial Y}{\partial x} & \frac{\partial Y}{\partial y} \end{pmatrix} . \tag{366}$$

Next to the transformation law (364), he wrote the unimodularity condition, i.e., the condition that the determinant of the transformation matrix equals 1 (cf. equations (240)–(241)):

$$p^x_{11} + p^x_{22} = 0 . \tag{367}$$

With the help of equation (366), this condition turns into

$$\frac{\partial X}{\partial x} + \frac{\partial Y}{\partial y} = 0 , \tag{368}$$

which is automatically satisfied if there is a generating function $\psi(x, y)$ determining X and Y via

$$X = \frac{\partial \psi}{\partial y}, \qquad Y = -\frac{\partial \psi}{\partial x} . \tag{369}$$

With the help of equation (366), the coefficients p^x_{ij} can also be expressed in terms of ψ:[178]

$$\begin{aligned} p_{11} &= \frac{\partial^2 \psi}{\partial x \partial y}, \qquad & p_{12} &= \frac{\partial^2 \psi}{\partial y^2}, \\ p_{21} &= -\frac{\partial^2 \psi}{\partial x^2}, \qquad & p_{22} &= -\frac{\partial^2 \psi}{\partial x \partial y}. \end{aligned} \tag{370}$$

Using these expressions, Einstein rewrote the first line of the transformation (364) as

$$dx' = \left(1 + \frac{\partial^2 \psi}{\partial x \partial y}\right) dx + \frac{\partial^2 \psi}{\partial y^2} dy . \tag{371}$$

He did not bother to write down the corresponding equation for dy'. Equation (371) in turn can be rewritten as

$$dx' = dx + \frac{\partial}{\partial x}\left(\frac{\partial \psi}{\partial y}\right) dx + \frac{\partial}{\partial y}\left(\frac{\partial \psi}{\partial y}\right) dy = dx + d\left(\frac{\partial \psi}{\partial y}\right) = dx + dX . \tag{372}$$

178 In these equations, p_{ij} should be p^x_{ij}.

This last expression already follows directly, of course, from equation (365). Writing δx and δy for X and Y and using equation (369), Einstein wrote

$$\delta x = \frac{\partial\psi}{\partial y}, \qquad \delta y = -\frac{\partial\psi}{\partial x}. \tag{373}$$

He could thus write transformation (365) as

$$\begin{aligned} x' &= x + \frac{\partial\psi}{\partial y}, \\ y' &= y - \frac{\partial\psi}{\partial x}, \end{aligned} \tag{374}$$

which appear at the bottom of p. 12R enclosed in a box.

In another part of the notebook, on pp. 41L–R, immediately following Einstein's earliest considerations on gravitation (see sec. 2), Einstein did the same calculation as on the bottom half p. 12R but this time pursued it a little further. At the bottom of p. 41L, he examined some specific choices for the generating function ψ. At the top of p. 41R, he chose one of the coordinates to be the time coordinate and compared the transformation for a particular choice of ψ with the transformation to a uniformly accelerated frame of reference, a transformation familiar from his papers on the static gravitational field. It seems plausible that the calculation on pp. 41L–R is just a continuation of the one on p. 12R. The calculation breaks off after Einstein failed to recover the transformation to a uniformly accelerating frame of reference in this manner. 41L–R

Below the horizontal line in the middle of p. 41L, under the heading, "Simplest Substitutions, whose Determinant = 1" ("Einfachste Substitution, deren Determinante = 1"), Einstein, as on p. 12R, began by writing down the transformation equations (364) for coordinate differentials, albeit in a more compact form than on p. 12R and leaving open the dimension of the space(-time) under consideration: 41L

$$dx'_\nu = dx_\nu + \sum p^x_{\nu\sigma} dx_\sigma . \tag{375}$$

As on p. 12R (see equation (366)), Einstein used the relations $p^x_{\mu\nu} = \partial X_\mu / \partial x_\nu$. Underneath equation (375), Einstein wrote: "X_ν are homogeneous and of second degree in the coordinates. Only two coordinates are being transformed" ("X_ν sind homogen u. zweiten Grades in den Koordinaten. Es werden nur zwei Koordinaten transformiert"). These comments suggest that Einstein was interested at this point in the special case of uniform acceleration, which he explicitly considered at the top of p. 41R. In that case only two coordinates, x and t, transform non-trivially. Moreover, the function X in $x' = x + X$ (cf. equation (365)) has to be proportional to t^2 and cannot have a constant term to get the desired form $x' = x + f(x)t^2$. In other words, X has to be "of second degree" and "homogeneous."

Einstein began by writing down the condition of unimodularity (cf. equations (367)–(368))

$$\frac{\partial X}{\partial x} + \frac{\partial Y}{\partial y} = 0\,, \tag{376}$$

which is satisfied automatically if there is a generating function $\psi(x, y)$ such that (cf. equation (369))

$$X = \frac{\partial \psi}{\partial y}, \qquad Y = -\frac{\partial \psi}{\partial x}. \tag{377}$$

As before (see equation (370)), the relation between the matrix p^x_{ij} and the function ψ is given by

$$\begin{aligned} p^x_{11} &= \frac{\partial^2 \psi}{\partial x \partial y}, & p^x_{12} &= \frac{\partial^2 \psi}{\partial y^2}, \\ p^x_{21} &= -\frac{\partial^2 \psi}{\partial x^2}, & p^x_{22} &= -\frac{\partial^2 \psi}{\partial x \partial y}. \end{aligned} \tag{378}$$

Up to this point the argument on p. 41L is identical to the argument on p. 12R.

Einstein now considered two specific choices for the generating function, namely $\psi = r^3$ and $\psi = r^2 x$. For these two cases he evaluated the four elements of the matrix in equation (378). Using that $r \equiv \sqrt{x^2 + y^2}$, so that $\partial r / \partial x = x / r$ and $\partial r / \partial y = y / r$, one recovers the results given by Einstein at this point, except for an overall factor of 3 in the case of $\psi = r^3$. Einstein effectively did the calculation with $\psi = r^3 / 3$. This function gives:

$$\begin{aligned} \frac{\partial \psi}{\partial x} &= rx & \frac{\partial^2 \psi}{\partial x^2} &= r + \frac{x^2}{r} \\ & & \frac{\partial^2 \psi}{\partial x \partial y} &= \frac{xy}{r} \\ \frac{\partial \psi}{\partial y} &= ry & \frac{\partial^2 \psi}{\partial y^2} &= r + \frac{y^2}{r} \end{aligned} \tag{379}$$

The function $\psi = r^2 x$ gives:

$$\begin{aligned} \frac{\partial \psi}{\partial x} &= r^2 + 2x^2 & \frac{\partial^2 \psi}{\partial x^2} &= 6x \\ & & \frac{\partial^2 \psi}{\partial x \partial y} &= 2y \\ \frac{\partial \psi}{\partial y} &= 2xy & \frac{\partial^2 \psi}{\partial y^2} &= 2x \end{aligned} \tag{380}$$

To the right of equations (376)–(378), Einstein wrote down two matrices

$$\begin{pmatrix} 1 & y \\ \cdot & 1 \end{pmatrix}, \qquad \begin{pmatrix} y & x \\ \langle \varphi(x) \rangle 0 & -y \end{pmatrix}, \tag{381}$$

which appear to be related to the considerations at the top of p. 41R. Apart from the 21-component, the second matrix is proportional to the matrix p^x_{ij} for $\psi = r^2x$, which is found upon substitution of equation (380) into equation (378):

$$p^x_{ij} = 2\begin{pmatrix} y & x \\ -3x & -y \end{pmatrix}. \tag{382}$$

As we shall see below, Einstein probably changed the 21-component of the second of the two matrices (381) to zero in the course of his calculations on p. 41R.

At the top of p. 41R (the first four lines in the top left corner and the first two in the top right corner), Einstein examined the transformation generated by $\psi = y^2x$, a modification of the function $\psi = r^2x$ considered at the bottom of p. 41L. Replacing y by t, he then compared this transformation to the transformation to a uniformly accelerated frame that he had considered in the context of his 1912 theory for static gravitational fields. 41R

Inserting equation (378) into equation (375) for two dimensions and setting $\psi = r^2x$, one finds, using equation (380):

$$\begin{aligned} dx' &= dx + p^x_{11}dx + p^x_{12}dy \\ &= dx + \frac{\partial^2\psi}{\partial x\partial y}dx + \frac{\partial^2\psi}{\partial y^2}dy \\ &= dx + 2ydx + 2xdy, \\ dy' &= dy + p^x_{21}dx + p^x_{22}dy \\ &= dy - \frac{\partial^2\psi}{\partial x^2}dx - \frac{\partial^2\psi}{\partial x\partial y}dy \\ &= dy - 6xdx - 2ydy. \end{aligned} \tag{383}$$

This result corresponds to the matrix p^x_{ij} in equation (382) above.

Einstein wrote at the top of p. 41R:

$$\begin{aligned} dx' &= dx + \alpha(ydx + xdy), \\ dy' &= dy - \alpha ydy. \end{aligned} \tag{384}$$

This corresponds to the matrix

$$p^x_{ij} = \alpha\begin{pmatrix} y & x \\ 0 & -y \end{pmatrix}, \tag{385}$$

which, except for the factor α, is the second of the two matrices (381) on p. 41L. It seems that Einstein adjusted his choice of the function ψ to get the matrix p^x_{ij} in equation (385) instead of the one in equation (382). It is easily seen that to achieve this $\psi = r^2x$ should be replaced by $\psi = y^2x$. If that is done, one finds (cf. equation (379)–(380)):

$$\frac{\partial\psi}{\partial x} = y^2 \qquad \frac{\partial\psi}{\partial y} = 2yx \qquad \frac{\partial^2\psi}{\partial x^2} = 0 \qquad \frac{\partial^2\psi}{\partial x\partial y} = 2y \qquad \frac{\partial^2\psi}{\partial y^2} = 2x \tag{386}$$

Using these results in equation (383) and writing α instead of 2, one finds Einstein's equation (384).

Einstein now replaced y by t[179] and integrated equations (384), using that $tdx + xdt = d(xt)$:

$$\begin{aligned} x' &= x + \alpha xt, \\ t' &= t - \alpha\frac{t^2}{2}. \end{aligned} \tag{387}$$

He compared this transformation to the transformation to a uniformly accelerating frame

$$\begin{aligned} x' &= x + \frac{1}{2}c\frac{\partial c}{\partial x}t^2, \\ t' &= ct, \end{aligned} \tag{388}$$

which he had obtained in the context of his theory for static fields of 1912 (Einstein 1912b, 456) and which he had already used on p. 11L (see equation (311)). The expressions for t' in equations (387) and (388) are quite different. So even after changing the generating function from $\psi = r^2x$ (with $r^2 \equiv x^2 + t^2$) to $\psi = t^2x$, Einstein was unable to recover the transformation to a uniformly accelerating frame of reference by integrating the infinitesimal unimodular transformation (383) generated by ψ. On the remainder of p. 41R, Einstein went through a calculation showing that motion constrained to a curved surface in three-dimensional space is along a geodesic (see sec. 4.5.8). The purpose of this calculation may simply have to been to reassure himself after the disappointing results of p. 12R and pp. 41L-R that at least the conceptual basis of his theory was sound.

41R

4.5.8 *Geodesic Motion along a Surface (41R)*

On p. 41R, starting with the expression $m(d^2x/dt^2)$, Einstein considered the motion of a particle in three-dimensional space constrained to move on a two-dimensional surface, but otherwise free of external forces. He proved that the trajectory of such a particle is a geodesic of the surface by showing that the line element of the trajectory is an extremal on the surface. Einstein had earlier recognized that the equation of motion of a particle in a static gravitational field follows from a variational principle

179 The second line of equation (387) originally had y instead of t.

for a four-dimensional line element.[180] At this point, he presumably realized that a similar result holds for the equation of motion in a general gravitational field.[181] In his earlier lecture notes on mechanics, Einstein had also treated constrained motion along a surface, but the concept of a geodesic line is not to be found in any of his published writings up to this point.[182] It is not entirely clear why Einstein considered the problem of constrained motion in this context. Einstein had long been familiar with the link between the physical concept of constrained motion and the geometric concept of a geodesic line from a course on infinitesimal geometry that he had taken as a student at the ETH with Carl Friedrich Geiser.[183]

Einstein started from the x-component of Newton's equation of motion for a particle of mass m, constrained to move on the surface $f(x, y, z) = 0$.

$$m\frac{d^2x}{dt^2} = \lambda\frac{\partial f}{\partial x}. \tag{389}$$

The analogous equations for the y– and z–components are indicated by dashed lines. As is clear from the accompanying figure, reproduced below, the right-hand side of this equation is the normal force that constrains the particle to move along the surface $f = 0$. This normal force is proportional to the gradient of f, which defines the normal direction, and which must be multiplied by a Lagrange multiplier λ determined by the magnitude of the force.

Einstein next absorbed m into a new Lagrange multiplier λ', and then changed independent variables, substituting the arc length s for the time t.[184] He thus arrived at

$$\frac{d^2x}{ds^2} = \lambda''\frac{\partial f}{\partial x}, \tag{390}$$

where $\lambda'' = (dt^2/ds^2)\lambda'$.

The rest of the proof is intended to show that this is the equation of a geodesic line on the curved surface $f = 0$.[185] Einstein wrote down the equation for the surface,

$$f = 0, \tag{391}$$

---

180 See the "Nachtrag" to (Einstein 1912b).

181 See Einstein to Ludwig Hopf, 16 August 1912: "The work on gravitation is going splendidly. Unless everything is just an illusion, I have now found the most general equations" ("Mit der Gravitation geht es glänzend. Wenn nicht alles trügt, habe ich nun die allgemeinsten Gleichungen gefunden." CPAE 5, Doc. 416).

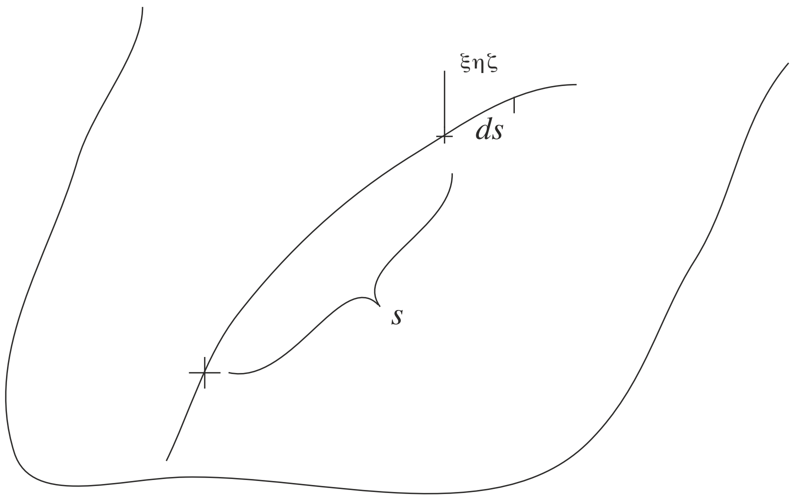
182 See CPAE 3, Doc. 1, [34–38], [75–76].

183 Einstein had registered for this course in winter semester 1897/1898 (see CPAE 1, Appendix E). Marcel Grossmann's notes on these lectures contain a page with very similar calculations (Bibliothek ETH, Zurich, Hs. 421:15).

184 There are no forces parallel to the surface, hence the speed of the point particle is constant.

185 Cf. Grossmann's notes on Geyser's lectures as well as very similar passages in Einstein's lecture notes on mechanics (CPAE 3, Doc. 4, [pp. 75ff.]).

drew a horizontal line, and sketched the figure below.



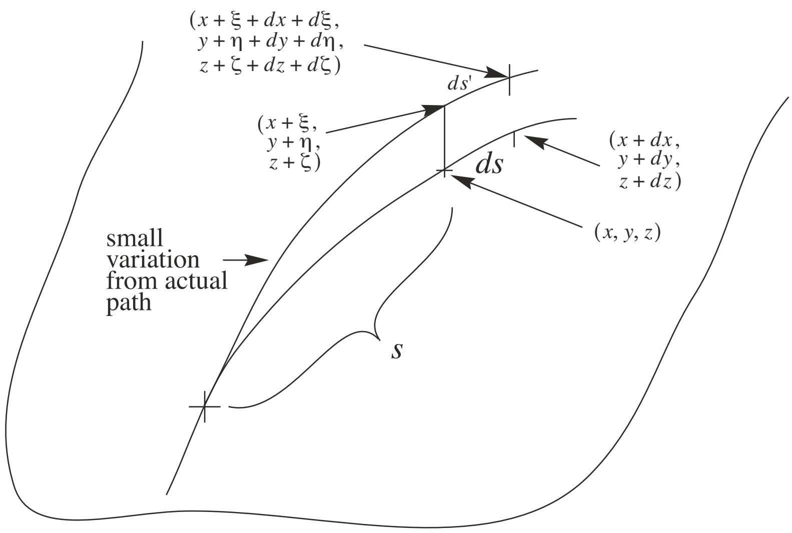
He then wrote down the coordinates:

$$x + \xi,$$
$$y + \eta,$$
$$z + \zeta. \tag{392}$$

The coordinates $(x(s), y(s), z(s))$ refer to points on the actual path. To prove that this path is a geodesic, he considered a nearby curve, produced by small variations (ξ, η, ζ). Thus a point with coordinates $(x + \xi, y + \eta, z + \zeta)$ is a point on this nearby curve. Next, Einstein considered a nearby point on the actual path, with coordinates $(x + dx, y + dy, z + dz)$, and a corresponding point on the path obtained through variation. He only wrote down the *x*-coordinate of the latter point:

$$x + \xi + \cancel{\frac{dx}{ds}}dx + d\xi\,. \tag{393}$$

Since Einstein was considering constrained motions, the path obtained through variation must also lie in the surface $f = 0$, i.e., $(x + \xi, y + \eta, z + \zeta)$ must also be the coordinates of a point on the surface $f = 0$. One can bring out the meaning of Einstein's figure more clearly by adding the path obtained through variation and labeling the various points.



Einstein now calculated the square of the line element ds' for the path obtained through variation, discarding terms of order $d\xi^2$:

$$\begin{aligned} ds'^2 &= (dx + d\xi)^2 + \cdot + \cdot \\ &= ds^2 + 2(dx d\xi + . + .) \\ &= ds^2\Big(1 + 2\Big(\frac{dx}{ds}\frac{d\dot{\xi}}{ds} + . + .\Big)\Big). \end{aligned} \tag{394}$$

He then took the square root of this equation:

$$ds' = ds\Big(1 + \Big(\frac{dx}{ds}\frac{d\dot{\xi}}{ds} + . + .\Big)\Big) \ . \tag{395}$$

Substituting $\dot{x}$ for dx/ds and $\dot{\xi}$ for $d\xi/ds$, he wrote the variation in the line element as:

$$ds' - ds = (\dot{x}\dot{\xi} + \cdot + \cdot)ds \, . \tag{396}$$

For a geodesic $\delta\int ds$ vanishes, which means that:

$$\int(\dot{x}\dot{\xi} + \cdot + \cdot)ds = 0 \, . \tag{397}$$

Through integration by parts Einstein transformed this equation into:

$$\int(\ddot{x}\xi + \cdot + \cdot)ds = 0 \, , \tag{398}$$

which will hold if the acceleration, with components ($\ddot{x}$, $\ddot{y}$, $\ddot{z}$), is perpendicular to the variation, with components (ξ , η , ζ). Since these variations are arbitrary (except that they have to lay within the surface), they should be perpendicular to the normal to the surface:

$$\frac{\partial f}{\partial x}\xi + \frac{\partial f}{\partial y}\eta + \frac{\partial f}{\partial z}\zeta = 0 \quad . \tag{399}$$

Hence, as Einstein wrote, “if” (“wenn”) condition (399) holds, then the actual path will be a geodesic: “from which the assertion” (“woraus die Behauptung”).

13L–R 4.6 *Emergence of the Entwurf Strategy (13L–R)*

On pp. 6L–12R (and pp. 41L–R), Einstein had tried in vain to find field equations invariant under a broad enough class of transformations—autonomous or non-autonomous[186]—to meet the requirements of the relativity principle and the equivalence principle (see sec. 1.1). He had pursued a combination of what we have called the mathematical and the physical strategy (see sec. 1.2).

Mathematically, the generally-covariant second Beltrami invariant (80), with (some power of) the determinant G of the metric playing the role of the arbitrary scalar function φ in its definition, looked like an especially promising point of departure. Field equations constructed out of the Beltrami invariant in this way are invariant under arbitrary (autonomous) unimodular transformations. On p. 12R and p. 41R, however, just prior to the entries on pp. 13L–R, Einstein had reached the conclusion that the important special case of an (autonomous) transformation to a uniformly accelerating frame of reference is not a unimodular transformation, not even infinitesimally. This must have been an important setback.

From a physics point of view, the core operator (118), which for weak fields reduces to the d’Alembertian acting on the metric, looked most promising. The drawback was that the core operator does not transform as a tensor under any autonomous non-linear transformations. It might, however, transform as a tensor under a class of non-autonomous non-linear transformation that would include the important special cases of rotation and uniform acceleration in Minkowski spacetime (see pp. 11L–12L). And even if this turned out not to be true for the core operator taken by itself, it might still be true for the sum of the core operator and some correction terms. Such correction terms were needed anyway to guarantee energy-momentum conservation (see p. 9L and sec. 4.4, especially the passage following equation (199)). With the help of such terms, it might furthermore be possible to connect field equations based on the core operator to the Beltrami invariants, which would throw light on their covariance properties. The Hertz restriction—the restriction to Hertz transformations, i.e., non-autonomous transformations under which the Hertz expression (231) transforms

186 See the introduction to sec. 4.3 for discussion of the distinction and footnote 94 for the origin of the terminology.

as a vector—played an important role in connecting the core operator to the Beltrami invariants. On p. 11L, Einstein had found that, at least infinitesimally, rotation in Minkowski spacetime was a Hertz transformation. He had also found, however, that uniform acceleration in Minkowski spacetime is not.

Still, the physically motivated core operator clearly held more promise overall than the mathematically more elegant Beltrami invariants. It is not surprising therefore that Einstein, on pp. 13L–R, bracketed the problem of the covariance of the field equations for the time being. He now began looking for field equations based on the core operator initially demanding only that such equations be invariant under arbitrary autonomous unimodular linear transformations. Presumably, he would check later whether these equations were also invariant under non-autonomous non-linear transformations such as uniform rotation and acceleration in Minkowski spacetime as was required by the equivalence principle. That Einstein restricted himself to unimodular transformations suggests that he eventually still wanted to connect the field equations to the Beltrami invariants.

On p. 13L, Einstein began an inventory of expressions involving the metric and its derivatives that transform as vectors and tensor under linear transformations and out of which he could therefore construct the correction terms to the core operator on the left-hand side of the field equations. On p. 13R, he substituted the core operator for the stress-energy tensor of matter in expression (74) for the energy-momentum balance between matter and field derived on p. 5R. In this way, it seems, Einstein hoped to identify the correction terms to the core operator that would guarantee the compatibility of the field equations with energy-momentum conservation. A variant of this strategy would subsequently lead to the *Entwurf* field equations. On p. 13R, however, Einstein quickly gave up on this line of reasoning. Marcel Grossmann then handed him a new mathematical quantity, which was far more promising than the Beltrami invariants. At the top of the very next page, p. 14L, the Riemann tensor makes its first appearance in the notebook. Pp. 14L–24L along with pp. 42L–43L are given over to attempts to extract field equations from this quantity along the lines of the mathematical strategy (see sec. 5). Only after these attempts had failed did Einstein return to the strategy we see emerging on pp. 13L–R (see p. 24R and pp. 26L–R and secs. 5.6.1 and 6).

4.6.1 *Bracketing the Generalization to Non-linear Transformations: Provisional Restriction to Linear Unimodular Transformations (13L)*

On p. 13L, under the heading, "Differential Covariants for Linear Substitutions, if one sets $G = 1$" ("Differentialkovarianten für lineare Substitutionen, falls $G = 1$ gesetzt wird"), Einstein started an inventory of quantities constructed out of the metric tensor and its derivatives that transform as vectors or tensors under (unimodular) linear transformations. All quantities that made it onto the list on p. 13L involve one and only one first-order derivative of the metric. Hence, they all fall under the heading "First Order" ("Erster Ordnung") on the third line of p. 13L. Originally, this heading was numbered "1)" but the number was subsequently deleted and no quantities involving second-order derivatives of the metric are listed. This may be because Einstein was 13L interested in constructing a stress-energy tensor for the gravitational field, which is a quantity quadratic in first-order derivatives of the metric (see sec.).

In addition to the “order” (“Ordnung”), Einstein also considered the “degree” (“Grad”) and the “multiplicity” (“Mannigfaltigkeit”) of the expressions he constructed. The degree of an expression is the number of factors of $g_{\mu\nu}$ and $\gamma_{\mu\nu}$ it contains. Its multiplicity is simply its rank.[187] Einstein denoted every free index of the vectors and tensors he constructed either by a dot (for a contravariant index) or a dash (for a covariant index). A contravariant vector is accordingly called a “point vector” (“Punktvektor”), a covariant vector a “plane vector” (“Ebenenvektor”).[188] Similarly, a tensor with two covariant and one contravariant index, for instance, is called a “ · · – tensor.” Einstein distinguished two ways of forming such vectors and tensors, “internal” (“Innere”) and “exterior” (“Aussere”) differentiation. In the case of “inner” differentiation, the index of the derivative operator is contracted with one of the indices of the components of the metric, so that a four-divergence is formed. In the case of “outer” differentiation, the index of the derivative operator is different from the indices of the components of the metric, so that a four-gradient is formed. With this explanation of Einstein’s terminology the list on p. 13L becomes largely self-explanatory.

The first item on the list is the Hertz expression, a point vector of first order and first degree obtained through “internal” differentiation:

$$\sum \frac{\partial \gamma_{\mu\nu}}{\partial x_\nu} \qquad \text{Punktvektor.} \tag{400}$$

Note that $\gamma_{\mu\nu}$ cannot be replaced by $g_{\mu\nu}$ in this expression since that would involve contraction over two covariant indices. The next items on the list are therefore

$$\sum \frac{\partial \gamma_{\mu\nu}}{\partial x_i} \qquad \cdot \ \cdot \ - \ \text{Tensor,} \tag{401}$$

and[189]

$$\sum \frac{\partial g_{\mu\nu}}{\partial x_i} \qquad - \ - \ \cdot \ \text{Tensor,} \tag{402}$$

both obtained through “external” differentiation.

On the next line, Einstein turned to expressions of “first order” and “second degree” (“zweiten Grades”) and began by writing down the four different possible

---

187 Einstein used this same terminology on p. 8L (see equation (163)).

188 This terminology may have been inspired by Grassmann’s “Ausdehnungslehre” (Grassmann 1862). For evidence of Einstein’s reading of Grassmann in this period, see Einstein to Michele Besso, 13 May 1911 (CPAE 5, Doc. 267), and Einstein to Conrad Habicht, 7 July 1913 (CPAE 5, Doc. 450). Later in the notebook Einstein used this same terminology for tensors as well (see, e.g., pp. 17L–R).

189 Expression (402) is, in fact, in Einstein’s terminology, a $---$Tensor . Einstein inadvertently may have thought for a moment that the character of the three indices in expression (402) would be just the opposite of those in expression (401).

vectors of this kind that can be obtained by contracting expressions (400)–(402) with $g_{\mu\nu}$ and $\gamma_{\mu\nu}$ in various ways:[190]

$$\sum g_{\mu\sigma}\frac{\partial\gamma_{\mu\nu}}{\partial x_\nu} \qquad \text{Ebenenvektor}, \tag{403}$$

$$\sum g_{\mu\nu}\frac{\partial\gamma_{\mu\nu}}{\partial x_i} \qquad \text{Ebenenvektor}, \tag{404}$$

$$\sum \gamma_{\mu\nu}\frac{\partial g_{\mu\nu}}{\partial x_i} \qquad \text{Ebenenvektor}, \tag{405}$$

$$\sum \gamma_{\mu i}\frac{\partial g_{\mu\nu}}{\partial x_i} \qquad \text{Ebenenvektor}, \tag{406}$$

Expressions (404) and (405) are connected by a curly bracket. Not only are they equal to one another (because of the relation $\partial/\partial x_i(g_{\mu\nu}\gamma_{\mu\nu}) = 0$), they both vanish because of the restriction to unimodular coordinates (for which $G = 1$) and the relation

$$g_{\mu\nu}\frac{\partial\gamma_{\mu\nu}}{\partial x_i} = \frac{1}{G}\frac{\partial G}{\partial x_i} \tag{407}$$

(cf. equation (87)).

Finally, under the heading, "In addition the tensors of third multiplicity" ("Dazu die Tensoren dritter Mannigfaltigkeit"), Einstein wrote down all third-rank tensors of first order and second degree that can be constructed out of expressions (400)–(402). The Hertz expression (400) can be turned into a tensor of this kind through multiplication with either the covariant or the contravariant metric,

$$\sum {}^{g_{i\kappa}}_{\gamma_{i\kappa}}\frac{\partial\gamma_{\mu\nu}}{\partial x_\nu}, \tag{408}$$

giving, in Einstein's terminology, a "$- - \cdot$ tensor" and a "$\cdot\ \cdot\ \cdot$ tensor," respectively. Expressions (401) and (402) can be turned into third-rank tensors of first order and second degree by contracting them with $g_{\mu\nu}$ and $\gamma_{\mu\nu}$. This leads to the last four expressions on p. 13L:

$$\sum g_{\kappa\mu}\frac{\partial\gamma_{\mu\nu}}{\partial x_i} \qquad - \cdot - \ \text{[Tensor]}, \tag{409}$$

$$\sum \gamma_{\kappa i}\frac{\partial\gamma_{\mu\nu}}{\partial x_i} \qquad \cdot\ \cdot\ \cdot \ \text{[Tensor]}, \tag{410}$$

$$\sum \gamma_{\kappa\mu}\frac{\partial g_{\mu\nu}}{\partial x_i} \qquad \cdot - - \ \text{[Tensor]}, \tag{411}$$

190 Einstein omitted the summation sign in expression (406).

$$\sum \gamma_{\kappa i} \frac{\partial g_{\mu\nu}}{\partial x_i} \qquad \cdot -- \text{[Tensor]}. \tag{412}$$

4.6.2 *Trying to Find Correction Terms to the Core Operator to Guarantee Compatibility of the Field Equations With Energy-Momentum Conservation (13R)*

13R The starting point of the considerations on p. 13R is equation (74) for the energy-momentum balance between matter and gravitational field derived on p. 5R. This equation is equivalent to the vanishing of the covariant derivative of the matter stress-energy tensor $T_{\mu\nu}$. In unimodular coordinates (for which $\sqrt{G} = 1$) the left-hand side of equation (74) can be written as

$$\sum_{\mu\nu} \frac{\partial}{\partial x_\mu}(g_{m\nu} T_{\mu\nu}) - \frac{1}{2} \sum \frac{\partial g_{\mu\nu}}{\partial x_m} T_{\mu\nu}. \tag{413}$$

Inserting the core operator (see, e.g., expression (324) on p. 11R),

$$\gamma_{\alpha\beta} \frac{\partial^2 \gamma_{\mu\nu}}{\partial x_\alpha \partial x_\beta}, \tag{414}$$

for the contravariant stress-energy tensor $T_{\mu\nu}$ in expression (413) and adding an equality sign, one finds the first line of p. 13R:

$$\sum_{\alpha\beta\mu\nu} \frac{\partial}{\partial x_\mu}\left(g_{m\nu} \gamma_{\alpha\beta} \frac{\partial^2 \gamma_{\mu\nu}}{\partial x_\alpha \partial x_\beta}\right) - \frac{1}{2} \sum \frac{\partial g_{\mu\nu}}{\partial x_m} \gamma_{\alpha\beta} \frac{\partial^2 \gamma_{\mu\nu}}{\partial x_\alpha \partial x_\beta} = \quad . \tag{415}$$

On the next line, Einstein wrote “Third-order derivatives do not appear, if

$$\sum_\mu \frac{\partial \gamma_{\mu\nu}}{\partial x_\mu} = 0 \text{.”} \tag{416}$$

(“Dritte Ableitungen treten nicht auf, wenn … = 0 ist.”). Equation (416) is the by now familiar Hertz restriction. On the remainder of p. 13R, Einstein rewrote expression (415) using this restriction.

Why was Einstein interested in expression (415)? The simplest answer is that he wanted to find what further restrictions, if any, would be needed to guarantee that the field equations

$$\gamma_{\alpha\beta} \frac{\partial^2 \gamma_{\mu\nu}}{\partial x_\alpha \partial x_\beta} = \kappa T_{\mu\nu} \tag{417}$$

—understood either as exact or as weak-field equations—be compatible with energy-momentum conservation, i.e., with the vanishing of the covariant derivative of $T_{\mu\nu}$, or, in unimodular coordinates, the vanishing of expression (413). If the field equations (417) hold, the vanishing of expression (413) for $T_{\mu\nu}$ implies that expression (415) for the core operator (414) must also vanish. It is unlikely, however, that this was the point of Einstein’s considerations on p. 13R.

Einstein already knew that the exact field equations cannot be obtained by simply setting the core operator equal to the stress-energy tensor of matter as in equation (417). Energy-momentum conservation requires an additional term on the left-hand side that can be interpreted as gravitational energy-momentum density (see p. 9L and the discussion in sec. 4.4, especially equations (199)–(208)). In other words, Einstein expected the exact field equations to have the form

$$\gamma_{\alpha\beta}\frac{\partial^2\gamma_{\mu\nu}}{\partial x_\alpha \partial x_\beta} - \kappa t_{\mu\nu} = \kappa T_{\mu\nu}, \tag{418}$$

where the quantity $t_{\mu\nu}$, which represents gravitational energy-momentum, is assumed to be quadratic in first-order derivatives of the metric. In a weak-field approximation, this additional term can be neglected and equations (418) reduce to equations (417). On p. 13R, however, terms quadratic in first-order derivatives of the metric are *not* neglected. This strongly suggests that Einstein was implicitly using field equations of the form of equation (418). This in turn would mean that Einstein expected expression (415) to be equal, not to zero, but to

$$\sum\frac{\partial}{\partial x_\mu}(g_{m\mu}\kappa t_{\mu\nu}) - \frac{1}{2}\sum\frac{\partial g_{\mu\nu}}{\partial x_m}\kappa t_{\mu\nu}. \tag{419}$$

The equality of expressions (415) and (419) would guarantee the compatibility of energy-momentum conservation (in the form of the vanishing of expression (413)) and field equations of the form (418). Presumably, what Einstein tried to do on p. 13R was to rewrite expression (415) in the form of equation (419) and identify $t_{\mu\nu}$. For one thing, this would explain his concern with the elimination of third-order derivatives from expression (415). Since $t_{\mu\nu}$ only contains first-order derivatives of the metric, expression (419) will contain no derivatives higher than of second order. As we pointed out in sec. 4.6.1, this may also be why Einstein only listed quantities of "first order" ("Erster Ordnung") on p. 13L.

Note that on this reading Einstein must have come to realize that gravitational energy-momentum has a special status. On p. 9L he had still demanded that the quantity $t_{\mu\nu}$ representing gravitational energy-momentum satisfy equation (74) posited for *all* energy-momentum on p. 5R (cf. footnote 128). In that case the right-hand side of equation (415) would simply be equal to zero rather than to expression (419).

When the Hertz restriction (416) is imposed, the left-hand side of equation (415) reduces to

$$\frac{\partial}{\partial x_\mu}(g_{m\nu}\gamma_{\alpha\beta})\frac{\partial^2\gamma_{\mu\nu}}{\partial x_\alpha \partial x_\beta} - \frac{1}{2}\frac{\partial g_{\mu\nu}}{\partial x_m}\gamma_{\alpha\beta}\frac{\partial^2\gamma_{\mu\nu}}{\partial x_\alpha \partial x_\beta}, \tag{420}$$

which no longer contains any third-order derivatives of the metric. Regrouping terms, one can rewrite this expression in the way Einstein wrote it on the third line of p. 13R:

$$\sum\left(\frac{\partial g_{m\nu}}{\partial x_\mu} - \frac{1}{2}\frac{\partial g_{\mu\nu}}{\partial x_m}\right)\gamma_{\alpha\beta}\frac{\partial^2\gamma_{\mu\nu}}{\partial x_\alpha \partial x_\beta} + \sum_{\mu\nu\alpha\beta} g_{m\nu}\frac{\partial\gamma_{\alpha\beta}}{\partial x_\mu}\frac{\partial^2\gamma_{\mu\nu}}{\partial x_\alpha \partial x_\beta}. \tag{421}$$

Einstein concentrated on the second summation in this expression. Using

$$\frac{\partial}{\partial x_\mu}\Big(\sum_\nu g_{\lambda\nu}\gamma_{\mu\nu}\Big) = 0 \tag{422}$$

and the Hertz restriction (416), he arrived at:

$$\sum \gamma_{\mu\nu}\frac{\partial g_{\lambda\nu}}{\partial x_\mu} = 0\,. \tag{423}$$

He tried to rewrite the second summation in expression (421) in such a way that he could apply equation (423) to the term $(\partial g_{m\nu}/\partial x_\beta)\gamma_{\mu\nu}$, but quickly realized that this was not feasible. He began by pulling out the differentiation with respect to x_α, thus arriving at:[191,192]

$$\frac{\partial}{\partial x_\alpha}\left(g_{m\nu}\frac{\partial\gamma_{\alpha\beta}}{\partial x_\mu}\frac{\partial\gamma_{\mu\nu}}{\partial x_\beta}\quad\right) - \frac{\partial g_{\mu\nu}}{\partial x_\alpha}\frac{\partial\gamma_{\alpha\beta}}{\partial x_\mu}\frac{\partial\gamma_{\mu\nu}}{\partial x_\beta}\,. \tag{424}$$

He then used relation (422) to rewrite the first term as:

$$-\frac{\partial}{\partial x_\alpha}\left(\frac{\partial g_{m\nu}}{\partial x_\beta}\gamma_{\mu\nu}\frac{\partial\gamma_{\alpha\beta}}{\partial x_\mu}\right)\,. \tag{425}$$

Here the calculation breaks off.

Einstein did not return to considerations of energy-momentum conservation until p. 19R. At that point he had further deepened his understanding of the special status of gravitational energy-momentum. Interpreting the second term in expression (413) as the gravitational force density, he tried to rewrite that term as the divergence of the quantity $t_{\mu\nu}$ representing the gravitational energy-momentum density (see the discussion following equation (484) in sec. 5.4.2). Later in the notebook (on p. 24R and, more systematically, on pp. 26L-R), he used this insight to derive field equations that are automatically compatible both with energy-momentum conservation and with Newtonian theory for static weak fields (see sec. 5.6.1 and sec. 6). This led him to the *Entwurf* field equations. The notion that energy-momentum conservation requires coordinate restrictions over and above the ones needed to recover Newtonian theory for static weak fields stayed with Einstein right up until his introduction of generally-covariant field equations in November 1915.[193]

191 Einstein drew a line connection expression (424) with the second summation in expression (421).

192 In this equation $\partial g_{\mu\nu}/\partial x_\alpha$ should be $\partial g_{m\nu}/\partial x_\alpha$. Because of the Hertz restriction there is no term with $\partial^2\gamma_{\alpha\beta}/\partial x_\alpha \partial x_\mu$.

193 Einstein later compared the Hertz restriction with coordinate restrictions in (Einstein and Grossmann 1914), which not only circumscribe the covariance of the *Entwurf* theory but also guarantee energy-momentum conservation (see Einstein to Paul Hertz, 22 August 1915 [CPAE 8, Doc. 111]). For further discussion of the role of coordinate restrictions in determining covariance properties, recovering Newtonian theory, and guaranteeing energy-momentum conservation, see “Untying the Knot ...” sec. 1.1 (in this volume).

5. EXPLORATION OF THE RIEMANN TENSOR (14L–25R, 42L–43L)

5.1 *Introduction* (14L–25R, 42L–43L)

14L–25R, 42L–43L

A new stage in Einstein's search for gravitational field equations began on p. 14L of the notebook with the systematic exploration of the Riemann tensor along the lines of the mathematical strategy.[194] In the course of this exploration, Einstein considered various gravitational field equations based on the Ricci tensor that he would publish in his communications to the Prussian Academy of November 1915 (Einstein 1915a, b, d). He even considered, albeit only in linear approximation, the crucial trace term that occurs in the final version of the field equations. However, the episode that, from a modern point of view, begins so promisingly on p. 14L with the introduction of the Riemann tensor ends disappointingly on p. 26L with the derivation of the *Entwurf* field equations along the lines of the physical strategy. What happened on these pages that made Einstein abandon the mathematical strategy and return to the physical strategy?

The analysis of pp. 14L–25R and related material on pp. 42L–43L, the last three pages of the part starting from the other end of the notebook, reveals a pattern that holds the key to our answer to this question. Einstein's starting point invariably is some expression of broad if not general covariance constructed out of the Riemann tensor. To extract from these expressions field equations that reduce to the Poisson equation of Newtonian theory in the special case of weak static fields, Einstein introduced various coordinate restrictions.[195] With the help of these he could eliminate unwanted terms with second-order derivatives of the metric. The left-hand sides of the resulting field equations consist of a term with a core operator (i.e., a term that, in linear approximation, reduces to the d'Alembertian acting on the metric), and terms quadratic in first-order derivatives of the metric, which vanish in linear approximation. Einstein then checked whether these field equations and the coordinate restrictions used in their construction satisfy his other heuristic requirements,[196] in particular whether they are compatible with energy-momentum conservation and whether they are covariant under a wide enough class of coordinate transformations (autonomous or non-autonomous[197]) to be compatible with the equivalence principle and a generalized relativity principle. All candidates considered by Einstein failed at least one of these tests. Finding coordinate restrictions of sufficiently broad covariance turned out to be particularly difficult.

---

194 See sec. 1.2 for a characterization of the difference between the mathematical and the physical strategy.

195 See sec. 4.1 for a discussion of the notion of a coordinate restriction.

196 See sec. 1.1 for a discussion of what we identified as the four major heuristic principles: the correspondence principle, the conservation principle, the relativity principle and the equivalence principle.

197 See sec. 4.1 for the definitions of the notions of autonomous and non-autonomous and transformations.

Eventually, Einstein switched back to the physical strategy. He developed the considerations of energy-momentum conservation on p. 13R into a method for generating field equations guaranteed both to have the desired form in linear approximation, and to be compatible with energy-momentum conservation. Field equations constructed in this manner have the same general form as the candidate field equations that Einstein had extracted from the Riemann tensor by eliminating unwanted second-order derivative terms with the help of coordinate restrictions. This suggested that the field equations generated by this new method could also be produced by the mathematical strategy. It remained unclear, of course, from which generally-covariant expression they could be extracted and whether the necessary coordinate restrictions would be of sufficiently broad covariance and would themselves be compatible with energy-momentum conservation. However, it must have been encouraging that the physical strategy yielded field equations, satisfactory on all other counts, of exactly the form that Einstein had come to expect while pursuing the mathematical strategy in his exploration of the Riemann tensor. And as far as the unknown covariance properties of the equations were concerned, the mathematical strategy, for all its promise on this score, had not allowed Einstein to make any real progress either. It thus becomes understandable that Einstein eventually gave up the idea of constructing field equations out of the Riemann tensor and instead adopted the *Entwurf* equations.

5.2 *General Survey (14L–25R, 42L–43L)*

14L At the top of p. 14L Einstein wrote down the fully covariant form of the Riemann tensor. He wrote Grossmann's name right next to it, suggesting that it was Grossmann who had drawn his attention to this tensor and its importance. Einstein proceeded to form the covariant form of the Ricci tensor by contracting the Riemann tensor over two of its four indices. He looked at the terms involving second-order derivatives of the metric and immediately noticed that in addition to a core-operator term the Ricci tensor contains three other second-order derivative terms that should not occur in the Newtonian limit. So the natural way of extracting a second-rank tensor from the Riemann tensor did not seem to produce a suitable candidate for the left-hand side of the field equations (sec. 5.3.1[198]).

14R–16R On pp. 14R–16R, Einstein explored a different way of extracting a two-index object from the Riemann tensor that, if not generally covariant, at least promised to be a tensor under unimodular transformations. In close analogy to his treatment of the Beltrami invariant on p. 9L, he computed the curvature scalar by fully contracting the Riemann tensor. Setting the determinant of the metric equal to unity, he succeeded in rewriting the curvature scalar as the contraction of the metric with a symmetric two-index object denoted (on p. 16L) by T_{ik}. Einstein presumably expected that T_{ik}

198 In this general survey we shall give a reference at the end of each brief summary of a calculation or a line of reasoning in this part of the notebook to the subsection where the calculation or the argument just summarized is discussed in much greater detail.

would turn out to be a tensor under unimodular transformations. Unfortunately, T_{ik} still contains two terms with unwanted second-order derivatives of the metric in addition to a core-operator term (sec. 5.3.2).

On the following pages (pp. 17L–18R), Einstein investigated the relation between the two expressions that he had formed out of the Riemann tensor. Since T_{ik} gives the curvature scalar when contracted with the covariant metric, it is itself a contravariant object. To facilitate comparison between T_{ik} and the covariant Ricci tensor, Einstein (on p. 17L) first formed the contravariant version of the latter. He tried to simplify the resulting expression using that covariance under unimodular transformations is all that matters for the comparison with T_{ik}. This meant that he could assume the determinant of the metric to be a constant. He abandoned this calculation as "too involved." On p. 17R he formed the covariant version of T_{ik} instead and started to bring the covariant Ricci tensor into a form in which it could be compared with this version of T_{ik}, again assuming the determinant of the metric to be a constant. This calculation also turned out to be complicated, and was abandoned as well (sec. 5.3.3). 17L–18R

Since neither the Ricci tensor nor T_{ik} had the form required by the correspondence principle, Einstein began to investigate the possibility of obtaining suitable candidates for the left-hand side of the field equations by restricting the range of admissible coordinates. On p. 19L he showed that the terms in the Ricci tensor with unwanted second-order derivatives can be eliminated by imposing the harmonic coordinate restriction (sec. 5.4.1).[199] He then checked whether these field equations and this coordinate restriction are compatible with his other heuristic requirements. 19L

On p. 19R, Einstein examined in linear approximation the harmonic restriction and the field equations constructed with the help of it. Einstein confirmed that the weak-field field equations are compatible with his conservation principle: with the help of these equations, the term giving the gravitational force density in the energy-momentum balance between matter and gravitational field can be written as the divergence of a quantity representing gravitational stress-energy density. To guarantee compatibility between the weak-field equations and energy-momentum conservation, however, Einstein had to impose an additional coordinate restriction, a linear approximation of the Hertz restriction.[200] Together, the harmonic restriction and the Hertz restriction imply that the trace of the weak-field metric has to vanish. In turn, this implies that the trace of the stress-energy tensor for matter has to vanish (sec. 5.4.2). 19R

To avoid these consequences, Einstein (on p. 20L) modified the weak-field equations by adding a term proportional to the trace of the weak-field metric. This term was 20L

199 Considering p. 19L in isolation, one would think that Einstein was simply applying the harmonic coordinate condition in the modern sense to recover the Poisson equation in the limit of weak static fields. This interpretation is even compatible with the whole passage dealing with the Ricci tensor in harmonic coordinates (pp. 19L–21R). The interpretation is incompatible, however, with Einstein's usage of coordinate conditions elsewhere in the notebook, both on pages preceding and on pages following the examination of the harmonic coordinate condition (cf. especially p. 23L). These other pages suggest that what Einstein had in mind throughout the notebook were coordinate restrictions rather than coordinate conditions in the modern sense. Cf., e.g., our discussion at the end of sec. 5.5.4.

200 See sec. 4.1 for the introduction of this coordinate restriction.

introduced in such a way that Einstein could now use the harmonic restriction to satisfy both the conservation principle and the correspondence principle. From a purely mathematical point of view, the resulting weak-field equations are the Einstein field equations of the final theory of November 1915 in linear approximation. As Einstein
21L checked briefly on p. 20L and more carefully on p. 21L, with these modified field equations the gravitational force density can still be written as the divergence of gravitational stress-energy density (sec. 5.4.3).

The modified weak-field equations, however, do not allow the spatially flat metric that Einstein continued to use to represent static fields.[201] Given this disparity, Einstein reconsidered his presupposition concerning the form of the metric of weak static
21R fields. On p. 21R he presented a seductive but ultimately fallacious argument that corroborated his prior beliefs on this point. The argument was based on the dynamics of point particles (recapitulated on p. 20R) and Galileo's principle that all bodies fall with the same acceleration in a given gravitational field. This powerful physical argument seemed to rule out the harmonic restriction (secs. 5.4.4 and 5.4.6).

22R Einstein, however, was not ready to give up his attempt to extract the left-hand side of the field equations from the Riemann tensor. On p. 22R, at the suggestion of Grossmann perhaps, whose name once again appears at the top of the page, he turned to a different coordinate restriction that might help him achieve his goal. First, he noticed that the Ricci tensor can be split into two parts, each of which by itself transforms as a tensor under unimodular transformations. Einstein took one of these as his new candidate for the left-hand side of the field equations. We call this part the November tensor, because it is the left-hand side of the field equations published in the first of Einstein's four papers of November 1915[202] (sec. 5.5.1). The November tensor still contains terms with unwanted second-order derivatives of the metric. Einstein eliminated these by imposing the Hertz restriction. The calculations on p. 19R had shown that the Hertz restriction is compatible with energy-momentum conservation in the weak-field case without the need for an additional trace term in the weak-field equations.[203]

22L Given Einstein's understanding of the status of coordinate restrictions at the time, the logical next step was to investigate the group of transformations allowed by the Hertz restriction. On the facing page (p. 22L) Einstein did indeed derive the condition for non-autonomous transformations leaving the Hertz restriction invariant. Earlier in the notebook (pp. 10L–11L), he had already found that the Hertz restriction rules out uniform acceleration in the important special case of Minkowski spacetime (sec. 5.5.3).[204] He nonetheless held on to the Hertz restriction for the time being.

---

201 This form of the static metric is also incompatible with the harmonic coordinate restriction, but it is unclear whether Einstein realized that at this point.

202 (Einstein 1915a).

203 Unlike the harmonic restriction, the Hertz restriction as well as the restriction to unimodular transformations are compatible with Einstein's assumptions concerning the form of the weak-field static metric (cf. footnote 201 above).

The elimination of terms with unwanted second-order derivatives of the metric from the November tensor had left Einstein (at the bottom of p. 22R) with a candidate for the left-hand side of the field equations containing numerous terms quadratic in first-order derivatives. On p. 23L he added another coordinate restriction to the Hertz 23L
restriction to eliminate some of these terms. Just as he had obtained the November tensor by splitting the Ricci tensor into two parts each of which transforms as a tensor under unimodular transformations, Einstein (on p. 23L) obtained yet another candidate for the left-hand side of the field equations by splitting the November tensor into various parts, each of which transforms as a tensor under a class of unimodular transformations under which a quantity which we shall call the ϑ-expression transforms as a tensor. Restricting the allowed transformations to such ϑ-transformations, Einstein found that he could eliminate all but one of the terms quadratic in first-order derivatives of the metric from the November tensor. He furthermore discovered that the restriction to ϑ-transformations sufficed to eliminate the terms with unwanted second-order derivatives as well. There was no need to add the Hertz restriction to the ϑ-restriction. Einstein thus abandoned the Hertz restriction and focused on the ϑ-restriction instead.

On p. 23R Einstein began to investigate the covariance properties of the ϑ-expres- 23R
sion. As with the Hertz restriction, he derived the condition for non-autonomous transformations leaving the ϑ-expression invariant (sec. 5.5.5). He did not attempt, however, to find the most general (non-autonomous) transformations satisfying this condition. Instead, he focused on the important special case of a vanishing ϑ-expression. The ϑ-expression vanishes, for instance, for the Minkowski metric in standard diagonal form. According to Einstein's heuristic principles, it should therefore also vanish for the Minkowski metric in accelerated frames of reference (sec. 5.5.6). It was thus natural for Einstein to investigate what metric fields are allowed by the condition that the ϑ-expression vanish.

On pp. 42L–43L,[205] Einstein addresses just this problem, further limiting himself 42L–43L
to time-independent metric fields and hence to uniformly accelerated frames of reference. The main result of his investigation was both promising and puzzling. He discovered that the ϑ-expression vanishes for a metric, which we shall call the ϑ–metric, whose covariant components are the contravariant components of the rotation metric, i.e., the Minkowski metric in rotating coordinates (sec. 5.5.6). Einstein tried to come to terms with this result in various ways.

First, on the bottom half of p. 42R and again at the bottom of p. 43LA, he inserted 42R–43LA
the suggestive ϑ–metric into the Lagrangian for a point particle moving in a metric field and began to compute the Euler-Lagrange equations. Presumably, the idea was to check whether the components of the ϑ–metric and its derivatives could be given

204 Contrary to what Einstein had concluded on p. 11L (see sec. 4.5.2), the Hertz restriction also rules out finite transformation to rotating frames in Minkowski spacetime.

205 These pages are found toward the end of the part starting from the other end of the notebook. At the beginning of sec. 5.5.6 we address the question of the temporal order of the calculations on pp. 42L–43L at one end of the notebook and those on pp. 23L–24L at the other.

a physical interpretation in terms of centrifugal and Coriolis forces. Einstein broke off this calculation without reaching a definite conclusion (sec. 5.5.7).

24L A variant of this approach can be found on p. 24L, the page immediately following the introduction of the ϑ-restriction. This time he took the energy-momentum balance between matter and gravitational field as his starting point rather than the Lagrangian for a point particle moving in a metric field (sec. 5.5.9).

43LA Einstein tried a different approach on p. 43LA. He replaced the covariant components of the metric in the ϑ-expression by the corresponding contravariant ones. Since the original ϑ-expression vanishes for the rotation metric if only its co- and contravariant components are interchanged, the new version will vanish for the rotation metric itself. This approach did not work either. After three failed attempts to come to terms with the ϑ-metric, Einstein gave up the ϑ-restriction, though not the idea of extracting field equations from the November tensor (sec. 5.5.8).

24R Although the actual calculation has not been preserved in the notebook, there are strong indications that Einstein at this point did for field equations extracted from the November tensor what he had done earlier (on pp. 19R, 20L, and 21L) for field equations extracted directly from the Ricci tensor. He confirmed that in linear approximation these field equations can be used to write the gravitational force density as the divergence of a quantity representing gravitational stress-energy density. At the top of p. 24R, Einstein wrote down an expression that is most naturally interpreted as the end result of this calculation and noted that it vanishes for the rotation metric. Given some errors in the expression for the rotation metric on p. 24R, one can understand how Einstein reached this conclusion. When these errors are corrected, one sees that the expression, in fact, does not vanish for the rotation metric (see the first half of sec. 5.6.1).

It is the next step that Einstein took on p. 24R that marks the beginning of Einstein's return to the physical strategy abandoned on p. 13R. So far Einstein had only verified in linear approximation that various candidate field equations allowed him to write the gravitational force density as the divergence of the gravitational stress-energy density. It was not at all clear whether this result would also hold exactly. On p. 24R Einstein introduced an ingenious new approach to this problem. Rather than starting from some candidate field equations and using them to rewrite the gravitational force density as a divergence without neglecting terms of one order or another, Einstein started from the expression for the divergence of the stress-energy density obtained in linear approximation and determined which higher-order terms need to be added to the linearized field equations such that this divergence becomes exactly equal to the gravitational force density. This is the strategy that Einstein used to derive the *Entwurf* equations (see pp. 26L–R).[206] It is in this context that it was very important for Einstein to check whether the expression at the top of p. 24R would vanish exactly for the rotation metric. This is a necessary condition for the rotation metric to be a

206 For a comparison between this method of finding candidate field equations and the method used on pp. 19L–23L, see the introduction to sec. 5.6.

solution of field equations constructed with the help of the new strategy. Once he had derived the new field equations, he checked directly whether the rotation metric is a solution and discovered that it is not. However, he also discovered that he had erroneously cancelled two terms in his construction of the new field equations which opened up the possibility that the rotation metric would be a solution of the corrected equations (see the second half of sec. 5.6.1).

On p. 25L Einstein tried to recover his new candidate field equations from the 25L
November tensor by imposing an appropriate coordinate restriction. It seems to have been Einstein's hope at this point that he could derive the same field equations following either the physical or the mathematical strategy (sec. 5.6.2).

Material at the top of p. 25R shows that Einstein considered a variant of the ϑ- 25R
restriction, which we shall call the $\hat{\vartheta}$-restriction, to extract field equations from the November tensor. It is unclear whether he was still trying to recover the field equations found on p. 24R in this way. Einstein went back to his calculations on p. 23L, where he had applied the original ϑ-restriction to the November tensor, and indicated the changes that would need to be made if the ϑ-restriction were replaced by the $\hat{\vartheta}$-restriction. Returning to p. 25R, he checked whether the $\hat{\vartheta}$-expression vanishes for the rotation metric, as required by the equivalence and relativity principles. He discovered that it does not, which is probably why he abandoned the $\hat{\vartheta}$-restriction (sec. 5.6.3). This marks the end of Einstein's pursuit of the mathematical strategy in the notebook.

On the remainder of p. 25R, Einstein started tinkering with the field equations he had found on p. 24R to ensure that the rotation metric would be a solution. Einstein convinced himself that a slightly modified version of the equations satisfies this requirement. These modified equations, however, are not well-defined mathematically (they involve contractions over pairs of covariant indices among other things). It is unclear whether Einstein came to recognize this (sec. 5.6.4). He abandoned these ill-defined equations when he found that it is not possible to rewrite the gravitational force density as the divergence of gravitational stress-energy density with their help (sec. 5.6.5).

On the next page (p. 26L), Einstein made a fresh start with the physical strategy. A possible indication that he had meanwhile abandoned his hope of recovering field equations found in this manner from the Riemann tensor is that he no longer set the determinant of the metric equal to unity, as he had still done on pp. 24R–25R in the hope perhaps to connect the field equations of p. 24R to the November tensor. The exploration of the Riemann tensor had nonetheless been fruitful (independently of the developments of November 1915). Even though it had failed to produce satisfactory field equations with a well-defined covariance group, it had given Einstein a clear idea of the structure any such field equations would have after unwanted terms in the Ricci tensor or the November tensor had been eliminated by imposing the appropriate coordinate restrictions. He had found a strategy to generate field equations of this form that automatically satisfy the correspondence and conservation principles. Given the difficulties Einstein had run into trying to make sure that the relativity and equivalence principles are satisfied as well, one can understand that he decided to bracket those problems for the time being.

14L–18R

5.3 *First Attempts at Constructing Field Equations out of the Riemann Tensor* (14L–18R)

On p. 14L the Riemann tensor appears for the first time in the notebook. Einstein immediately considered the Ricci tensor as a candidate for the left-hand side of the field equations, but ran into the problem that it contains unwanted terms with second-order derivatives of the metric. On the following pages (pp. 14R–18R), Einstein extracted another candidate for the left-hand side of the field equations from the curvature scalar and compared the result to the Ricci tensor. Unfortunately, this new candidate also contains unwanted second-order derivative terms. Moreover, its relation to the Ricci tensor remained unclear.

14L

5.3.1 *Building a Two-Index Object by Contraction: the Ricci Tensor (14L)*

At the top of p. 14L Einstein wrote down the definitions of the Christoffel symbols of the first kind,

$$\begin{bmatrix} \mu\ \nu \\ l \end{bmatrix} = \frac{1}{2}\left(\frac{\partial g_{\mu l}}{\partial x_\nu} + \frac{\partial g_{\nu l}}{\partial x_\mu} - \frac{\partial g_{\mu\nu}}{\partial x_l}\right), \tag{426}$$

and of the fully covariant Riemann tensor,

$$\begin{aligned}(i\kappa, lm) = {} & \frac{1}{2}\left(\frac{\partial^2 g_{im}}{\partial x_\kappa \partial x_l} + \frac{\partial^2 g_{\kappa l}}{\partial x_i \partial x_m} - \frac{\partial^2 g_{il}}{\partial x_\kappa \partial x_m} - \frac{\partial^2 g_{\kappa m}}{\partial x_i \partial x_l}\right) \\ & + \sum_{\rho\sigma} \gamma_{\rho\sigma}\left(\begin{bmatrix} i\ \ m \\ \sigma \end{bmatrix}\begin{bmatrix} \kappa\ \ l \\ \rho \end{bmatrix} - \begin{bmatrix} i\ \ l \\ \sigma \end{bmatrix}\begin{bmatrix} \kappa\ \ m \\ \rho \end{bmatrix}\right).\end{aligned} \tag{427}$$

Next to the Riemann tensor Einstein wrote “Grossmann Tensor vierter Mannigfaltigkeit,” indicating that it was his friend and colleague Marcel Grossmann who introduced him to these mathematical objects. Grossmann’s sources for the Christoffel symbols and the Riemann tensor[207] were in all probability (Christoffel 1869) and (Riemann 1867).[208] These are the sources cited for the Riemann tensor in Grossmann’s section of the *Entwurf* paper,[209] where it appears in almost identical form.[210] A comparison of the notation used for the Christoffel symbols and the Riemann tensor

207 It should be emphasized that by calling the expressions (426) and (427) “Christoffel symbols” and “Riemann tensor,” we do not mean to suggest that these were the terms used by contemporary authors.

208 For further discussion of Riemann’s paper, see, e.g., (Reich 1994, sec. 2.1.3). Grossmann was aware of Riemann’s and Beltrami’s work in non-Euclidean geometry at least as early as 1904, as the introductory paragraph of (Grossmann 1904) shows.

by various contemporary authors[211] further supports this conjecture.[212] The square brackets for the Christoffel symbols were introduced by Christoffel himself, and, according to the survey in (Reich 1994), this notation was taken up again only in 1912 by Friedrich Kottler and then in the *Entwurf* paper itself, which cites (Kottler 1912) (Einstein and Grossmann 1913, 23 and 30). The notation for the Riemann tensor used in Christoffel's paper is $(gkhi)$, i.e., it does not have the comma, which appears in equation (427). The comma was used by Riemann whose notation was $(\iota\iota', \iota''\iota''')$. Also the occurrence of the word "Mannigfaltigkeit," a phrase used by Riemann that does not appear anywhere in Christoffel's paper, suggests that Einstein and Grossmann consulted Riemann's paper in addition to Christoffel's.

There is yet another indication, however, that the actual expressions were directly taken from Christoffel's paper. In the top right corner, Einstein wrote the expression

$$\frac{\partial}{\partial x_\kappa}\begin{bmatrix} i & l \\ & m \end{bmatrix} - \frac{\partial}{\partial x_i}\begin{bmatrix} \kappa & l \\ & m \end{bmatrix}, \tag{428}$$

which gives exactly the four second-order derivative terms of the Riemann tensor in equation (427), if the definition of the Christoffel symbols in equation (426) is inserted. Just this derivation of the Riemann tensor in terms of derivatives and products of the Christoffel symbols was given on the relevant page (p. 54) of (Christoffel 1869), where the Riemann tensor was first published.

The Riemann tensor formed a promising starting point for Einstein's search for gravitational field equations. The first task was to construct from the fourth-rank Riemann tensor an expression that could be used as the left-hand side of field equations with the second-rank stress-energy tensor on the right-hand side.

---

209 (Einstein and Grossmann 1913, 35). In the case of Riemann the reference is to "Riemann, Ges. Werke, S. 270". The page number does not make sense for the first edition of 1876 nor for the second edition of 1892. In the first edition Riemann's "Commentatio" starts on p. 370; in the second on p. 91. It seems probable that Einstein and Grossmann used the first edition and that "270" is a misprint and should be "370." Among other things, the second edition differs from the first in the German notes to the latin text of the "Commentatio." Unfortunately, the reference in the *Entwurf* paper is Einstein's most explicit reference to Riemann. In (Einstein and Fokker 1914, p. 325), for instance, he simply referred to the "bekannten Riemann-Christoffelschen Tensor." In (Einstein 1914b, 1053) and in (Einstein 1916, 799), the phrase "Riemann-Christoffel tensor" occurs in (sub)section headings without any reference to the literature.

210 In the *Entwurf* paper, the indices ρ and σ in the second part of equation (427) are interchanged (Einstein and Grossmann 1913, 35).

211 See (Reich 1994, 232) for a survey of the tensor analytic notation employed by various authors.

212 Some indirect confirmation of the conjecture comes from the draft of a letter from Felix Klein to Einstein of 20 March 1918. Commenting on a set of lecture notes he promised to send Einstein, Klein wrote: "You will probably immediately agree with what I have to say about Riemann, Beltrami, and Lipschitz; it seems to me that Grossmann at the time instructed you too much from the point of view of the school of Christoffel more narrowly" ("Was ich von Riemann, Beltrami, und Lipschitz erzähle, wird wohl gleich ihren Beifall haben; es scheint mir, dass Grossmann Sie s. Z. zu einseitig vom Standpunkte der engeren Christoffelschen Schule aus instruiert hat." CPAE 8, Doc. 487, note 26).

The most natural way to obtain such an expression is to contract over two indices, thus forming the second-rank Ricci tensor. This is exactly what Einstein did on the next line

$$\sum \gamma_{\kappa l}(i\kappa, lm)\,. \tag{429}$$

He put a question mark next to this expression, presumably to indicate that he wanted to check whether this would be an acceptable candidate for the left-hand side of the field equations

It can be seen immediately that there is a problem with the correspondence principle. Upon contraction, the first of the four second-order derivative terms in equation (427) gives the core-operator term

$$\sum \gamma_{\kappa l}\frac{\partial^2 g_{im}}{\partial x_\kappa \partial x_l}\,, \tag{430}$$

but the other three terms should not occur in the Newtonian limit. In fact, at the bottom of the page, underneath a horizontal line, Einstein wrote down these three bothersome terms

$$\sum_\kappa \left(\frac{\partial^2 g_{\kappa\kappa}}{\partial x_i \partial x_m} - \frac{\partial^2 g_{i\kappa}}{\partial x_\kappa \partial x_m} - \frac{\partial^2 g_{m\kappa}}{\partial x_\kappa \partial x_i}\right)\,, \tag{431}$$

set them equal to zero, and remarked that they "should vanish" ("Sollte verschwinden").[213]

Einstein not only considered the second-order derivative terms in the Ricci tensor, he also started to rewrite the first-order derivative terms. First, he introduced the relation[214]

$$\begin{aligned}\sum \gamma_{\kappa l}\begin{bmatrix}\kappa & l\\ & \rho\end{bmatrix} &= \sum \gamma_{\kappa l}\left[\frac{\partial g_{\kappa\rho}}{\partial x_l} + \frac{\partial g_{l\rho}}{\partial x_\kappa} - \frac{\partial g_{\kappa l}}{\partial x_\rho}\right]\\ &= -\frac{\partial \lg G}{\partial x_\rho} + 2\sum_{\kappa l}\gamma_{\kappa l}\frac{\partial g_{\kappa\rho}}{\partial x_l}\,,\end{aligned} \tag{432}$$

where G is the determinant of the metric.[215] He inserted (the corrected version of[216]) this relation into one of the two terms in the Ricci tensor with a product of Christoffel symbols

213 In deriving this expression, Einstein apparently assumed a weak-field metric with components $\mathrm{diag}(1, 1, 1, 1)$ to zeroth order, which would explain why the contracting metric does not appear explicitly in the equation.

214 A factor $1/2$ is missing in the expressions following the equality signs.

215 The introduction of G into the expression for the Ricci tensor would come to play an important role on p. 22R.

216 See footnote 214.

$$
\begin{aligned}
&\sum \gamma_{\kappa l}\gamma_{\rho\sigma}\begin{bmatrix} i & m \\ & \sigma \end{bmatrix}\begin{bmatrix} \kappa & l \\ & \rho \end{bmatrix} \\
&= \frac{1}{4}\sum \gamma_{\rho\sigma}\left(\frac{\partial g_{i\sigma}}{\partial x_m} + \frac{\partial g_{m\sigma}}{\partial x_i} - \frac{\partial g_{im}}{\partial x_\sigma}\right)\left[-\frac{\partial \lg G}{\partial x_\rho} + 2\sum_{\kappa l}\gamma_{\kappa l}\frac{\partial g_{\kappa\rho}}{\partial x_l}\right].
\end{aligned}
\tag{433}
$$

In the next step, without writing down their definition, Einstein used the Christoffel symbols of the second kind,

$$
\begin{Bmatrix} i & m \\ & \rho \end{Bmatrix} = \gamma_{\rho\sigma}\begin{bmatrix} i & m \\ & \sigma \end{bmatrix} = \frac{1}{2}\gamma_{\rho\sigma}\left(\frac{\partial g_{i\sigma}}{\partial x_m} + \frac{\partial g_{m\sigma}}{\partial x_i} - \frac{\partial g_{im}}{\partial x_\sigma}\right), \tag{434}
$$

again following the notation of Christoffel's 1869 paper. With the help of equations (432) and (434), and the following relation between Christoffel symbols of the first and the second kind,

$$
\begin{aligned}
\gamma_{\kappa l}\begin{bmatrix} \kappa & m \\ & \rho \end{bmatrix} &= \frac{1}{2}\gamma_{\kappa l}\left(\frac{\partial g_{\kappa\rho}}{\partial x_m} + \frac{\partial g_{m\rho}}{\partial x_\kappa} - \frac{\partial g_{\kappa m}}{\partial x_\rho}\right) \\
&= \gamma_{\kappa l}\frac{\partial g_{\kappa\rho}}{\partial x_m} + \frac{1}{2}\gamma_{\kappa l}\left(\frac{\partial g_{m\rho}}{\partial x_\kappa} - \frac{\partial g_{\kappa m}}{\partial x_\rho} - \frac{\partial g_{\kappa\rho}}{\partial x_m}\right) \\
&= \gamma_{\kappa l}\frac{\partial g_{k\rho}}{\partial x_m} - \gamma_{\kappa l}\begin{bmatrix} m & \rho \\ & \kappa \end{bmatrix} \\
&= \gamma_{\kappa l}\frac{\partial g_{\kappa\rho}}{\partial x_m} - \begin{Bmatrix} m & \rho \\ & l \end{Bmatrix},
\end{aligned}
\tag{435}
$$

he rewrote the relevant part of the Ricci tensor as[217]

$$
\begin{aligned}
&\sum \gamma_{\kappa l}\gamma_{\rho\sigma}\left(\begin{bmatrix} i & m \\ & \sigma \end{bmatrix}\begin{bmatrix} \kappa & l \\ & \rho \end{bmatrix} - \begin{bmatrix} i & l \\ & \sigma \end{bmatrix}\begin{bmatrix} \kappa & m \\ & \rho \end{bmatrix}\right) \\
&= \sum_\rho - \begin{Bmatrix} i & m \\ & \rho \end{Bmatrix}\frac{\partial \lg G}{\partial x_\rho} + 2\sum_{\kappa l\rho}\begin{Bmatrix} i & m \\ & \rho \end{Bmatrix}\gamma_{\kappa l}\frac{\partial g_{\kappa\rho}}{\partial x_l} \\
&\quad - \sum_{\rho l\kappa}\begin{Bmatrix} i & l \\ & \rho \end{Bmatrix}\left(\frac{\partial g_{\kappa\rho}}{\partial x_m}\right)\gamma_{\kappa l} + \sum_{\rho l}\left\{\begin{Bmatrix} i & l \\ & \rho \end{Bmatrix}\begin{Bmatrix} \rho & m \\ & l \end{Bmatrix}\right\}.
\end{aligned}
\tag{436}
$$

217 A factor $\frac{1}{2}$ is missing in the first two terms on the right-hand side (cf. footnote 214).

In summary, p. 14L introduces a set of new invariant-theoretical quantities taken from the mathematical literature that looked promising for constructing gravitational field equations. In his first exploration of these new possibilities, however, Einstein had also hit upon a serious obstacle, viz. the problem of unwanted second-order derivative terms in addition to a core-operator term.

14R–16R 5.3.2 *Extracting a Two-Index Object from the Curvature Scalar (14R–16R)*

On the following five pages, Einstein computed the curvature scalar from the Riemann tensor. He simplified the calculation by setting the determinant of the metric equal to unity. As becomes clear toward the end (on p. 16L), the aim of the calculation was to produce an object which transforms as a scalar under unimodular transformations and which is the contraction of the metric and a new two-index object. Einstein presumably expected that this object would transform as a second-rank tensor under unimodular transformations. It would thus be another candidate for the left-hand side of the field equations. Unfortunately, the object, like the Ricci tensor, contains unwanted second-order derivative terms.

At the top of p. 14R, Einstein fully contracted the Riemann tensor to form the curvature scalar:

$$\varphi = \gamma_{im}\gamma_{\kappa l}(i\kappa, lm)\,. \tag{437}$$

Inserting expression (427) for the fully covariant Riemann tensor on p. 14L into this equation, one arrives at the expression

$$\varphi = \sum_{im\kappa l}\gamma_{im}\gamma_{\kappa l}\left(\frac{\partial^2 g_{im}}{\partial x_\kappa \partial x_l} - \frac{\partial^2 g_{il}}{\partial x_\kappa \partial x_m}\right) + \sum_{\rho\sigma im\kappa l}\gamma_{\rho\sigma}\gamma_{im}\gamma_{\kappa l}\left(\begin{bmatrix} i\ m \\ \sigma \end{bmatrix}\begin{bmatrix} \kappa\ l \\ \rho \end{bmatrix} - \begin{bmatrix} i\ l \\ \sigma \end{bmatrix}\begin{bmatrix} \kappa\ m \\ \rho \end{bmatrix}\right) \tag{438}$$

16R

14R 15L

15R (incorrect)
16L (correct)

15R (top)

given at the top of p. 14R. The lines and arrows underneath this expression provide a flow-chart for the calculation on the next five pages. They are meant to show at a glance exactly which terms in the expression for φ are dealt with on which pages.

14R Einstein first dealt with the second summation in equation (438). On p 14R he expanded and simplified the first of the two terms in this summation,

$$\gamma_{\rho\sigma}\gamma_{im}\gamma_{\kappa l}\begin{bmatrix} i\ m \\ \sigma \end{bmatrix}\begin{bmatrix} \kappa\ l \\ \rho \end{bmatrix}\,. \tag{439}$$

This is equal to 1/4 times the sum of the underlined terms on p. 14R

$$\gamma_{\rho\sigma}\frac{\partial \lg G}{\partial x_\sigma}\frac{\partial \lg G}{\partial x_\rho}+3\frac{\partial \lg G}{\partial x_\sigma}\frac{\partial \gamma_{\sigma\alpha}}{\partial x_\alpha}+\frac{\partial \lg G}{\partial x_\sigma}\frac{\partial \gamma_{\rho\alpha}}{\partial x_\alpha}+4g_{\kappa\rho}\frac{\partial \gamma_{\kappa\alpha}}{\partial x_\alpha}\frac{\partial \gamma_{\rho\beta}}{\partial x_\beta}\,. \tag{440}$$

On p. 15L he expanded and simplified the second term, 15L

$$\gamma_{\rho\sigma}\gamma_{im}\gamma_{\kappa l}\begin{bmatrix} i\ l \\ \sigma \end{bmatrix}\begin{bmatrix} \kappa\ m \\ \rho \end{bmatrix}, \tag{441}$$

the result being 1/4 times the sum of the underlined terms on p. 15L

$$\begin{aligned}\gamma_{\kappa l}\frac{\partial \gamma_{m\rho}}{\partial x_l}\frac{\partial g_{m\rho}}{\partial x_\kappa}+\gamma_{im}\frac{\partial \gamma_{\kappa\rho}}{\partial x_i}\frac{\partial g_{\kappa\rho}}{\partial x_m}+\gamma_{\rho\sigma}\frac{\partial \gamma_{\kappa m}}{\partial x_\sigma}\frac{\partial g_{\kappa m}}{\partial x_\rho}\\ +g_{\kappa\rho}\frac{\partial \gamma_{\kappa m}}{\partial x_\sigma}\frac{\partial \gamma_{\rho\sigma}}{\partial x_m}+g_{m\rho}\frac{\partial \gamma_{m\kappa}}{\partial x_\sigma}\frac{\partial \gamma_{\rho\sigma}}{\partial x_\kappa}\,.\end{aligned} \tag{442}$$

Einstein further simplified expressions (440) and (442) by setting $G = 1$ and by grouping together identical terms in expression (442). At the top of p. 15R, he wrote: 15R
"The second sum [in equation (438)] thus reduces, in the case that one is allowed to set $G = 1$, to" ("Die zweite Summe reduziert sich also in dem Falle, dass $G = 1$ gesetzt werden darf, auf"):

$$\frac{1}{4}\cdot\left(4g_{\kappa\rho}\frac{\partial \gamma_{\kappa\alpha}}{\partial x_\alpha}\frac{\partial \gamma_{\rho\beta}}{\partial x_\beta}+3\gamma_{\kappa l}\frac{\partial \gamma_{m\rho}}{\partial x_\kappa}\frac{\partial g_{m\rho}}{\partial x_l}+2g_{\kappa\rho}\frac{\partial \gamma_{\kappa m}}{\partial x_\sigma}\frac{\partial \gamma_{\rho\sigma}}{\partial x_m}\right). \tag{443}$$

With the help of some auxiliary calculations on p. 16R, in which he used that 16R
$g_{im}d\gamma_{im} = d\lg G = 0$ for $G = 1$, Einstein then rewrote the two terms in the first summation in equation (438). Underneath equation (443) on p. 15R, he wrote: "If the 15R
determinant $G = 1$, one has furthermore" ("Wenn Determinante $G = 1$, es ist ferner"):

$$\sum\gamma_{im}\gamma_{\kappa l}\frac{\partial^2 g_{im}}{\partial x_\kappa \partial x_l} = \sum g_{im}\gamma_{\kappa l}\frac{\partial^2 \gamma_{im}}{\partial x_\kappa \partial x_l}, \tag{444}$$

and

$$-\sum\gamma_{im}\gamma_{\kappa l}\frac{\partial^2 g_{il}}{\partial x_\kappa \partial x_m} = -\sum g_{il}\frac{\partial \gamma_{im}}{\partial x_\kappa}\frac{\partial \gamma_{\kappa l}}{\partial x_m}-\sum g_{il}\frac{\partial \gamma_{l\alpha}}{\partial x_\alpha}\frac{\partial \gamma_{i\beta}}{\partial x_\beta}+\sum g_{il}\gamma_{im}\frac{\partial^2 \gamma_{\kappa l}}{\partial x_\kappa \partial x_m}, \tag{445}$$

Einstein erroneously thought that the first term on the right-hand side of equation (445), a term that originally had an extra factor 2, cancelled by the third term of expression (443), which originally did not have the factor $1/4$ in front of it. Moreover, in the last term in equation (445), x_m originally seems to have been x_l. All in all, Einstein arrived at the (erroneous) expression

$$\begin{aligned}&3\sum g_{\kappa\rho}\frac{\partial\gamma_{\kappa\alpha}}{\partial x_\alpha}\frac{\partial\gamma_{\rho\beta}}{\partial x_\beta}+\sum g_{im}\gamma_{\kappa l}\frac{\partial^2\gamma_{im}}{\partial x_\kappa\partial x_l}\\&+\sum g_{il}\gamma_{im}\frac{\partial^2\gamma_{\kappa l}}{\partial x_\kappa\partial x_l}+3\sum\gamma_{\kappa l}\frac{\partial\gamma_{m\rho}}{\partial x_\kappa}\frac{\partial g_{m\rho}}{\partial x_l}.\end{aligned}\tag{446}$$

for the curvature scalar in coordinates such that $G = 1$.[218] Einstein proceeded to rewrite this expression as the contraction of $g_{\kappa\lambda}$ with a new two-index object that would be a candidate for the left-hand side of the field equations. It is immediately clear, however, that these new candidate field equations would suffer from the same problem that Einstein had already encountered on p. 14L with the Ricci tensor. The second term in expression (446) gives rise to a core-operator term in the field equations, but the third term produces additional second-order derivative terms. Directly underneath the troublesome third term in expression (446), Einstein wrote

$$\sum\frac{\partial^2\gamma_{\kappa l}}{\partial x_\kappa\partial x_l} = 0 \quad ? \tag{447}$$

He may have considered imposing an additional constraint to make sure that the offending term vanishes. In passing we note that equation (447) is a weakening of the Hertz restriction

$$\frac{\partial\gamma_{kl}}{\partial x_l} = 0\,, \tag{448}$$

which Einstein had already used on p. 10L–11L, i.e., before he began his exploration of the Riemann tensor.

As we have seen, the calculations on p. 15R contain several errors. Einstein cor-
16L rected some of these, but then made a fresh start at the top of the next page (p. 16L). Using again that $g_{m\rho}d\gamma_{m\rho} = d\lg G = 0$ if $G = 1$, he rewrote the second term of expression (443) as

$$\frac{3}{4}\gamma_{\kappa l}\frac{\partial\gamma_{m\rho}}{\partial x_\kappa}\frac{\partial g_{m\rho}}{\partial x_l} = -\frac{3}{4}g_{m\rho}\gamma_{\kappa l}\frac{\partial^2\gamma_{m\rho}}{\partial x_\kappa\partial x_l}\,. \tag{449}$$

Substituting this result into expression (443) and adding the various terms in expressions (443)–(445), he arrived at the (correct) expression

$$\frac{1}{4}g_{\kappa\rho}\gamma_{\alpha\beta}\frac{\partial^2\gamma_{\kappa\rho}}{\partial x_\alpha\partial x_\beta}-\frac{1}{2}g_{\kappa\rho}\frac{\partial\gamma_{\kappa\alpha}}{\partial x_\beta}\frac{\partial\gamma_{\rho\beta}}{\partial x_\alpha}+\frac{\partial^2\gamma_{\alpha\beta}}{\partial x_\alpha\partial x_\beta} \tag{450}$$

for the Riemann curvature scalar in coordinates such that $G = 1$. In this expression one only needs to rewrite the last term[219] in order to pull out a common factor $g_{\kappa\rho}$ and to extract the two-index object

218 The corrected version is given below as expression (450).

$$T_{i\kappa} = \frac{1}{4}\gamma_{\alpha\beta}\frac{\partial^2\gamma_{i\kappa}}{\partial x_\alpha \partial x_\beta} - \frac{1}{2}\frac{\partial\gamma_{i\alpha}}{\partial x_\beta}\frac{\partial\gamma_{\kappa\beta}}{\partial x_\alpha} + \frac{1}{2}\gamma_{\kappa\beta}\frac{\partial^2\gamma_{i\alpha}}{\partial x_\alpha \partial x_\beta} + \frac{1}{2}\gamma_{i\alpha}\frac{\partial^2\gamma_{\kappa\beta}}{\partial x_\alpha \partial x_\beta}. \tag{451}$$

Einstein apparently assumed that $T_{i\kappa}$ is a contravariant tensor under unimodular transformations since it produces a scalar under unimodular transformation when contracted with $g_{i\kappa}$.[220]

Unfortunately, $T_{i\kappa}$ still contained unwanted second-order derivative terms in addition to a core-operator term. Immediately below the expression for T_{ik}, Einstein wrote down the identity

$$\frac{\partial}{\partial x_\alpha}\frac{\partial}{\partial x_\beta}(\gamma_{i\alpha}\gamma_{\kappa\beta}) = \gamma_{i\alpha}\frac{\partial^2\gamma_{\kappa\beta}}{\partial x_\alpha \partial x_\beta} + \gamma_{\kappa\beta}\frac{\partial^2\gamma_{i\alpha}}{\partial x_\alpha \partial x_\beta} + \frac{\partial\gamma_{i\alpha}}{\partial x_\alpha}\frac{\partial\gamma_{\kappa\beta}}{\partial x_\beta} + \frac{\partial\gamma_{i\alpha}}{\partial x_\beta}\frac{\partial\gamma_{\kappa\beta}}{\partial x_\alpha}, \tag{452}$$

probably in an attempt to eliminate the unwanted second-order derivative terms from T_{ik}. He did not pursue this attempt any further.

The upshot then is that a promising alternative way of generating a candidate for the left-hand side of the field equations from the Riemann tensor eventually led to the same problem that Einstein had encountered with the Ricci tensor on p. 14L. In deriving these new candidate field equations he had already imposed the condition $G = 1$, but additional constraints would be needed to eliminate unwanted second-order derivative terms. He may have considered one such constraint, equation (447), a weakening of the Hertz restriction.

5.3.3 *Comparing $T_{i\kappa}$ and the Ricci Tensor (17L–18R)* 17L–18R

On pp. 17L–18R, Einstein investigated the relation between the two two-index objects he had constructed out of the Riemann tensor on the preceding pages, the Ricci tensor and the object T_{ik}, which promised to be a tensor under unimodular transformations. On p. 17R he compared the contravariant forms of the two expressions, which meant that he had to raise the covariant indices of the Ricci tensor. On pp. 17R–18R he com-

---

219 Immediately above it, the term was restored to the form $g_{\kappa\rho}\gamma_{\kappa\alpha}\frac{\partial^2\gamma_{\rho\beta}}{\partial x_\alpha \partial x_\beta}$, in which it was originally written in equation (445). Einstein symmetrized this expression in κ and ρ: $g_{\kappa\rho}\left(\frac{1}{2}\gamma_{\kappa\alpha}\frac{\partial^2\gamma_{\rho\beta}}{\partial x_\alpha \partial x_\beta} + \frac{1}{2}\gamma_{\rho\beta}\frac{\partial^2\gamma_{\alpha\kappa}}{\partial x_\alpha \partial x_\beta}\right)$, so that the object left after pulling out the common factor $g_{\kappa\rho}$ will be symmetric in κ and ρ. A simpler way to rewrite the last term of expression (450) as the contraction of $g_{\kappa\rho}$ with an expression that is symmetric in κ and ρ is to multiply it by $\frac{1}{4}g_{\kappa\rho}\gamma_{\kappa\rho}$ (cf. the top of p. 17R, the last term in expression (458)).

220 To prove that $T_{i\kappa}$ is a contravariant tensor (under unimodular transformations), one would, of course, have to show that its contraction with an *arbitrary* covariant tensor produces a scalar (under unimodular transformations). Notice, however, that Einstein did make sure that $T_{i\kappa}$ is symmetric (see the preceding note). Any anti-symmetric contribution to $T_{i\kappa}$ would vanish upon contraction with the metric, no matter whether $T_{i\kappa}$ is a tensor or not.

pared the covariant forms, which meant that he had to lower the contravariant indices of T_{ik}. The resulting expressions quickly became so cumbersome that Einstein abandoned both calculations.

17L On p. 17L, Einstein formed the contravariant Ricci tensor and started to expand the first-order derivative terms. The calculations are very similar to the ones on the preceding pages. Directly underneath the heading "Point tensor of gravitation" ("Punkttensor der Gravitation"), he wrote down the symbol for the fully covariant Riemann tensor introduced on p. 14L, and noted that this is a "plane tensor" ("Ebenentensor"),

$$(i\kappa, lm) = \text{Ebenentensor vierter Mannigfaltigkeit}\,. \tag{453}$$

He then formed the contravariant Ricci tensor, and noted that this is a "point tensor" ("Punkttensor"),

$$\sum_{i\kappa lm} \gamma_{\kappa l}\gamma_{ip}\gamma_{mq}(i\kappa, lm) = \text{Punkttensor}\,. \tag{454}$$

The prefixes "Punkt-" and "Ebene-" were introduced on p. 13L to distinguish between contravariant and covariant indices, albeit only in the context of linear transformations for which $G = 1$.[221] On p. 13L only the terms "Punktvektor" and "Ebenenvektor" occur explicitly. The terms "Punkttensor" and "Ebenentensor" occur here on p. 17L for the first time, although they were implied on p. 13L by the convention of using dots and dashes to denote contravariant ("Punkt") and covariant ("Ebene") indices, respectively, for vectors as well as tensors.[222] Einstein did not use this terminology in any of his published writings.

Einstein went on to expand and simplify the "point tensor" (454). He first wrote down an incorrect expression for the Riemann tensor,

$$(i\kappa, lm) = \frac{\partial^2 g_{im}}{\partial x_\kappa \partial x_l} - \frac{\partial^2 g_{il}}{\partial x_\kappa \partial x_m} + \sum_{\rho\sigma}\gamma_{\rho\sigma}\left\{\begin{bmatrix} i\ m \\ \sigma \end{bmatrix}\begin{bmatrix} \kappa\ l \\ \rho \end{bmatrix} - \begin{bmatrix} i\ l \\ \sigma \end{bmatrix}\begin{bmatrix} \kappa\ m \\ \rho \end{bmatrix}\right\}. \tag{455}$$

The four second-order derivative terms in equation (427) for the Riemann tensor cannot be grouped together in this way. They can, when the Riemann tensor is fully contracted to form the curvature scalar, as was done on p. 14R. This suggests that Einstein read off equation (455) from expression (438) for the curvature scalar which formed the starting point of his calculations on pp. 14R–16R. The error does not affect the rest of the calculation, since Einstein did not get beyond rewriting the first-order derivative terms in equation (455).

The further manipulation of the contravariant Ricci tensor on this page is very similar to that of the covariant Ricci tensor and of the curvature scalar on the preceding pages. He first considered the first term in the second summation in equation (455) contracted with $\gamma_{kl}\gamma_{ip}\gamma_{mq}$ (see equation (454)), which he expanded to

---

221 See the discussion of p. 13L in sec. 5.6.1.

222 On pp. 28L and 29L, the terms "Punkttensor" and "kontravarianter Tensor" are used interchangeably.

$$\frac{1}{4}\gamma_{\kappa l}\gamma_{ip}\gamma_{mq}\gamma_{\rho\sigma}\left(\frac{\partial g_{i\sigma}}{\partial x_m}+\frac{\partial g_{m\sigma}}{\partial x_i}-\frac{\partial g_{im}}{\partial x_\sigma}\right)\left(\frac{\partial g_{\kappa\rho}}{\partial x_l}+\frac{\partial g_{l\rho}}{\partial x_\kappa}-\frac{\partial g_{\kappa l}}{\partial x_\rho}\right). \tag{456}$$

He then pulled one pair of γ factors ($\gamma_{ip}\gamma_{mq}$) inside the first set of parentheses and the other pair into the second. In the resulting terms he used the relation $g_{ij}\gamma_{ik} = \delta_{jk}$ to move the differentiation over from the g 's to the γ 's. He also set the determinant of metric equal to unity again, writing "$\lg G$ set to zero" ("$\lg G = 0$ gesetzt "), in which case the last term in the second set of parentheses in expression (456) vanishes upon contraction with $\gamma_{\kappa l}$. The next step was to move the factor $\gamma_{\rho\sigma}$ from the second set of parentheses to the first. The expression could then be simplified further through contractions of the form $g_{ij}\gamma_{ik} = \delta_{jk}$. In this way, Einstein rewrote expression (456) as

$$\frac{1}{4}\left(\gamma_{mq}\frac{\partial\gamma_{\rho p}}{\partial x_m}+\gamma_{ip}\frac{\partial\gamma_{\rho q}}{\partial x_i}-\gamma_{\rho\sigma}\frac{\partial\gamma_{pq}}{\partial x_\sigma}\right)\left(g_{\kappa\rho}\frac{\partial\gamma_{\kappa l}}{\partial x_l}+g_{l\rho}\frac{\partial\gamma_{\kappa l}}{\partial x_\kappa}\right). \tag{457}$$

Einstein drew a horizontal line, and turned to the second term in the second summation in equation (455), proceeding in much the same way as he had with the first term. However, after two lines he broke off the calculation with the comment, written at the bottom of the page, that it was "too involved" ("zu umständlich").

On pp. 17R–18R, Einstein took a slightly different approach. Under the heading "Plane tensor constructed in two different ways" ("Auf zwei Arten Ebenentensor gebildet"), he tried to bring the covariant versions of the Ricci tensor and the object $T_{i\kappa}$ of p. 16L (with a minor modification) into a form in which they could be compared to one another. The calculation extends over the following three pages (17R–18R). Again it breaks off before yielding a definite result. 17R–18R

Einstein began by considering what he called the "first way" ("1. Art") of forming a covariant tensor. This refers to the object $T_{i\kappa}$ extracted from the curvature scalar on pp. 14R–16R. He presumably went back to p. 16L, to expression (450) for the curvature scalar under the condition that $G = 1$. Rewriting the last term in this expression in a slightly different way than was done on p. 16L, one arrives at[223]

$$\frac{1}{4}g_{\kappa\rho}\gamma_{\alpha\beta}\frac{\partial^2\gamma_{\kappa\rho}}{\partial x_\alpha\partial x_\beta}-\frac{1}{2}g_{\kappa\rho}\frac{\partial\gamma_{\kappa\alpha}}{\partial x_\beta}\frac{\partial\gamma_{\rho\beta}}{\partial x_\alpha}+\frac{1}{4}g_{\kappa\rho}\gamma_{\kappa\rho}\frac{\partial^2\gamma_{\alpha\beta}}{\partial x_\alpha\partial x_\beta}, \tag{458}$$

which is $1/4$ times the contraction of $g_{\kappa\rho}$ with the expression in square brackets at the top of p. 17R, 17R

$$\gamma_{\alpha\beta}\frac{\partial^2\gamma_{\kappa\rho}}{\partial x_\alpha\partial x_\beta}-2\frac{\partial\gamma_{\kappa\alpha}}{\partial x_\beta}\frac{\partial\gamma_{\rho\beta}}{\partial x_\alpha}+\gamma_{\kappa\rho}\frac{\partial^2\gamma_{\alpha\beta}}{\partial x_\alpha\partial x_\beta}. \tag{459}$$

223 See footnote 219 above.

With minor modifications, this is the object T_{ik} defined on p. 16L. Apart from an overall factor of 4 and the labeling of its free indices, it only differs from the object constructed on p. 16L in its last term.

The expression at the top of p. 17R gives the covariant version of this object,

$$\sum g_{\kappa\sigma} g_{\rho\tau}\left[\gamma_{\alpha\beta}\frac{\partial^2\gamma_{\kappa\rho}}{\partial x_\alpha \partial x_\beta} - 2\frac{\partial\gamma_{\kappa\alpha}}{\partial x_\beta}\frac{\partial\gamma_{\rho\beta}}{\partial x_\alpha} + \gamma_{\kappa\rho}\frac{\partial^2\gamma_{\alpha\beta}}{\partial x_\alpha \partial x_\beta}\right]. \tag{460}$$

On the next two lines, Einstein wrote down equivalent expressions for all three terms in expression (460), using the relation $g_{ij}\gamma_{ik} = \delta_{jk}$ and relations that can be derived from it through differentiation.[224]

Einstein now turned to the "second way" ("2. Art") of forming a covariant tensor, which is contracting the Riemann tensor to form the Ricci tensor. Einstein had already started this calculation on p. 14L, but he made a fresh start on this page.

Using the expressions for the terms with products of Christoffel symbols on the facing page (p. 17L, cf. expression (456)), he wrote the Ricci tensor as

$$\begin{aligned}
&\frac{1}{2}\sum\gamma_{\kappa l}\left(\frac{\partial^2 g_{im}}{\partial x_\kappa \partial x_l} + \frac{\partial^2 g_{\kappa l}}{\partial x_i \partial x_m} - \frac{\partial^2 g_{il}}{\partial x_\kappa \partial x_m} - \frac{\partial^2 g_{\kappa m}}{\partial x_i \partial x_l}\right) \\
&+\frac{1}{4}\sum_{kl\rho\sigma}\gamma_{\kappa l}\gamma_{\rho\sigma}\left(\frac{\partial g_{i\sigma}}{\partial x_m} + \frac{\partial g_{m\sigma}}{\partial x_i} - \frac{\partial g_{im}}{\partial x_\sigma}\right)\left(\frac{\partial g_{\kappa\rho}}{\partial x_l} + \frac{\partial g_{l\rho}}{\partial x_\kappa} - \frac{\partial g_{\kappa l}}{\partial x_\rho}\right) \\
&-\frac{1}{4}\sum_{kl\rho\sigma}\gamma_{\kappa l}\gamma_{\rho\sigma}\left(\frac{\partial g_{i\sigma}}{\partial x_l} + \frac{\partial g_{l\sigma}}{\partial x_i} - \frac{\partial g_{il}}{\partial x_\sigma}\right)\left(\frac{\partial g_{\kappa\rho}}{\partial x_m} + \frac{\partial g_{m\rho}}{\partial x_\kappa} - \frac{\partial g_{\kappa m}}{\partial x_\rho}\right).
\end{aligned} \tag{461}$$

Further simplification of this expression was facilitated by his previous investigation of the Ricci tensor on p. 14L. Without further calculation Einstein noted that the first two terms in the second set of parentheses on the second line are identical and "can be combined" ("vereinigt sich"), and that the third term "vanishes" ("fällt weg"), which indicates that he once again imposed the condition that the determinant G is a constant. These results had explicitly been derived on p. 14L (cf. equation (432)). Einstein could therefore immediately rewrite the second line as

$$\frac{1}{2}\sum_{kl\rho\sigma}\gamma_{\kappa l}\gamma_{\rho\sigma}\left(\frac{\partial g_{i\sigma}}{\partial x_m} + \frac{\partial g_{m\sigma}}{\partial x_i} - \frac{\partial g_{im}}{\partial x_\sigma}\right)\frac{\partial g_{\kappa\rho}}{\partial x_l} = -\frac{1}{2}\sum\frac{\partial\gamma_{\rho\sigma}}{\partial x_\rho}\left(\frac{\partial g_{i\sigma}}{\partial x_m} + \frac{\partial g_{m\sigma}}{\partial x_i} - \frac{\partial g_{im}}{\partial x_\sigma}\right). \tag{462}$$

His increased facility in handling these expressions is also manifest in his treatment of the third line of expression (461). He started to rewrite the contraction of

224 It takes a short calculation very similar to the one rehearsed in the last four lines of p. 16R to show that the first term can be rewritten as $-\gamma_{\alpha\beta}\frac{\partial^2 g_{\tau\sigma}}{\partial x_\alpha \partial x_\beta} + 2\gamma_{\alpha\beta}\gamma_{\kappa\rho}\frac{\partial g_{\rho\tau}}{\partial x_\beta}\frac{\partial g_{\kappa\sigma}}{\partial x_\alpha}$. Einstein might have done this calculation on a separate piece of paper.

$\gamma_{\kappa l}\gamma_{\rho\sigma}$ with the term in the second set of parentheses, but then he noticed that he could simplify the expression more easily by a symmetry argument. He marked four terms in the expression by wiggly lines,

$$-\frac{1}{4}\sum_{kl\rho\sigma}\gamma_{\kappa l}\gamma_{\rho\sigma}\left(\frac{\partial g_{i\sigma}}{\partial x_l}+\frac{\partial g_{l\sigma}}{\partial x_i}-\frac{\partial g_{il}}{\partial x_\sigma}\right)\left(\frac{\partial g_{\kappa\rho}}{\partial x_m}+\frac{\partial g_{m\rho}}{\partial x_\kappa}-\frac{\partial g_{\kappa m}}{\partial x_\rho}\right). \tag{463}$$

The first two underlined terms are antisymmetric in l and σ, the last two in κ and ρ. Upon contraction with $\gamma_{\kappa l}\gamma_{\rho\sigma}$, the two remaining terms become symmetric in κ and ρ and l and σ, respectively. So, the product of the underlined terms in one set of parentheses with the non-underlined term in the other produces the contraction of an expression anti-symmetric with an expression symmetric in the same pair of indices. These products thus vanish, and expression (463) reduces to

$$-\frac{1}{4}\sum\gamma_{\kappa l}\gamma_{\rho\sigma}\frac{\partial g_{l\sigma}}{\partial x_i}\frac{\partial g_{\kappa\rho}}{\partial x_m}-\frac{1}{4}\sum\gamma_{\kappa l}\gamma_{\rho\sigma}\left(\frac{\partial g_{i\sigma}}{\partial x_l}-\frac{\partial g_{il}}{\partial x_\sigma}\right)\left(\frac{\partial g_{m\rho}}{\partial x_\kappa}-\frac{\partial g_{\kappa m}}{\partial x_\rho}\right), \tag{464}$$

the expression written at the bottom of the page. Einstein rewrote the first term as

$$\frac{1}{4}\frac{\partial\gamma_{\rho\sigma}}{\partial x_i}\frac{\partial g_{\sigma\rho}}{\partial x_m}. \tag{465}$$

At the top of the next page (p. 18L), he inserted the results found on p. 17R (see equation (462) and expression (464)–(465)) into expression (461) for the Ricci tensor and arrived at[225] 18L

$$\begin{aligned}&\frac{1}{2}\gamma_{\kappa l}\left(\frac{\partial^2 g_{im}}{\partial x_\kappa\partial x_l}+\frac{\partial^2 g_{\kappa l}}{\partial x_i\partial x_m}-\frac{\partial^2 g_{il}}{\partial x_\kappa\partial x_m}-\frac{\partial^2 g_{\kappa m}}{\partial x_i\partial x_l}\right)\\ &-\frac{1}{2}\frac{\partial\gamma_{\sigma\rho}}{\partial x_\rho}\left(\frac{\partial g_{i\sigma}}{\partial x_m}+\frac{\partial g_{m\sigma}}{\partial x_i}-\frac{\partial g_{im}}{\partial x_\sigma}\right)+\frac{1}{4}\frac{\partial\gamma_{\rho\sigma}}{\partial x_i}\frac{\partial g_{\rho\sigma}}{\partial x_m}\\ &-\frac{1}{4}\gamma_{\kappa l}\gamma_{\rho\sigma}\left(\frac{\partial g_{i\sigma}}{\partial x_l}-\frac{\partial g_{l\sigma}}{\partial x_i}\right)\left(\frac{\partial g_{m\rho}}{\partial x_\kappa}-\frac{\partial g_{\kappa m}}{\partial x_\rho}\right).\end{aligned} \tag{466}$$

On pp. 18L–R, Einstein tried to simplify this expression still further. Thus, in the first two lines of "auxiliary calculations" ("Nebenrechnungen") on p. 18R, he showed that the condition that the determinant of the metric is constant allows one to rewrite the second term in the first line of expression (466) as 18L–R

225 The expression $\frac{\partial g_{i\sigma}}{\partial x_l}-\frac{\partial g_{l\sigma}}{\partial x_i}$ on the third line should be $\frac{\partial g_{i\sigma}}{\partial x_l}-\frac{\partial g_{il}}{\partial x_\sigma}$, i.e., the expression should be anti-symmetric in l and σ, not in i and l.

$$-\frac{1}{2}\frac{\partial\gamma_{\kappa\rho}}{\partial x_i}\frac{\partial g_{\kappa\rho}}{\partial x_m}. \tag{467}$$

This expression (without the minus sign) is given on the last line of p. 18L. By the same token, the last term on the second line is equivalent to

$$-\frac{1}{4}\gamma_{\rho\sigma}\frac{\partial^2 g_{\rho\sigma}}{\partial x_i \partial x_m}, \tag{468}$$

the expression written underneath this term, except for the fact that Einstein mistakenly wrote $\gamma_{\rho\sigma}$ instead of $g_{\rho\sigma}$.

Einstein performed another auxiliary calculation on p. 18R to expand the third line of expression (466). Multiplying the end result of this auxiliary calculation by minus $1/4$, one arrives at the expression

$$-\frac{1}{4}\left(\frac{\partial\gamma_{l\sigma}}{\partial x_m}\frac{\partial g_{i\sigma}}{\partial x_l}+\frac{\partial\gamma_{\rho\sigma}}{\partial x_i}\frac{\partial g_{m\rho}}{\partial x_\sigma}-\frac{\partial\gamma_{l\sigma}}{\partial x_m}\frac{\partial g_{l\sigma}}{\partial x_i}\right)-\frac{1}{4}\gamma_{\kappa l}\gamma_{\rho\sigma}\frac{\partial g_{i\sigma}}{\partial x_l}\frac{\partial g_{m\rho}}{\partial x_\kappa}, \tag{469}$$

written directly underneath this term in expression (466) on p. 18L. At this point Einstein gave up and crossed out his calculations on pp. 18L and 18R in their entirety. The relation between the Ricci tensor and the two-index object T_{ik}, the two candidates for the left-hand side of the field equations that Einstein had considered in his first explorations of the Riemann tensor, still remained unclear.

19L–21R

5.4 *Exploring the Ricci tensor in Harmonic Coordinates (19L–21R)*

The mathematical strategy which Einstein had been following since the introduction of the Riemann tensor on p. 14L had still not yielded a viable candidate for the left-hand side of the gravitational field equations. The two expressions considered on the preceding pages (14L–18R), the Ricci tensor and the object T_{ik}, both contain, in addition to the desired core-operator term, unwanted second-order derivative terms. On p. 18R Einstein used the restriction to unimodular coordinates, which allowed him to set $G = 1$, to eliminate one of these unwanted terms (cf. eqs. (466)-(467)). On p. 19L Einstein eliminated all unwanted second-order derivative terms through an appropriate choice of coordinates. In this way he reduced the Ricci tensor to the sum of a core-operator term and terms quadratic in the first-order derivatives of the metric. Today these coordinates are called "harmonic coordinates" and the corresponding condition is called the "harmonic coordinate condition." Let us reiterate that Einstein understood coordinate conditions not in the modern sense of selecting at least one member from each possible equivalence class of metrics,[226] but as restrictions on the covariance group of the field equations. That is why we adopted the phrase coordinate restric-

226 Two metrics are in the same equivalence class if and only if a coordinate transformation exists that maps one onto the other.

tion.[227] With the help of the harmonic restriction, Einstein was finally able to extract from the Riemann tensor field equations that satisfy the correspondence principle.

On p. 19R, Einstein began to examine these new field equations and the harmonic restriction in a weak-field approximation. He made sure the new field equations were compatible with energy-momentum conservation by checking that the gravitational force on a cloud of dust particles can be written as the divergence of a quantity representing gravitational stress-energy density. To ensure compatibility between the field equations and energy-momentum conservation, Einstein imposed an extra condition on the metric tensor. The combination of this extra condition—a linearized version of the Hertz restriction—and the harmonic restriction leads to the unacceptable result that the trace of the weak-field metric has to be a constant.

On p. 20L, Einstein modified the weak-field equations to avoid this implication. By adding a term with the trace of the weak-field metric, he ensured that the requirements of the correspondence principle and the conservation principle are both satisfied by imposing the harmonic restriction. From a purely mathematical point of view, the left-hand side of these modified weak-field equations is, in fact, the linearized Einstein tensor. At the bottom of p. 20L and on p. 21L, Einstein used the new weak-field equations to rewrite the gravitational force as a divergence, thus convincing himself that the addition of the trace term to the field equations does not destroy their compatibility with energy-momentum conservation. There was, however, a different problem. The modified weak-field equations of p. 20L are incompatible with Einstein's presupposition concerning static gravitational fields.[228]

It is no coincidence therefore that, on p. 20R, Einstein reexamined the dynamics of point particles moving in a metric field. The purpose of this calculation becomes clear on p. 21R. Einstein checked whether his presupposition concerning static gravitational fields was actually justified. On p. 21R, using elements of his discussion of the dynamics of point particles on p. 20R, he developed what appeared to be a very strong argument in support of his views on the static field. This (fallacious) argument led him to give up the harmonic restriction and the field equations constructed with its help.

5.4.1 *Extracting Field Equations from the Ricci Tensor Using Harmonic Coordinates (19L)* 19L

On p. 19L, as is announced in the first line: "Renewed calculation of the plane tensor[229]" ("Nochmalige Berechnung des Ebenentensors"), Einstein re-calculated the quantity given by the expression

227 For further discussion, see sec. 4.1.

228 This presupposition is also incompatible with the harmonic coordinate restriction, but it is unclear whether Einstein realized that at this point.

229 See sec. 5.3.3, for a discussion of Einstein's usage of the term "Ebenentensor" for covariant tensors.

$$\left.\begin{array}{l}\frac{1}{2}\left(\frac{\partial^2 g_{im}}{\partial x_\kappa \partial x_l}+\frac{\partial^2 g_{\kappa l}}{\partial x_i \partial x_m}-\frac{\partial^2 g_{il}}{\partial x_\kappa \partial x_m}-\frac{\partial^2 g_{\kappa m}}{\partial x_i \partial x_l}\right) \\ -\frac{1}{4}\sum_{\kappa l \rho\sigma}\gamma_{\rho\sigma}\left(\frac{\partial g_{i\rho}}{\partial x_l}+\frac{\partial g_{l\rho}}{\partial x_i}-\frac{\partial g_{il}}{\partial x_{\sigma\rho}}\right)\left(\frac{\partial g_{\kappa\sigma}}{\partial x_m}+\frac{\partial g_{m\sigma}}{\partial x_\kappa}-\frac{\partial g_{\kappa m}}{\partial x_\sigma}\right)\end{array}\right|\gamma_{\kappa l}. \tag{470}$$

This expression gives an incomplete version of the fully covariant Riemann tensor (see, e.g., equation (427)) contracted with the contravariant metric γ_{kl}. In other words, it is an incomplete version of the Ricci tensor, which Einstein had already investigated on p. 14L and on pp. 17R–18R.[230] What is missing is another term with products of Christoffel symbols. This was not just an oversight on Einstein's part. The missing term vanishes if the condition

$$\gamma_{\kappa l}\begin{bmatrix}\kappa & l \\ & i\end{bmatrix} = \gamma_{\kappa l}\left(2\frac{\partial g_{il}}{\partial x_\kappa}-\frac{\partial g_{\kappa l}}{\partial x_i}\right) = 0 \tag{471}$$

is imposed, which Einstein actually wrote down two lines further down. This is the condition that we today call the harmonic coordinate condition. The point of introducing this condition and the purpose of the whole calculation becomes clear in the line immediately following expression (470). On this line, Einstein wrote down the first of the four second-order derivative terms in expression (470),

$$\frac{1}{2}\gamma_{kl}\frac{\partial^2 g_{im}}{\partial x_\kappa \partial x_l}, \tag{472}$$

which has the form of a core operator, underlined it, and noted that it would "remain" ("bleibt stehen"). Exactly how this comes about is recapitulated in the calculation that follows.

Einstein indicated that he wanted to take the derivative of the condition with free index i in equation (471) with respect to x_m and then do the same with the indices i and m interchanged. Adding the two equations, Einstein got[231]

---

230 Expression (470) differs from the expressions for the Riemann tensor and the Ricci tensor given earlier (cf. equations (427), (455), and (461)) in that the indices ρ and σ in the terms with products of Christoffel symbols (such as the last line of expression (470)) have been switched, in accordance with the labeling of these indices in the *Entwurf* paper (Einstein and Grossmann 1913, 35). This could be an indication that some time elapsed between these earlier calculations and the ones starting on p. 19L.

231 In the very last terms in equations (473) and (474), the derivative should be with respect to x_m rather than x_i.

$$2\gamma_{\kappa l}\left(\frac{\partial^2 g_{il}}{\partial x_\kappa \partial x_m} + \frac{\partial^2 g_{\kappa m}}{\partial x_i \partial x_l} - \frac{\partial^2 g_{\kappa l}}{\partial x_i \partial x_m}\right)$$
$$+ \frac{\partial \gamma_{\kappa l}}{\partial x_m}\left(2\frac{\partial g_{il}}{\partial x_\kappa} - \frac{\partial g_{\kappa l}}{\partial x_i}\right) + \frac{\partial \gamma_{\kappa l}}{\partial x_i}\left(2\frac{\partial g_{m\kappa}}{\partial x_l} - \frac{\partial g_{\kappa l}}{\partial x_i}\right) = 0. \tag{473}$$

This relation allowed Einstein to replace the three bothersome second-order derivative terms in expression (470) by an expression

$$-\frac{1}{2}\gamma_{\kappa l}(\quad) = \frac{1}{4}\left(\frac{\partial \gamma_{\kappa l}}{\partial x_m}\left(2\frac{\partial g_{il}}{\partial x_\kappa} - \frac{\partial g_{\kappa l}}{\partial x_i}\right) + \frac{\partial \gamma_{\kappa l}}{\partial x_i}\left(2\frac{\partial g_{m\kappa}}{\partial x_l} - \frac{\partial g_{\kappa l}}{\partial x_i}\right)\right) \tag{474}$$

containing only first-order derivatives.

Einstein now turned his attention to the "second term" ("zweites Glied"), i.e., the second line of expression (470). He invoked the same symmetry argument as on p. 17R (cf. the discussion following expression (463) above). He marked the symmetric and anti-symmetric terms in the index pairs (l, ρ) and (κ, σ) by straight and wiggly lines, respectively, and immediately wrote down the only two non-vanishing contributions, one coming from the symmetric parts,

$$-\frac{1}{4}\gamma_{\rho\sigma}\frac{\partial g_{l\rho}}{\partial x_i}\frac{\partial g_{\kappa\sigma}}{\partial x_m}\gamma_{\kappa l}, \tag{475}$$

and one coming from the anti-symmetric parts,

$$-\frac{1}{4}\gamma_{\rho\sigma}\left(\frac{\partial g_{i\rho}}{\partial x_l} - \frac{\partial g_{il}}{\partial x_\rho}\right)\left(\frac{\partial g_{m\sigma}}{\partial x_\kappa} - \frac{\partial g_{\kappa m}}{\partial x_\sigma}\right)\gamma_{\kappa l}$$
$$= -\frac{1}{2}\gamma_{\rho\sigma}\gamma_{\kappa l}\frac{\partial g_{i\rho}}{\partial x_l}\frac{\partial g_{m\sigma}}{\partial x_\kappa} + \frac{1}{2}\gamma_{\rho\sigma}\gamma_{\kappa l}\frac{\partial g_{il}}{\partial x_\rho}\frac{\partial g_{m\sigma}}{\partial x_\kappa}. \tag{476}$$

He now underlined the terms that form the Ricci tensor: the core-operator term in expression (472), the right-hand side of equation (474), expression (475), and the right-hand side of equation (476). A short auxiliary calculation showed that expression (475) and the last term in equation (474) cancel each other.[232] He added the remaining underlined terms and concluded that "the covariant tensor [the Ricci tensor], multiplied by 2, thus takes the form" ("Der mit 2 multiplizierte Ebenentensor erhält also die Form")

232 The two terms only cancel if the final index in equation (474) is corrected (see the preceding note). Einstein probably realized the index was wrong at this point, although he did not correct it.

$$\begin{aligned}&\gamma_{\kappa l}\frac{\partial^2 g_{im}}{\partial x_\kappa \partial x_l}-\frac{1}{2}\frac{\partial\gamma_{\kappa l}}{\partial x_m}\frac{\partial g_{\kappa l}}{\partial x_i}+\frac{\partial\gamma_{\kappa l}}{\partial x_m}\frac{\partial g_{il}}{\partial x_\kappa}+\frac{\partial\gamma_{\kappa l}}{\partial x_i}\frac{\partial g_{m\kappa}}{\partial x_l}\\ &-\gamma_{\rho\sigma}\gamma_{\kappa l}\frac{\partial g_{i\rho}}{\partial x_l}\frac{\partial g_{m\sigma}}{\partial x_\kappa}+\gamma_{\rho\sigma}\gamma_{\kappa l}\frac{\partial g_{il}}{\partial x_\rho}\frac{\partial g_{m\sigma}}{\partial x_\kappa}.\end{aligned} \tag{477}$$

This is the result Einstein wanted on this page: in harmonic coordinates, the Ricci tensor is the sum of a core-operator term, which is the only remaining second-order derivative term, and terms quadratic in first-order derivatives. He seems to have checked this carefully, for at the bottom of the page he wrote: “Result certain. Valid for coordinates which satisfy the equation $\Delta\varphi = 0$” (“Resultat sicher. Gilt für Koordinaten, die der Gl. $\Delta\varphi = 0$ genügen”). These coordinates were well-known and were called “isothermal coordinates” in the contemporary literature.[233] They are now commonly known as “harmonic coordinates.” Einstein’s notation suggests that he (or Grossmann) found the coordinate condition $\Delta\varphi = 0$ in the literature.[234]

19R

5.4.2 *Discovering a Conflict between the Harmonic Coordinate Restriction, the Weak-Field Equations, and Energy-Momentum Conservation (19R)*

On p. 19L, the mathematical strategy had finally born fruit. Einstein had found a way of constructing field equations out of the Ricci tensor that satisfy the correspondence principle. He now had to check whether these field equations and the harmonic restriction used in their construction also satisfy his other heuristic requirements. Unfortunately, in the course of checking the conservation principle on p. 19R, he discovered a problem.

The considerations on p. 19R are all in the context of a first-order, weak-field approximation. The metric is assumed to be the sum of a diagonalized Minkowski metric and small deviations from this metric. With the help of an imaginary time coordinate, introduced explicitly further down on the page, the zeroth order metric can be written as $\mathrm{diag}(1, 1, 1, 1)$.

Einstein began by writing down the harmonic restriction (see equation (471)) in this weak-field approximation, writing: “For the first approximation our additional condition is” (“Für die erste Annäherung lautet unsere Nebenbedingung”)[235]

---

233 See, e.g., (Bianchi 1910, sec. 36-37) or (Wright 1908, sec. 57).

234 Einstein had already used this condition in a different but equivalent form on p. 6L (see equation (83)), but it is unclear whether he recognized the equivalence.

235 Einstein’s notation here and in the following is awkward. He used the same summation index κ for two different summations, and did not explicitly distinguish between the diagonal zeroth-order metric and the first-order metric with small deviations from it. Introducing the more explicit notation $g_{\mu\nu} \equiv \delta_{\mu\nu} + \bar{g}_{\mu\nu}$ where $\delta_{\mu\nu}$ is the Kronecker delta and $\bar{g}_{\mu\nu} \ll 1$, one can rewrite equation (478) more carefully as: $\delta^{\kappa\kappa'}\left(2\frac{\partial\bar{g}_{i\kappa}}{\partial x_{\kappa'}}-\frac{\partial\bar{g}_{\kappa\kappa'}}{\partial x_i}\right)$.

$$\sum_{\kappa} \gamma_{\kappa\kappa}\left(2\frac{\partial g_{i\kappa}}{\partial x_{\kappa}} - \frac{\partial g_{\kappa\kappa}}{\partial x_i}\right) = 0. \tag{478}$$

Einstein conjectured that this condition "can *perhaps* be decomposed into" ("Zerfällt *vielleicht* in") the following two conditions,[236]

$$\sum \gamma_{\kappa\kappa}\frac{\partial g_{i\kappa}}{\partial x_{\kappa}} = 0,^{237} \tag{479}$$

a condition equivalent, at least in first-order approximation, to the Hertz restriction, and

$$\sum \gamma_{\kappa\kappa} g_{\kappa\kappa} = \text{konst},^{238} \tag{480}$$

a condition on the trace of the weak-field metric.

As will become clear below, Einstein wanted to ensure energy-momentum conservation by imposing the linearized Hertz restriction (479). On p. 19L he had introduced the harmonic restriction (478) to satisfy the correspondence principle. The combination of these two restrictions implies equation (480), which was unacceptable to Einstein. Einstein became aware of these implications in the course of his considerations concerning energy-momentum conservation on the remainder of this page.

On the next line, Einstein wrote down the "equations" ("Gleichungen")

$$\sum \gamma_{\kappa\kappa}\frac{\partial^2 g_{im}}{\partial x_{\kappa}^2} = \kappa\rho_0 \frac{dx_i}{ds}\frac{dx_m}{ds} g_{ii} g_{mm},^{239} \tag{481}$$

which are just the field equations of p. 19L in first-order approximation. The left-hand side is the core-operator term of the reduced Ricci tensor (see expression (477)). To first order, this is the only term that contributes. The right-hand side gives the covariant version of the stress-energy tensor for pressureless dust, multiplied by the gravitational constant κ .

For Einstein, energy-momentum conservation required that the density of the four-force of the gravitational field on the pressureless dust can be written as the four-divergence of a quantity that can be interpreted as representing gravitational stress-energy

236 The combination of these two new conditions is, in fact, stronger than the original condition.

237 Using the notation introduced in footnote 235, one would write equation (479) as: $\delta^{\kappa\kappa'}\frac{\partial \bar{g}_{i\kappa}}{\partial x_{\kappa'}} = 0$.

238 Using the notation introduced in footnote 235, one would write this equation as: $\delta^{\kappa\kappa'}\bar{g}_{\kappa\kappa'} = 0$. In other words, the condition is that the trace of $\bar{g}_{i\kappa}$ vanish. This is the form in which the condition is written at the top of the next page (p. 20L): $\sum g^{x}_{\kappa\kappa} = 0$.

239 Einstein erroneously wrote $\frac{dx_m}{dx_s}$ instead of $\frac{dx_m}{ds}$. Using the notation introduced in footnote 235, one would write this equation as: $\delta^{\kappa\kappa'}\frac{\partial^2 \bar{g}_{im}}{\partial x_{\kappa}\partial x_{\kappa'}} = \kappa\rho_0\frac{dx_{i'}}{ds}\frac{dx_{m'}}{ds}\delta_{ii'}\delta_{mm'}$.

density. Einstein checked whether his new field equations would actually allow him to rewrite the gravitational force density in the form of such a divergence.[240]

The expression for the force density

$$\frac{1}{2}\sqrt{G}\sum\frac{\partial g_{\mu\nu}}{\partial x_m}T^b_{\mu\nu}\,, \tag{482}$$

was introduced on p. 5R (see expression (67) above). Here $T^b_{\mu\nu}$ is the (contravariant) stress-energy tensor for pressureless dust. The force density gives the rate at which four-momentum is transferred from the gravitational field to the pressureless dust. As such it enters into the energy-momentum balance between matter and gravitational field for which Einstein had introduced the equation

$$\sum_{\nu n}\frac{\partial}{\partial x_n}(\sqrt{G}g_{m\nu}T_{\nu n})-\frac{1}{2}\sqrt{G}\sum\frac{\partial g_{\mu\nu}}{\partial x_m}T_{\mu\nu}=0 \tag{483}$$

on the next line on p. 5R (see equation (71); the superscript b was silently dropped). This equation is equivalent to the statement that the covariant divergence of the stress-energy tensor $T_{\mu\nu}$ vanishes. If (minus) the force density can be written as the divergence of gravitational stress-energy density, then equation (483) can be written as the vanishing of the *ordinary* divergence of the sum of quantities representing the stress-energy density of matter (pressureless dust in this case) and of the gravitational field, respectively. To find out whether some candidate field equations allow such rewriting of the force density, one substitutes their left-hand side (divided by κ) for the stress-energy tensor in the second term of equation (483) and tries to rewrite the resulting expression as a divergence.

In the first-order approximation considered on p. 19R, the second term in equation (483) reduces to

$$-\frac{1}{2}\frac{\partial g_{im}}{\partial x_\sigma}T_{im}\,. \tag{484}$$

This expression, while not actually written down on p. 19R, provides the link between equation (481), giving the field equations in first-order approximation, and the equation written on the next line. Eliminating T_{im} from expression (484) with the help of equation (481) and neglecting a factor -2κ, one arrives at the equation that Einstein did write,[241]

$$\sum_{\kappa im}\gamma_{\kappa\kappa}\frac{\partial^2 g_{im}}{\partial x_\kappa^2}\frac{\partial g_{im}}{\partial x_\sigma} = \sum_{\kappa im}\gamma_{\kappa\kappa}\left[\frac{\partial}{\partial x_\kappa}\left(\frac{\partial g_{im}}{\partial x_\kappa}\frac{\partial g_{im}}{\partial x_\sigma}\right)-\frac{1}{2}\frac{\partial}{\partial x_\sigma}\left(\frac{\partial^2 g_{im}{}^2}{\partial x_\kappa}\right)\right]. \tag{485}$$

240 A completely analogous consideration can be found in the *Entwurf* paper (Einstein and Grossmann 1913, 15). In his second static theory of 1912, Einstein likewise made sure that the (ordinary three-)force density could be written as the divergence of a gravitational stress tensor (Einstein 1912b, 456).

Since the right-hand side of this equation does indeed have the form of a divergence, Einstein concluded that "energy-momentum conservation holds in the relevant approximation" ("Energie- u. Impulssatz gilt mit der in Betr[acht] kommenden Annäherung").

Einstein still had to check whether the harmonic restriction (see equation (478)) and the two conditions into which it had tentatively been split (see equations (479) and (480)) are also compatible with energy-momentum conservation. Presumably, the second part of the comment immediately following Einstein's conclusion that "energy-momentum conservation holds in the relevant approximation" refers to this issue: "uniqueness and additional conditions" ("Eindeutigkeit u. Nebenbedingungen"). As to the first part of this comment, Einstein apparently hoped that his heuristic requirements would uniquely determine the field equations. This may well have been the motive for his (inconclusive) investigation on pp. 17L–18R of the relation between the Ricci tensor and the two-index object extracted from the curvature scalar on pp. 14R–16R. Einstein, however, did not actually address the uniqueness problem on p. 19R. The question regarding the additional conditions appears to have been more pressing.

He once again wrote down the linearized field equations, this time in the more compact form[242]

$$\Box g_{im} = \kappa\rho_0 \frac{dx_i}{d\tau}\frac{dx_m}{d\tau} . \tag{486}$$

The considerations on the remainder of the page suggest that Einstein discovered the following problem. In first-order approximation, covariant derivatives become ordinary derivatives and the energy-momentum balance between pressureless dust and gravitational field reduces to the conservation law

$$\frac{\partial}{\partial x_m}\left(\rho_0 \frac{dx_i}{d\tau}\frac{dx_m}{d\tau}\right) = 0 . \tag{487}$$

The easiest way to ensure that the field equations (486) are compatible with equation (487) is to impose the linearized Hertz restriction[243]

---

241 Einstein apparently substituted the covariant object on the left-hand side of equation (481) for the contravariant tensor T_{im} in expression (484). Correcting this and using the notation introduced in footnote 235, one would write equation (485) as

$$\delta^{\kappa\kappa'}\frac{\partial^2\bar{\gamma}_{im}}{\partial x_\kappa \partial x_{\kappa'}}\frac{\partial \bar{g}_{im}}{\partial x_\sigma} = \delta^{\kappa\kappa'}\left[\frac{\partial}{\partial x_\kappa}\left(\frac{\partial\bar{\gamma}_{im}}{\partial x_{\kappa'}}\frac{\partial \bar{g}_{im}}{\partial x_\sigma}\right) - \frac{1}{2}\frac{\partial}{\partial x_\sigma}\left(\frac{\partial\bar{\gamma}_{im}}{\partial x_{\kappa'}}\frac{\partial \bar{g}_{im}}{\partial x_\kappa}\right)\right].$$

242 The equation has covariant indices on the left- and contravariant indices on the right-hand side. Since indices are raised and lowered with the Kronecker delta in this approximation, this does not really matter. Using the notation introduced in footnote 235, one could write the equation more carefully as: $\Box\bar{\gamma}_{im} = \kappa\rho_0\frac{dx_i}{d\tau}\frac{dx_m}{d\tau}$. Einstein also wrote dx_τ instead of $d\tau$ again (cf. footnote 239).

$$\frac{\partial g_{im}}{\partial x_m} = 0 . \tag{488}$$

Equations (486) and (488) immediately imply equation (487):

$$\kappa \frac{\partial}{\partial x_m}\left(\rho_0 \frac{dx_i}{d\tau}\frac{dx_m}{d\tau}\right) = \frac{\partial}{\partial x_m}(\Box g_{im}) = \Box\left(\frac{\partial g_{im}}{\partial x_m}\right) = 0 . \tag{489}$$

As we have seen (cf. equation (479)), the linearized Hertz restriction is just one of the two restrictions in the tentative decomposition of the harmonic restriction given at the top of the page. But now a problem arises, which lies neither with the Hertz restriction nor with the harmonic restriction taken by itself, but with the combination of the two. Together these two restrictions imply equation (480), which says that the trace of the linearized metric has to be a constant.

This was objectionable for two reasons. First, through the field equations, it imposes the condition that the trace of the stress-energy tensor vanish (cf. equation (486)), which is clearly violated in the case under consideration, viz. pressureless dust.[244] Secondly, the trace of the metric $\mathrm{diag}(\pm 1, \pm 1, \pm 1, c^2(x,y,z))$, which Einstein used to represent static fields (see p. 6R and p. 39L), is obviously not a constant.

To avoid these problems, Einstein considered giving up the Hertz restriction. That means that it is no longer guaranteed that the divergence of the stress-energy tensor vanishes. The calculations at the bottom of p. 19R suggest that Einstein was prepared to consider the possibility that this divergence is non-vanishing. The result of these calculations, however, convinced him that was not an option. And from this he inferred that the Hertz restriction, which is the natural way of forcing the divergence of the stress-energy tensor to vanish, also had to be retained.

Einstein wrote down the "continuity condition" ("Kontinuitätsbedingung") for a cloud of pressureless dust, with "density of material points" ("Dichte materieller Punkte") $\rho_0/\sqrt{1-\mathfrak{q}^2/c^2}$,[245]

$$-\frac{\partial}{ic\partial t}\left(\frac{\rho_0 ic}{\sqrt{1-\frac{\mathfrak{q}^2}{c^2}}}\right) = \frac{\partial}{\partial x}\left(\frac{\rho_0 \mathfrak{q}_x}{\sqrt{1-\frac{\mathfrak{q}^2}{c^2}}}\right) + \frac{\partial}{\partial y}\left(\frac{\rho_0 \mathfrak{q}_y}{\sqrt{1-\frac{\mathfrak{q}^2}{c^2}}}\right) + \frac{\partial}{\partial z}\left(\frac{\rho_0 \mathfrak{q}_z}{\sqrt{1-\frac{\mathfrak{q}^2}{c^2}}}\right) . \tag{490}$$

---

243 Or rather (see the preceding note) that: $\partial\bar{\gamma}_{im}/\partial x_m = 0$. This relation is equivalent to equation (488), which with the help of the notation introduced in footnote 235 can be written as: $\delta^{mm'}(\partial\bar{g}_{im}/\partial x_{m'}) = 0$.

244 In the second of his four communications to the Prussian Academy of November 1915, Einstein briefly revived the condition that the stress-energy tensor for matter be traceless (Einstein 1915b, 799) as it is for the electromagnetic stress-energy tensor. At that point he suggested that this constraint can be reconciled with the non-vanishing of the trace of the stress-energy tensor for pressureless dust by assuming that gravity plays an essential role in the constitution of matter. For discussion, see, e.g., "Untying the Knot …" sec. 7 (in this volume).

245 Einstein only wrote an abbreviated form of this equation indicating the last two terms by dots.

To combine the continuity equation with equation (487), expressing energy-momentum conservation for a cloud of pressureless dust, Einstein introduced the four-velocity[246]

$$\mathfrak{w}_i \equiv \frac{dx_i}{d\tau} = \frac{\dot{x}_i}{\sqrt{1-\frac{\mathfrak{q}^2}{c^2}}}, \tag{491}$$

in which the relation $d\tau = dt\sqrt{1-\mathfrak{q}^2/c^2}$ (with $\mathfrak{q} \equiv (\dot{x}, \dot{y}, \dot{z})$) between proper time and coordinate time has been used, and in which the dot indicates differentiation with respect to t. With the imaginary time coordinate $u = ict$, the $\mathfrak{w}_u$-component becomes $ic/\sqrt{1-\mathfrak{q}^2/c^2}$. Using this new notation, Einstein rewrote the continuity equation as[247]

$$\frac{\partial}{\partial x}(\rho_0\mathfrak{w}_x) + \frac{\partial}{\partial y}(\rho_0\mathfrak{w}_y) + \frac{\partial}{\partial z}(\rho_0\mathfrak{w}_\mathfrak{z}) + \frac{\partial}{\partial u}(\rho_0\mathfrak{w}_u) = 0 \ . \tag{492}$$

Similarly, the x-component of the divergence of the stress-energy tensor for pressureless dust (see equation (487)) can be rewritten as

$$\begin{aligned}\frac{\partial}{\partial x_j}(\rho_O\mathfrak{w}_x\mathfrak{w}_j) &= \frac{\partial}{\partial x}(\rho_0\mathfrak{w}_x\mathfrak{w}_x) + \frac{\partial}{\partial y}(\rho_0\mathfrak{w}_x\mathfrak{w}_y) + \frac{\partial}{\partial z}(\rho_0\mathfrak{w}_x\mathfrak{w}_z) + \frac{\partial}{\partial u}(\rho_0\mathfrak{w}_x\mathfrak{w}_u)\\ &= \left[\frac{\partial}{\partial x}(\rho_0\mathfrak{w}_x) + \frac{\partial}{\partial y}(\rho_0\mathfrak{w}_y) + \frac{\partial}{\partial z}(\rho_0\mathfrak{w}_z) + \frac{\partial}{\partial u}(\rho_0\mathfrak{w}_u)\right]\mathfrak{w}_x\\ &\quad + \rho_0\mathfrak{w}_x\frac{\partial\mathfrak{w}_x}{\partial x} + \rho_0\mathfrak{w}_y\frac{\partial\mathfrak{w}_x}{\partial y} + \rho_0\mathfrak{w}_z\frac{\partial\mathfrak{w}_x}{\partial z} + \rho_0\mathfrak{w}_u\frac{\partial\mathfrak{w}_x}{\partial u}.\end{aligned} \tag{493}$$

The term in square brackets vanishes because of the continuity equation. Bringing the remaining terms over to the left-hand side, one arrives at the final equation of p. 19R,[248]

$$\begin{aligned}&\frac{\partial}{\partial x}(\rho_0\mathfrak{w}_x\mathfrak{w}_x) + \frac{\partial}{\partial y}(\rho_0\mathfrak{w}_x\mathfrak{w}_y) + \frac{\partial}{\partial z}(\rho_0\mathfrak{w}_x\mathfrak{w}_z) + \frac{\partial}{\partial u}(\rho_0\mathfrak{w}_x\mathfrak{w}_u)\\ &-\rho_0\mathfrak{w}_x\frac{\partial\mathfrak{w}_x}{\partial x} - \rho_0\mathfrak{w}_y\frac{\partial\mathfrak{w}_x}{\partial y} - \rho_0\mathfrak{w}_z\frac{\partial\mathfrak{w}_x}{\partial z} - \rho_0 w_u\frac{\partial\mathfrak{w}_x}{\partial u} = 0.\end{aligned} \tag{494}$$

Einstein noticed that the last four terms on the left-hand side add up to

$$-\rho\frac{D\mathfrak{w}_x}{D\tau}. \tag{495}$$

---

246 He wrote $\mathfrak{w}_i$ and $\mathfrak{w}_m$ above the relevant terms on the right-hand side of equation (486); and he wrote $\mathfrak{w}_x$ and $\mathfrak{w}_u$ for the expressions $\dot{x}/\sqrt{\quad}$ and $ic/\sqrt{\quad}$ in the lower right-hand corner of the page.

247 In the notebook the third term is indicated only by a dot.

248 In the notebook the last two terms on both the first and the second line are indicated only by dots.

The vanishing of this expression and the corresponding y- and z-components is the condition that the dust particles move on geodesics in what in this first-order approximation is Minkowski spacetime. Looking back at equation (493), we thus see that the vanishing of the divergence of the stress-energy tensor follows directly from the continuity equation and the equations of motion for the dust cloud.[249] It was therefore not an option for Einstein to drop the divergence requirement. The easiest way to guarantee the divergence requirement was to impose the Hertz restriction (see equations (487)–(489)). So, Einstein wanted to hold on to the Hertz restriction to satisfy the conservation principle and at the same time he wanted to hold on to the harmonic restriction to satisfy the correspondence principle. As Einstein put it: "*both* restrictions are to be retained" ("*Beide* obige Bedingungen sind aufrecht zu erhalten"). He had hit upon a serious problem: the harmonic restriction plus the Hertz restriction implied the unacceptable condition that the trace of the linearized metric be a constant.

5.4.3 *Modifying the Weak-field Equations: the Linearized Einstein Tensor (20L, 21L)*

20L At the top of p. 20L, Einstein once again wrote down the two conditions[250]

$$\sum \frac{\partial g_{i\kappa}}{\partial x_\kappa} = 0, \qquad \sum g^{x}_{\kappa\kappa} = 0. \tag{496}$$

into which he had tentatively decomposed the harmonic restriction at the top of p. 19R (cf. equations (479)–(480)). The harmonic condition was used to reduce the Ricci tensor to the d'Alembertian acting on the metric in the weak-field case. The Hertz restriction was added to make sure that the divergence of the stress-energy tensor vanishes in the weak-field case. The combination of these two restrictions implies that the trace of the metric has to vanish. This is problematic for a couple of reasons. First, if the metric is traceless, the weak-field equations (486) tell us that the stress-energy tensor be traceless as well. This last inference can be avoided by modifying the weak-field equations so as to make their right-hand side traceless. This is precisely what Einstein did on the next line:[251]

249 This cannot have come as a great surprise for Einstein, for it was precisely from the equations of motion for a cloud of pressureless dust that he had derived the equation for energy-momentum conservation (see, e.g., equation (483)) in the first place (see p. 5R).

250 In the second equation Einstein used the superscript "x" to indicate that he was considering first-order deviations from the flat metric. The same convention was used elsewhere in the notebook (see pp. 41L, 10L, 10R, 12L, 12R).

251 Using the notation of footnote 235 along with the d'Alembertian $\Box$ and the stress-energy tensor $T_{i\kappa} = \rho_0 \frac{dx_i}{d\tau}\frac{dx_\kappa}{d\tau}$ (with trace T), one can rewrite this equation more carefully as: $\Box g_{i\kappa} = \kappa\left(T_{i'\kappa'} - \frac{1}{4}\delta^{i'\kappa'}T\right)\delta_{ii'}\delta_{\kappa\kappa'}$. Einstein omitted the gravitational constant κ, and instead of the Kronecker delta $\delta^{i'\kappa'}$, he wrote "for the same i and κ" ("für gleiche i u. κ") underneath the second term on the right-hand side.

$$\sum_{\sigma}\frac{\partial^2 g_{i\kappa}}{\partial x_\sigma^2} = \rho_0\frac{dx_i}{d\tau}\frac{dx_\kappa}{d\tau} - \left(\frac{1}{4}\rho_0\sum\frac{dx_\kappa}{d\tau}\frac{dx_\kappa}{d\tau}\right). \tag{497}$$

With these new weak-field equations, the condition on the trace of the metric tensor no longer implies any condition on the trace of the stress-energy tensor.[252] Still, Einstein must have found the result unsatisfactory, for he crossed out the two lines with equations (496)–(497).

Eq. (497) does indeed only solve part of the problem caused by the combination of the harmonic and the Hertz restrictions. It takes care of the problem that a traceless metric would imply a traceless energy-momentum tensor, but it does not address a second problem, namely that a metric of the form Einstein used to represent static fields is not traceless.

A more satisfactory way to solve the problems would be to modify the weak-field equations in such a way that one avoids the condition that the metric be traceless altogether. This can be done by modifying the field equations in such a way that the harmonic restriction ensures both the elimination of unwanted second-order derivative terms for the Ricci tensor and the vanishing of the divergence of the stress-energy tensor. The combination of the harmonic and the Hertz restriction is thus replaced by the harmonic restriction alone and the problematic condition that the metric be traceless no longer follows. This is exactly the way in which Einstein took care of the problem in the next line.

First he wrote down the harmonic restriction again in first-order approximation (cf. equation (478) and footnote 235 for a more careful notation),

$$\sum\left(\frac{\partial g_{i\kappa}}{\partial x_\kappa} - \frac{1}{2}\frac{\partial g_{\kappa\kappa}}{\partial x_i}\right) = 0, \tag{498}$$

underlining the left-hand side. He introduced the abbreviation

$$\sum g_{\kappa\kappa} = U \tag{499}$$

for the trace of the metric. Instead of adding a term with the trace of the stress-energy tensor to the right-hand side of the weak-field equations as he had done in equation (497), he now added a term with the trace of the metric to the left-hand side. He wrote down a few components of these modified "gravitational equations" ("Gravitationsgleichungen"), indicating the remaining components by a dot and three lines of dashes[253]

252 In the fourth communication of November 1915, Einstein (1915d) similarly added a term with the trace T of the energy-momentum tensor of matter to his field equations to avoid the condition $T = 0$. For discussion, see "Untying the Knot …" sec. 7 (in this volume).

$$\begin{array}{cccccc} \Delta\left(g_{11}-\frac{1}{2}U\right) = T_{11} & & \Delta g_{12} = T_{12} & . & \Delta g_{14} = T_{14} & \\ \rule{2em}{0.4pt}\ \rule{1em}{0.4pt}\ \rule{1.5em}{0.4pt}\ \rule{1.5em}{0.4pt} & & \rule{2em}{0.4pt}\ \rule{2em}{0.4pt} & & \rule{2.5em}{0.4pt}\ \rule{2em}{0.4pt} & \qquad (500) \\ \rule{2em}{0.4pt}\ \rule{1em}{0.4pt}\ \rule{1.5em}{0.4pt}\ \rule{1.5em}{0.4pt} & & \rule{2em}{0.4pt}\ \rule{2em}{0.4pt} & & \rule{2.5em}{0.4pt}\ \rule{2em}{0.4pt} & \\ \rule{2em}{0.4pt}\ \rule{1em}{0.4pt}\ \rule{1.5em}{0.4pt}\ \rule{1.5em}{0.4pt} & & \rule{2em}{0.4pt}\ \rule{2em}{0.4pt} & & \rule{2.5em}{0.4pt}\ \rule{2em}{0.4pt} & \end{array}$$

In modern notation, using the Kronecker delta, these equations can be written more compactly as[254]

$$\Box\left(g_{i\kappa}-\frac{1}{2}\delta_{i\kappa}U\right) = T_{i\kappa}\,. \tag{501}$$

One can now ensure compatibility between the weak-field equations and the vanishing of the divergence of the stress-energy tensor by imposing

$$\frac{\partial}{\partial x_\kappa}\left(g_{i\kappa}-\frac{1}{2}\delta_{i\kappa}U\right) = 0\,. \tag{502}$$

which is just the harmonic restriction (see equation (498)). The calculation for the modified weak-field equations and the harmonic restriction is completely analogous to the calculation for the original weak-field equations and the Hertz restriction (cf. equation (489)),

$$\frac{\partial T_{i\kappa}}{\partial x_\kappa} = \frac{\partial}{\partial x_\kappa}\left(\Box\left(g_{i\kappa}-\frac{1}{2}\delta_{i\kappa}U\right)\right) = \Box\left(\frac{\partial}{\partial x_m}\left(g_{i\kappa}-\frac{1}{2}\delta_{i\kappa}U\right)\right) = 0\,. \tag{503}$$

In other words, Einstein's modification of the weak-field equations removed the need for the Hertz restriction (equation (479)), and thereby the need for the troublesome trace condition (equation (480)).

The modified weak-field equations (500) have exactly the same form as the weak-field equations for the final theory of November 1915.[255] The left-hand side is the linearized version of the Einstein tensor $R_{\mu\nu}-(1/2)g_{\mu\nu}R$. There is no indication in the notebook that Einstein tried to find the exact equations corresponding to the weak-field equations with trace term.

---

253 The Δ used here apparently denotes the d'Alembertian operator $\Delta \equiv \sum_{i=1}^{4}\frac{\partial^2}{\partial x_i^2}$, $(x_4 = ict)$ which had been denoted by the $\Box$ on the preceding page (p. 19R). The notation $\Box$ does not occur anywhere else in the notebook. The use of Δ as the analogue of the Laplace operator or the Laplace-Beltrami operator in four dimensions can also be found at the bottom of p. 19L. For two dimensions it is used on p. 10R, for three dimensions on p. 40L.

254 Strictly speaking, one should write the right-hand side as $T_{i'\kappa'}\delta_{ii'}\delta_{\kappa\kappa'}$, since $T_{i\kappa}$ is a contravariant tensor (cf. footnote 251).

255 (Einstein 1915d).

Einstein did write the modified weak-field equations in an alternative form. Taking the trace on both sides of equation (500), he found

$$2\Delta U = \sum T_{\kappa\kappa} \tag{504}$$

(the factor 2 on the left-hand side should be -1).With the help of this relation, Einstein replaced the term with the trace of the metric on the left-hand side of equation (500) by a term with the trace of the stress-energy tensor on the right-hand side. He obtained, writing “from this equations” (“Hieraus Gleichungen”)

$$\begin{array}{llll} \Delta g_{11} = T_{11} + \frac{1}{2}\sum T_{\kappa\kappa} & \Delta g_{12} = T_{12} & . & \Delta g_{14} = T_{14} \\ \text{\_\_\_\_\_ \_\_\_ \_\_\_\_ \_\_\_\_} & \text{\_\_\_\_\_ \_\_\_\_\_} & & \text{\_\_\_\_\_\_\_ \_\_\_\_\_} \\ \text{\_\_\_\_\_ \_\_\_ \_\_\_\_ \_\_\_\_} & \text{\_\_\_\_\_ \_\_\_\_\_} & & \text{\_\_\_\_\_\_\_ \_\_\_\_\_} \\ \text{\_\_\_\_\_ \_\_\_ \_\_\_\_ \_\_\_\_} & \text{\_\_\_\_\_ \_\_\_\_\_} & & \text{\_\_\_\_\_\_\_ \_\_\_\_\_} \end{array} \tag{505}$$

Einstein partly corrected his error in equation (504). The plus sign in equation (505), however, should be a minus sign. In modern notation, the correct equations can be written more compactly as (cf. equation (501))

$$\Box g_{ij} = T_{ij} - \frac{1}{2}\delta_{ij}\left(\sum T_{\kappa\kappa}\right). \tag{506}$$

Einstein proceeded to check that the modified weak-field equations (in the form of equation (500) rather than in the form of equation (505)) still allow the gravitational four-force density to be written in the form of a divergence. In first-order approximation, the force density is given by (see expression (484)):

$$\frac{1}{2}T_{i\kappa}\frac{\partial g_{i\kappa}}{\partial x_\sigma}. \tag{507}$$

Using equation (501) to eliminate $T_{i\kappa}$ from this expression, one arrives at

$$-\frac{1}{2}\Delta\left(g_{i\kappa} - \frac{1}{2}\delta_{i\kappa}U\right)\frac{\partial g_{i\kappa}}{\partial x_\sigma} = -\frac{1}{2}\left(\Delta g_{i\kappa}\frac{\partial g_{i\kappa}}{\partial x_\sigma} - \frac{1}{2}\Delta U\frac{\partial U}{\partial x_\sigma}\right) \tag{508}$$

On p. 19R, Einstein had already established that the first term on the right-hand side can be written in the form of a divergence (see equation (485)). It only remained for him to verify that this is true for the second term as well. He started to rewrite this term at the bottom of p. 20L as

$$\begin{aligned} -\frac{1}{2}\sum \Delta U\frac{\partial g_{\kappa\kappa}}{\partial x_\sigma} &= \cancel{-\frac{1}{2}\sum \frac{\partial^2 g_{\alpha\alpha}}{\partial x_\beta \partial x_\beta}}\frac{\partial g_{\kappa\kappa}}{\partial x_\sigma} = -\frac{1}{2}\sum \Delta U\frac{\partial U}{\partial x_\sigma} \\ &= -\frac{1}{2}\sum\left(\frac{\partial^2 U}{\partial x^2} + . + . + .\right)\frac{\partial U}{\partial x_\sigma} \end{aligned} \tag{509}$$

Although the final expression still does not have the form of a divergence, Einstein concluded that it is "representable in the required form" ("Darstellbar in der verl[angten] Form"). Since the two terms on the right-hand side of equation (508) have the exact same structure and since the result had already been established for the first, this conclusion is obvious.

21L Nevertheless, Einstein made a fresh start with this whole calculation on p. 21L, this time considering both terms in equation (508).[256] At the top of p. 21L Einstein wrote down the right-hand side of equation (508)

$$-\frac{1}{2}\left|\sum_{i\kappa}\Delta g_{i\kappa}\frac{\partial g_{i\kappa}}{\partial x_\sigma}-\frac{1}{2}\sum_{\kappa}\Delta U\frac{\partial U}{\partial x_\sigma}\right. . \tag{510}$$

He rewrote the sum in the second term as

$$\frac{\partial^2 U}{\partial x_\nu^2}\frac{\partial U}{\partial x_\sigma} = \frac{\partial}{\partial x_\nu}\left(\frac{\partial U}{\partial x_\nu}\frac{\partial U}{\partial x_\sigma}\right)-\frac{1}{2}\frac{\partial}{\partial x_\sigma}\left(\left(\frac{\partial U}{\partial x_\nu}\right)^2\right), \tag{511}$$

and the one in the first as

$$\frac{\partial^2 g_{i\kappa}}{\partial x_\nu^2}\frac{\partial g_{i\kappa}}{\partial x_\sigma} = \frac{\partial}{\partial x_\nu}\left(\frac{\partial g_{i\kappa}}{\partial x_\nu}\frac{\partial g_{i\kappa}}{\partial x_\sigma}\right)-\frac{1}{2}\frac{\partial}{\partial x_\sigma}\left(\left(\frac{\partial g_{i\kappa}}{\partial x_\nu}\right)^2\right). \tag{512}$$

Inserting these results into expression (510), he wrote the gravitational force density in the required form of a coordinate divergence of gravitational stress-energy density,[257]

$$\begin{aligned}&\sum_{i\kappa\nu}\frac{\partial}{\partial x_\nu}\left(\frac{\partial g_{i\kappa}}{\partial x_\nu}\frac{\partial g_{i\kappa}}{\partial x_\sigma}\right)-\frac{1}{2}\sum_{i\kappa\nu}\frac{\partial}{\partial x_\sigma}\left(\left(\frac{\partial g_{i\kappa}}{\partial x_\nu}\right)^2\right)\\ &-\frac{1}{2}\sum_{\nu}\frac{\partial}{\partial x_\nu}\left(\frac{\partial U}{\partial x_\nu}\frac{\partial U}{\partial x_\sigma}\right)+\frac{1}{4}\sum\frac{\partial}{\partial x_\sigma}\left(\left(\frac{\partial U}{\partial x_\nu}\right)^2\right).\end{aligned} \tag{513}$$

At some point, Einstein deleted the second line and wrote that the trace "U must vanish" ("U muss verschwinden"). He probably meant that the *derivatives* of U must vanish. It is unclear why he resurrected this condition on the trace of the metric tensor. He later deleted the remark about U, but did not rescind the deletion of the second line of expression (513).

From the fragmentary calculations on the remainder of the page, one can infer that Einstein somehow wanted to produce an exact analogue of his first-order calculation. Exactly how and for what purpose remains unclear. Perhaps he wanted to verify that the exact field equations allow one to rewrite the gravitational force density as the

---

256 This seemingly redundant calculation may have been done in connection with the calculation at the bottom of p. 21R involving the gravitational stress tensor for Einstein's 1912 static theory.

257 Notice that he did not take into account the factor $-1/2$ in front of expression (510), which was probably only added later.

divergence of the gravitational stress-energy density as well; or he wanted to find the exact expression for the quantity representing gravitational stress-energy density on the basis of the approximative expression that could be read off from expression (513). Whatever the purpose of these calculations, he drew a horizontal line and wrote[258]

$$\frac{\partial g_{i\kappa}}{\partial x_\sigma}\gamma_{\mu\nu}\frac{\partial^2 g_{i\kappa}}{\partial x_\mu \partial x_\nu} = \frac{\partial}{\partial x_\nu}\left(\gamma_{\mu\nu}\frac{\partial g_{i\kappa}}{\partial x_\sigma}\frac{\partial g_{i\kappa}}{\partial x_\mu}\right) - \frac{\partial g_{i\kappa}}{\partial x_\mu}\frac{\partial}{\partial x_\nu}\left(\frac{\partial g_{i\kappa}}{\partial x_\sigma}\gamma_{\mu\nu}\right). \tag{514}$$

Contrary to its linearized analogue (equation (512)), the right-hand side of this equation cannot be written as a divergence. Einstein noted that the last term can be written as

$$-\gamma_{\mu\nu}\frac{\partial g_{i\kappa}}{\partial x_\mu}\frac{\partial^2 g_{i\kappa}}{\partial x_\nu \partial x_\sigma}, \tag{515}$$

if the Hertz restriction,

$$\sum \frac{\partial \gamma_{\mu\nu}}{\partial x_\nu} = 0, \tag{516}$$

is imposed. This still does not make it possible, however, to rewrite the right-hand side of equation (514) as a divergence. At this point, the calculation breaks off.[259]

5.4.4 *Reexamining the Presuppositions Concerning the Static Field (20R, 21R)*

On p. 19R Einstein had found that the compatibility between the field equations constructed out of the Ricci tensor on p. 19L and the correspondence and conservation principles required that the trace of the linearized metric be a constant. This condition was problematic for several reasons, one of which was that this requirement is not satisfied by a metric of the form $\mathrm{diag}(1, 1, 1, c^2(x, y, z))$ which Einstein used to represent weak static fields. On p. 20L Einstein showed that the condition could be avoided by adding a trace term to the weak-field equations, but then these weak-field equations themselves no longer allow a metric of the form $\mathrm{diag}(1, 1, 1, c^2)$ as a solution.[260,261]

---

258 Note that the expression is ill-defined since it contracts over pairs of covariant indices. The corresponding approximative calculations, of course, had the same problem, but there it was only a matter of awkward notation (see notes 241–242).

259 Einstein drew another horizontal line, started to write down, but then immediately deleted, the transformation law of what he referred to as the "second tensor" ("zweiter Tensor transformiert").

260 This is most easily seen when these modified weak-field equations are written in the form of equation (505). Consider a weak-field generated by some static mass distribution. The only non-vanishing component of the stress-energy tensor will be the 44 -component. If the weak-field equations are the ones in equation (505), the metric of a static weak field will deviate from the Minkowski metric in all its diagonal components, and not just in its 44 -component, as Einstein expected.

261 A metric of this form is also incompatible with the harmonic restriction with which the field equations of p. 19L were extracted from the Ricci tensor. It is unclear whether Einstein was aware of this problem at this point.

Before giving up the promising new field equations in the face of these problems, Einstein reexamined whether his presuppositions concerning the static field were actually justified. On p. 21R he found an argument that convinced him they were. The argument involves the dynamics of point particles in a gravitational field, which Einstein reviewed on p. 20R. Einstein argued that, unless the 44 -component is the only variable component of the metric of a static field, particles with different energy, and hence different inertial mass, fall with different accelerations in such fields. He thus saw himself forced to give up the field equations considered on pp. 19L–20L.

20R The discussion on p. 20R of the mechanics of point particles and continuous matter distributions in a gravitational field is essentially the same as the discussions on p. 5R and in (Einstein and Grossmann 1913, sec. 2 and 4). At the top of the page, Einstein wrote down the line element

$$\sum g_{\mu\nu} dx_\mu dx_\nu = d\tau^2 , \tag{517}$$

and introduced the Lagrangian for a point particle of unit mass moving in a given metric field

$$\eta = \frac{d\tau}{dt} = \sqrt{g_{\mu\nu}\dot{x}_\mu \dot{x}_\nu} . \tag{518}$$

This last equation is written here somewhat more compactly than in the notebook. Though he indicated most terms by dots, Einstein expanded the sum under the square root sign,[262] writing (x, y, z, t) for the coordinates x_μ. He continued to do so in most of the equations on this page. His argument is easier to follow, however, if the more compact notation is used.

On the basis of the Euler-Lagrange equations,

$$\frac{\partial\eta}{\partial x_i} - \frac{d}{dt}\left(\frac{\partial\eta}{\partial \dot{x}_i}\right) = 0 , \tag{519}$$

where $i = 1, 2, 3$, the quantities

$$-\frac{\partial\eta}{\partial x_i} = -\frac{1}{2}\frac{dt}{d\tau}\frac{\partial g_{\mu\nu}}{\partial x_i}\frac{dx_\mu}{dt}\frac{dx_\nu}{dt} \tag{520}$$

can be interpreted as the components of the force on the particle,[263] and the quantities

$$-\frac{\partial\eta}{\partial \dot{x}_i} = -\frac{1}{2}\frac{dt}{d\tau}\frac{\partial}{\partial \dot{x}_i}(g_{\mu\nu}\dot{x}_\mu \dot{x}_\nu) = -g_{i\mu}\frac{dx_\mu}{d\tau} \tag{521}$$

as the components of its momentum. The particle's energy is given by the Legendre transform

262 We use a hybrid summation convention here, as did Einstein in the notebook at various places.
263 In the notebook, the factor $1/2$ on the right-hand side is omitted.

$$\begin{aligned}\eta - \sum \frac{\partial\eta}{\partial \dot{x}_i}\dot{x}_i &= \frac{d\tau}{dt} - \left(g_{i\mu}\frac{dx_\mu}{d\tau}\right)\frac{dx_i}{dt} \\ &= \frac{1}{dtd\tau}(d\tau^2 - g_{i\mu}dx_i dx_\mu) \\ &= \frac{1}{dtd\tau}(g_{4\mu}dx_4 dx_\mu) \\ &= g_{4\mu}\frac{dx_\mu}{d\tau}.\end{aligned} \tag{522}$$

Equations (521)–(522) show that "minus momentum and energy form a four-vector" ("Negativer Impuls u. Energie bilden Vierervektor"), whose components can be written as

$$g_{\mu\rho}\frac{dx_\rho}{d\tau}. \tag{523}$$

For a particle that does not have unit mass, this expression has to be "multiplied by [its rest mass] m" ("Noch mit m zu mult[iplizieren]").

Instead of one particle, Einstein now considered a continuous mass distribution. Dividing the four-momentum in expression (523) (multiplied by m) by the volume V, which can be written as[264]

$$V = \frac{1}{\sqrt{G}}\frac{d\tau}{dt}V_0 = \frac{1}{\sqrt{G}}\frac{d\tau}{dt}\frac{m}{\rho_0}, \tag{524}$$

he introduced the four-momentum density[265]

$$g_{\mu\rho}\frac{dx_\rho}{d\tau}\frac{dx_4}{d\tau}\rho_0\sqrt{G}. \tag{525}$$

Einstein now drew a horizontal line and wrote down the contravariant stress-energy tensor for pressureless dust, or, as he called it, the "tensor of material flow" ("Tensor der materiellen Strömung")

$$T_{i\kappa} = \rho_0\frac{dx_i}{d\tau}\frac{dx_\kappa}{d\tau}. \tag{526}$$

He lowered one index and introduced the notation

$$T'_{i\kappa} = \sum_i g_{\nu i}T_{i\kappa}, \tag{527}$$

264 (Einstein and Grossmann 1913, 10). Einstein had already used this equation on p. 5R (see equation (61)).

265 In the notebook, "momentum density" ("Impulsdichte") and "energy density" ("Energiedichte") are introduced separately.

for the "resulting mixed tensor" ("Hieraus gemischter Tensor"), which he explicitly identified as "stress-energy tensor" ("Sp[annungs]-Energie-Tensor"). The components of the four-momentum density introduced in expression (525) are simply the $\mu 4$-component of $\sqrt{G}T_{\mu\nu}'$.

Finally, Einstein wrote down the divergence of this mixed tensor density

$$\frac{\partial}{\partial x_\kappa}(\sqrt{G}g_{\nu i}T_{i\kappa}) . \tag{528}$$

According to the energy-momentum balance between matter and gravitational field, which Einstein had derived on p. 5R on the basis of considerations closely analogous to those on this page, the sum of this divergence and the force density must vanish. The latter is given by the force per unit mass (see equation (520)) multiplied by m and divided by V (see equation (524)). The result is[266]

$$-\frac{1}{2}\sqrt{G}\rho_0\frac{\partial g_{\kappa\lambda}}{\partial x_\nu}\frac{dx_\kappa}{d\tau}\frac{dx_\lambda}{d\tau} = -\frac{1}{2}\sqrt{G}\frac{\partial g_{\kappa\lambda}}{\partial x_\nu}T_{\kappa\lambda}. \tag{529}$$

At the bottom of the page, however, Einstein only wrote down expression (528), the first term in the energy-momentum balanced.

21R At the top of p. 21R, he returned to the consideration of force and energy rather than of force and energy densities. He wrote down an expression equivalent to equation (520), the first component of which gives the "x-component of the ponderomotive force" ("x-Komponente der ponderomotorischen Kraft") on a point particle of unit mass,[267]

$$-\frac{\sum\frac{\partial g_{\mu\nu}}{\partial x_\sigma}\dot{x}_\mu\dot{x}_\nu}{\sqrt{g_{44}+g_{11}\dot{x}^2+\ldots}} , \tag{530}$$

and an expression equivalent to equation (522) for the "energy of the point" ("Energie des Punktes"),

$$\frac{\left(g_{14}\frac{dx}{dt}+g_{24}\frac{dy}{dt}+\cdot+g_{44}\right)}{\sqrt{g_{44}+g_{11}\dot{x}^2+\ldots}} . \tag{531}$$

Right next to this expression, Einstein explicitly wrote down the metric of a static field

266 In the expression for the force density or "force per unit volume" ("Kraft pro Volumeinheit") in the notebook, a factor $\rho_0/2$ is missing.

267 A crossed-out factor of $\rho_0\sqrt{G}$ occurs in the numerator of both this expression and the next, which suggests that Einstein at some point considered force and energy densities.

$$\begin{matrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & c^2 \end{matrix} \,, \qquad (532)$$

which would allow him to recover his 1912 static theory from a theory based on the metric tensor. He noted that the g_{i4}-components "definitely vanish in a static field" ("g_{14} g_{24} … verschwinden sicher im statischen Felde"). In the static case, the numerator in expression (531) for the particle's energy thus reduces to g_{44}. Special relativity tells us that energy is proportional to inertial mass. Galileo's principle, i.e., the principle that the gravitational acceleration is the same for all bodies, tells us that the gravitational force is proportional to inertial mass as well. Expressions (530) and (531), however, imply that, unless all spatial components of the metric are constants, the ratio of force and energy in a static field,

$$\frac{\sum \frac{\partial g_{\mu\nu}}{\partial x_\sigma} \dot{x}_\mu \dot{x}_\nu}{g_{44}} \,, \qquad (533)$$

will depend on the particle's velocity. As Einstein put it: "If the force is supposed to vary like the energy, then g_{11}, g_{22} etc. must vanish for the static field" ("Soll die Kraft sich ändern wie die Energie, so müssen im statischen Felde g_{11}, g_{22} etc. verschwinden").[268] Although not stated explicitly, Einstein's conclusion was that the metric of the static field has to be of the form (532).[269]

This argument has a certain *prima facie* plausibility, but it does not hold up under closer scrutiny. On p. 20R, Einstein had identified a particle's momentum p_i (see equation (521)) and the force F_i acting on it (see equation (520) and expression (530)) in such a way that the spatial components of the geodesic equation, the Euler-Lagrange equations for the Lagrangian in equation (518), can be written in a form reminiscent of Newton's second law

$$\frac{dp_i}{dt} = F_i \,. \qquad (534)$$

For Einstein's argument on p. 21R to be valid one would have to be able to substitute (modulo a proportionality constant) $E\dot{x}_i$ (where the energy E is given by equation (522) or expression (531)) for p_i in equation (534). This substitution, however, is not

---

268 More accurately, the *deviations* of "g_{11}, g_{22} etc." from their constant Minkowskian values must vanish in the static case.

269 Presumably, although this is not made explicit, Einstein only wanted to draw the conclusion that static fields must be of this form *in first-order approximation*. He did not seem the least bit disturbed when in June 1913, in an attempt to calculate the perihelion advance of Mercury on the basis of the *Entwurf* theory, he found that the metric field of the sun in second-order approximation is not spatially flat (cf. [p. 6] of the Einstein-Besso manuscript [CPAE 4, Doc. 14]).

allowed. In other words, there is no reason to think that the antecedent of Einstein's conditional (i.e., "the force varies like energy") is true, and the argument fails.

Einstein drew two figures next to expressions (530)–(531), presumably to illustrate his argument, although their purpose remains unclear. The upshot of Einstein's considerations, however, is unambiguous. He had found an argument, based on a fundamental postulate of classical mechanics and completely independent of the gravitational field equations, that seemed to show that the metric of static gravitational fields has to be spatially flat. This confirmed his ideas about how to recover both Newton's theory and his own 1912 theory for static gravitational fields from the metric theory.[270] A metric of this form, however, was incompatible with the modified weak-field equations introduced on p. 20L.[271] Einstein therefore gave up the idea of constructing field equations out of the Ricci tensor with the help of the harmonic restriction.

21R

5.4.5 *Embedding the Stress Tensor for Static Gravitational Fields into the Metric Formalism (21R)*

For a metric of the form $\mathrm{diag}(-1, -1, -1, c^2)$ Einstein expected his new metric theory to reduce to his 1912 theory for static gravitational fields. That implied that one should also recover the expression for the gravitational stress tensor of the 1912 theory. On the bottom half of p. 21R, Einstein tried to translate the expression for this stress tensor into the language of the metric theory, replacing factors c, c^2, and $1/c^2$ by $\sqrt{-G}$, g_{44}, and γ_{44}, respectively.[272] Exactly what Einstein hoped to achieve remains unclear. His comments and the fact that he deleted the calculation in its entirety do make clear, however, that he was unhappy with the results. Part of the problem may have been that the expression for the stress tensor of the 1912 theory did not seem to agree with the corresponding components of the quantity representing gravitational stress-energy density constructed on p. 21L.

Under the heading "static special case" ("Statischer Spezialfall"), Einstein wrote down the X_x-component of the stress tensor of (the final version of) his 1912 static theory[273]

$$X_x = \frac{1}{c}\frac{\partial c}{\partial x}\frac{\partial c}{\partial x} - \frac{1}{2c}\mathrm{grad}^2 c \ . \tag{535}$$

---

270 In Einstein to Erwin Freundlich, 19 March 1915 (CPAE 8, Doc. 63), Einstein once again addressed the question "whether matter at rest can generate any other gravitational field than a g_{44}-field" ("ob ruhende Materie ein anderes Gravitationsfeld als ein g_{44}-Feld erzeugen kann"). "It cannot" ("Dies ist nicht der Fall"), he wrote. In support of this claim, he presented a calculation done in first-order approximation and based on the *Entwurf* field equations (i.e., on weak-field equations without a trace term). He made no reference to any other arguments for his claim.

271 It is also incompatible with the harmonic restriction (see footnote 261).

272 On p. 39L, Einstein had attempted a similar translation of the field equations of the 1912 theory (see sec. 2.2, equations (9)–(18)).

273 (Einstein 1912b, 456, equation (5)).

Using the relation $\frac{\partial c}{\partial x} = \frac{1}{c}\frac{\partial}{\partial x}\left(\frac{c^2}{2}\right)$, he rewrote this equation as

$$\begin{aligned} 4X_x &= \frac{1}{c^3}\frac{\partial c^2}{\partial x}\frac{\partial c^2}{\partial x} - \frac{1}{2c^3}\mathrm{grad}^2 c^2 \\ &= c\frac{1}{c^2}\frac{1}{c^2}\frac{\partial c^2}{\partial x}\frac{\partial c^2}{\partial x} - \frac{1}{2}\left(c\frac{1}{c^2}\frac{1}{c^2}\mathrm{grad}^2 c^2\right) \end{aligned} \tag{536}$$

He then translated this expression into the language of the metric theory, using that $c = \sqrt{-G}$, $g_{44} = c^2$, and $\gamma_{44} = 1/c^2$ for a metric of the form $\mathrm{diag}(-1, -1, -1, c^2)$. In this way, he arrived at

$$4X_x = \sqrt{-G}\left(\gamma_{44}\gamma_{44}\frac{\partial g_{44}}{\partial x}\frac{\partial g_{44}}{\partial x} - \frac{1}{2}\gamma_{44}\gamma_{44}\left(\sum_\nu \frac{\partial g_{44}}{\partial x_\nu}\frac{\partial g_{44}}{\partial x_\nu}\right)\right) \tag{537}$$

Underneath this expression Einstein wrote: “impossible because of divergence equation” (“Unmöglich wegen Divergenzgleichung”). The “divergence equation” is presumably the equation setting the gravitational four-force density equal to the divergence of gravitational stress-energy density. In the notation of the *Entwurf*, this equation can be written as[274]

$$-\frac{1}{2}\sqrt{-g}\frac{\partial g_{\mu\nu}}{\partial x_\sigma}T^{\mu\nu} = \frac{\partial}{\partial x_\nu}(\sqrt{-g}g_{\sigma\mu}t^{\mu\nu}), \tag{538}$$

where $T^{\mu\nu}$ and $t^{\mu\nu}$ denote the contravariant stress-energy tensor for matter and the corresponding pseudo-tensor for the gravitational field, respectively. On pp. 19R, 20L, and 21L, Einstein had checked, in a first-order approximation and using his candidate (weak) field equations to eliminate the stress-energy tensor for matter, whether the gravitational force density can be rewritten as a divergence. The tentative expression for gravitational stress-energy density constructed on p. 21R confronted Einstein with the converse problem, viz. whether the divergence of the gravitational stress-energy density actually gives the gravitational force density. It is not clear how Einstein could tell without further calculation that this is not possible—if that is in fact what he means by his remark “impossible because of divergence equation”—but Einstein may have reached this conclusion on the basis of a comparison of equation (537) with expression (513) on p. 21L, which (in linear approximation) gives the gravitational force density in the form of the divergence of gravitational stress-energy density.

The translation of equation (536) into equation (537) is not unique. Einstein, in fact, gave an alternative translation of the first term in the expression for the stress tensor

274 Cf., e.g., (Einstein and Grossmann 1913, 16–17, equations (12a) and (18)).

$$\sqrt{-G}\frac{\partial}{\partial x}\left(\frac{1}{c^2}\right)\frac{\partial c^2}{\partial x} = \sum_{i\kappa}\sqrt{G}\frac{\partial \gamma_{i\kappa}}{\partial x}\frac{\partial g_{i\kappa}}{\partial x}\,. \tag{539}$$

Apparently, this expression was not satisfactory either. Einstein deleted this whole calculation by a diagonal line, and wrote: "special case probably incorrect" ("Spezialfall wahrscheinlich unrichtig").

19L–21R

5.4.6 *Synopsis of the Problems with the Harmonic Restriction and the Linearized Einstein Tensor (19L–21R)*

The calculations on p. 21R mark the end of Einstein's consideration of field equations extracted from the Ricci tensor with the help of the harmonic restriction, and of the modified weak-field equations that we now recognize as the Einstein equations of the final theory in linearized form. To conclude this section, we summarize the chain of reasoning that produced this unfortunate turn of events.

On p. 19L, Einstein showed that the harmonic restriction can be used to eliminate unwanted second-order derivative terms from the Ricci tensor. On p. 19R, examining these new field equations in linear approximation, he found that the natural way to make sure that the weak-field equations are compatible with energy-momentum conservation is to impose a further coordinate restriction, viz. the Hertz restriction. The combination of the harmonic restriction and the Hertz restriction implies that the trace of the linearized metric must be a constant. To avoid this implication, Einstein (on p. 20L) added a trace term to the weak-field equations, effectively changing their left-hand side from the linearized Ricci tensor to the linearized Einstein tensor. This modification obviates the need for the Hertz restriction and thus for the condition on the trace of the weak-field metric. Part of the original problem, however, still persists, albeit in a different guise.

One of the difficulties with the restriction on the trace of the metric is that it is not satisfied by a metric of the form $\mathrm{diag}(-1, -1, -1, c^2)$. Einstein believed that his theory would not have a sensible Newtonian limit unless weak static fields can be represented by a metric of this form. At the same time, a metric of this form would allow him to recover his 1912 theory for static gravitational fields from the new metric theory. It was thus a serious problem that the restriction on the trace of the metric rules out a metric of this form.

Unfortunately, with the modification needed to avoid this restriction, the weak-field equations themselves no longer allow a solution with a metric of the form $\mathrm{diag}(-1, -1, -1, c^2)$. On p. 21R, Einstein therefore reexamined whether his presuppositions about the static field were justified. A fallacious argument convinced him that nothing less than Galileo's principle that all bodies fall with the same acceleration requires that the metric of static fields does indeed have the form he had been assuming.[275] This sealed the fate of the harmonic restriction and the linearized Einstein tensor. Einstein only returned to field equations including a trace term by a different route in November 1915.[276]

5.5 *Exploring the Ricci Tensor in Unimodular Coordinates (22L–24L, 42L–43L)*

On p. 22R Einstein considered a new way of extracting a candidate for the left-hand side of the field equations from the Riemann tensor. He started from a new expression for the Ricci tensor, now entirely in terms of the Christoffel symbols and their first-order derivatives. He split this tensor into two parts each of which separately transforms as a tensor under unimodular transformations. Restricting the allowed transformations to unimodular transformations, he took one of these parts as the new candidate for the left-hand side of the field equations and explored it on the following pages. The object returned in the field equations published in the first of Einstein's four communications to the Berlin Academy of November 1915.[277] We therefore call it the November tensor. 22R

The November tensor still contains terms with second-order derivatives of the metric in addition to a core-operator term. On p. 22R, Einstein imposed the Hertz restriction to eliminate those terms (see secs. 5.5.2–5.5.3). This coordinate restriction would return as the coordinate condition (in the modern sense) for the November tensor in (Einstein 1915a). The advantage of the Hertz restriction is that it serves two purposes. It can be used to eliminate unwanted terms with second-order derivatives of the metric, and it ensures that the divergence of the matter stress-energy tensor vanishes in a weak-field approximation (see p. 19R). On p. 19R, Einstein had been forced to introduce two separate conditions for these two purposes—the harmonic restriction and the Hertz restriction. The introduction of the Einstein tensor on p. 20L can be seen as an attempt to eliminate the Hertz restriction. The introduction of the November tensor can likewise be seen as an attempt to eliminate the harmonic restriction.

The expression extracted from the November tensor with the Hertz restriction contains a large number of terms quadratic in first-order derivatives of the metric. On p. 23L, Einstein added a further addition to eliminate most of these terms. He went back to the original form of the November tensor in terms of the Christoffel symbols and imposed a coordinate restriction with which he could eliminate two of the three terms of the Christoffel symbols. This coordinate restriction is to unimodular transformations under which an expression that we call the ϑ-expression transforms as a tensor. We call this restriction the ϑ-restriction. Einstein discovered that the ϑ-restriction not only eliminates many terms with first-order derivatives of the metric but that it also takes care of the unwanted second-order derivatives that he had eliminated earlier with the Hertz restriction. Einstein thus lifted the Hertz restriction and kept only the ϑ-restriction (p. 23L; discussed in sec. 5.5.4). He began to investigate which non-autonomous transformations are allowed by the ϑ-restriction. In the end, he aban- 23L

275 Einstein checked and confirmed this expectation once more in 1915 (see footnote 270). He only realized that a weak static field need not be represented by a spatially flat metric when he calculated the perihelion motion of Mercury in (Einstein 1915c).

276 For discussion, see "Untying the Knot …" secs. 5–6 (in this volume).

277 (Einstein 1915a).

doned the ϑ-restriction because it does not allow transformation to rotating frames in Minkowski spacetime (pp. 23R–24L, 42L–43L; discussed in secs. 5.5.5–5.5.9).

22R 5.5.1 *Extracting the November Tensor from the Ricci Tensor (22R)*

At the top of p. 22R, Einstein wrote down the covariant form of the Ricci tensor in the form[278]

$$T_{il} = \sum_{kl} \frac{\partial}{\partial x_l}\begin{Bmatrix} i & k \\ & k \end{Bmatrix} - \frac{\partial}{\partial x_k}\begin{Bmatrix} i & l \\ & k \end{Bmatrix} + \begin{Bmatrix} i & k \\ & \lambda \end{Bmatrix}\begin{Bmatrix} \lambda & l \\ & k \end{Bmatrix} - \begin{Bmatrix} i & l \\ & \lambda \end{Bmatrix}\begin{Bmatrix} \lambda & k \\ & k \end{Bmatrix}. \tag{540}$$

Contrary to the expressions for the Riemann and Ricci tensors earlier in the notebook, equation (540) is written entirely in terms of Christoffel symbols rather than in terms of the metric tensor and its derivatives. Apparently, it was Marcel Grossmann, whose name appears at the top of the page, who suggested this expression to Einstein. Grossmann may also have suggested some of the further manipulations of this expression on p. 22R. Recall that Grossmann's name was also written next to the first occurrence of the Riemann tensor in the notebook on p. 14L.

Two of the four terms in this new expression for the Ricci tensor can be combined to form a quantity that can easily be seen to be a tensor under unimodular transformations. Under unimodular transformations the determinant G of the metric transforms as a scalar. Hence, $\lg\sqrt{G}$ also transforms as a scalar under such transformations, and the ordinary derivative of this quantity as a vector. Einstein denoted this vector by T_i and wrote that "if G is a scalar, then [...] T_i a tensor of first rank" ("Wenn G ein Skalar ist, dann

$$\frac{\partial \lg\sqrt{G}}{\partial x_i} = T_i \tag{541}$$

Tensor 1. Ranges.").[279] Using the relation

$$\frac{\partial \lg\sqrt{-g}}{\partial x_i} = \begin{Bmatrix} i & \kappa \\ & \kappa \end{Bmatrix}, \tag{542}$$

one can identify two of the four terms of the Ricci tensor as the covariant derivative of the vector T_i. Einstein regrouped the terms in equation (540) accordingly,

278 The summation should be over k and λ rather than over k and l.

279 Note that this is probably the first time that vectors are called "tensors of first rank" (also note that this is all in the context of unimodular transformation only). In (Budde 1914), the generalization of tensors to arbitrary dimension and rank was credited to Grossmann's part of Einstein and Grossmann 1913: "Recently, Mr. Grossmann [...] has proposed a still further reaching generalization. He denotes quantities of arbitrary rank as "tensors," so that vectors, trivectors, and bitensors are also subsumed under the term "tensor;" the generalization consists in extending his definitions to structures of n th rank in m-dimensional space" (Budde 1914, 246). For further discussion, see the appendix to (Norton 1992) and (Reich 1994). The term "rank" ("Rang") appears in the notebook only on this page. On p. 14L the fourth rank Riemann tensor was called "Ebenentensor vierter Mannigfaltigkeit."

$$T_{il} = \left(\frac{\partial T_i}{\partial x_l} - \sum \begin{Bmatrix} i & l \\ & \lambda \end{Bmatrix} T_\lambda\right) - \sum_{k\lambda}\left(\frac{\partial}{\partial x_k}\begin{Bmatrix} i & l \\ & k \end{Bmatrix} - \begin{Bmatrix} i & k \\ & \lambda \end{Bmatrix}\begin{Bmatrix} \lambda & l \\ & k \end{Bmatrix}\right). \tag{543}$$

Under the first term in parentheses he wrote "tensor of second rank" ("Tensor 2. Ranges"); under the second he wrote "presumed tensor of gravitation T^x_{il}" ("Vermutlicher Gravitationstensor T^x_{il}"). Since taking the covariant derivative is a generally-covariant operation and T_i is a first-rank tensor under unimodular transformations, its covariant derivative is a second-rank tensor under unimodular transformations. The second term in parentheses in equation (543), the difference between the full generally-covariant Ricci tensor and the covariant derivative of T_i, will therefore also transform as a tensor under unimodular transformations. Einstein took this quantity,

$$T^x_{il} = \frac{\partial}{\partial x_k}\begin{Bmatrix} i & l \\ & k \end{Bmatrix} - \sum_{k\lambda}\begin{Bmatrix} i & k \\ & \lambda \end{Bmatrix}\begin{Bmatrix} \lambda & l \\ & k \end{Bmatrix} \tag{544}$$

as his new candidate for the left-hand side of the field equations. We shall refer to it as the November tensor.

5.5.2 *Extracting Field Equations from the November Tensor Using the Hertz Restriction (22L–R)*

By imposing the Hertz restriction in addition to the unimodularity restriction, one can reduce the November tensor to the required form of a core operator plus terms with products of first-order derivatives of the metric. This is what Einstein confirmed on the next two lines on p. 22R, as the first step in a "further rewriting of the tensor of gravitation" ("Weitere Umformung des Gravitationstensors"). 22R

The only terms in T^x_{il} with second-order derivatives of the metric occur in the term with derivatives of the Christoffel symbol. Einstein expanded this term to

$$\frac{\partial}{\partial x_k}\begin{Bmatrix} i & l \\ & k \end{Bmatrix} = \frac{1}{2}\frac{\partial}{\partial x_k}\left(\gamma_{k\alpha}\left(\frac{\partial g_{i\alpha}}{\partial x_l} + \frac{\partial g_{l\alpha}}{\partial x_i} - \frac{\partial g_{il}}{\partial x_\alpha}\right)\right), \tag{545}$$

and then eliminated all unwanted second-order derivative terms by assuming the Hertz restriction. "We presuppose" ("Wir setzen voraus"), he wrote, that

$$\sum_\kappa \frac{\partial \gamma_{\kappa\alpha}}{\partial x_\kappa} = 0, \tag{546}$$

adding that "then this [i.e., the right-hand side of equation (545)] is equal to" ("dann ist dies gleich"):

$$-\frac{1}{2}\sum \gamma_{k\alpha}\frac{\partial^2 g_{il}}{\partial x_\alpha \partial x_k} - \frac{1}{2}\sum\left(\frac{\partial \gamma_{k\alpha}}{\partial x_l}\frac{\partial g_{i\alpha}}{\partial x_k} + \frac{\partial \gamma_{k\alpha}}{\partial x_i}\frac{\partial g_{l\alpha}}{\partial x_k}\right). \tag{547}$$

In the notebook, the factors $1/2$ in front of both terms appear above the summation signs and were probably added later. The core-operator term in expression (547) comes from the last term on the right-hand side of equation (545). The first two terms in equation (545) turn into the products of first-order derivative terms in expression (547) with the help of the Hertz restriction and the relation

$$\gamma_{k\alpha}\frac{\partial g_{i\alpha}}{\partial x_l} = -\frac{\partial \gamma_{k\alpha}}{\partial x_l}g_{i\alpha}\,. \tag{548}$$

22L On the bottom half of p. 22L, we find what appears to be an earlier attempt at eliminating unwanted second-order derivative terms from the November tensor with the help of the Hertz restriction. Relabeling the indices in equation (545), one can write the first part of the November tensor as

$$\frac{\partial}{\partial x_l}\begin{Bmatrix} i & m \\ & l \end{Bmatrix} = \frac{1}{2}\frac{\partial}{\partial x_l}\left(\gamma_{\kappa l}\left(\frac{\partial g_{\kappa i}}{\partial x_m} + \frac{\partial g_{\kappa m}}{\partial x_i} - \frac{\partial g_{im}}{\partial x_\kappa}\right)\right)\,. \tag{549}$$

The first two terms on the right-hand side give rise to unwanted second-order derivative terms

$$\frac{1}{2}\gamma_{\kappa l}\left\{\frac{\partial^2 g_{\kappa i}}{\partial x_l \partial x_m} + \frac{\partial^2 g_{\kappa m}}{\partial x_l \partial x_i}\right\}. \tag{550}$$

Except for the factor $1/2$, this is just the expression that Einstein wrote directly underneath the horizontal line on p. 22L. The first term in expression (550), multiplied by a factor 2, can be rewritten as

$$-\frac{\partial \gamma_{kl}\partial g_{ki}}{\partial x_m \partial x_l} + \frac{\partial}{\partial x_m}\left(\gamma_{kl}\frac{\partial g_{ki}}{\partial x_l}\right)\,. \tag{551}$$

In the notebook, the first term, a product of first-order derivatives, is indicated only by a dot (and an expression similar to expression (551) for the second term in equation (550) is omitted altogether). The Hertz restriction ensures that the second term, which can be rewritten as

$$\frac{\partial}{\partial x_m}\left(-\frac{\partial \gamma_{kl}}{\partial x_l}g_{ki}\right)\,, \tag{552}$$

vanishes. As Einstein wrote directly underneath the second term in the expression (551) in the notebook: “suffices, if $\sum \partial\gamma_{\kappa l}/\partial x_l$ vanishes” (“Gen[ü]gt, wenn … verschwindet”).

22R We now return to p. 22R. On the bottom half of the page, Einstein turned his attention to terms in the November tensor (544) with products of first-order derivatives of the metric. He began by expanding the term with a product of Christoffel symbols:

$$\begin{aligned}\begin{Bmatrix} i & k \\ & \lambda \end{Bmatrix}\begin{Bmatrix} \lambda & l \\ & k \end{Bmatrix} &= \frac{1}{4}\gamma_{\lambda\alpha}\gamma_{k\beta}\left(\frac{\partial g_{i\alpha}}{\partial x_k} - \frac{\partial g_{ik}}{\partial x_\alpha} + \frac{\partial g_{\alpha k}}{\partial x_i}\right)\left(\frac{\partial g_{l\beta}}{\partial x_\lambda} - \frac{\partial g_{l\lambda}}{\partial x_\beta} + \frac{\partial g_{\lambda\beta}}{\partial x_l}\right) \\ &= -\frac{1}{4}\gamma_{\lambda\alpha}\gamma_{k\beta}\left(\frac{\partial g_{i\alpha}}{\partial x_k} - \frac{\partial g_{ik}}{\partial x_\alpha}\right)\left(\frac{\partial g_{l\lambda}}{\partial x_\beta} - \frac{\partial g_{l\beta}}{\partial x_\lambda}\right) + \frac{1}{4}\gamma_{\lambda\alpha}\gamma_{k\beta}\frac{\partial g_{\alpha k}}{\partial x_i}\frac{\partial g_{\lambda\beta}}{\partial x_l}.\end{aligned} \tag{553}$$

As in expression (547), the numerical factors were added later. In the second step, Einstein used the same symmetry argument that he had used on pp. 17R and 19L (see the discussion following expression (463)). In the following two lines, Einstein noted that the last term can be rewritten as

$$-\frac{\partial \gamma_{\lambda\alpha}}{\partial x_i}\frac{\partial g_{\lambda\alpha}}{\partial x_l} \tag{554}$$

"or" ("oder") as

$$-\frac{\partial \gamma_{\lambda\alpha}}{\partial x_l}\frac{\partial g_{\lambda\alpha}}{\partial x_i}. \tag{555}$$

Einstein also rewrote the terms with products of first-order derivatives in the first part (547) of the November tensor. Using the relation (in modern notation)

$$\begin{aligned} g^{\alpha\beta}{}_{,i}g_{\alpha l,\beta} &= \frac{1}{2}g^{\alpha\beta}{}_{,i}(g_{\alpha l,\beta} + g_{\beta l,\alpha} - g_{\alpha\beta,l}) + \frac{1}{2}g^{\alpha\beta}{}_{,i}g_{\alpha\beta,l} \\ &= g^{\alpha\beta}{}_{,i}\begin{bmatrix} \alpha & \beta \\ & l \end{bmatrix} + \frac{1}{2}g^{\alpha\beta}{}_{,i}g_{\alpha\beta,l} \end{aligned} \tag{556}$$

and relabeling indices, one can rewrite the expression in parentheses in expression (547) as[280]

$$\frac{\partial \gamma_{\alpha\beta}}{\partial x_l}\frac{\partial g_{i\beta}}{\partial x_\alpha} + \frac{\partial \gamma_{\alpha\beta}}{\partial x_i}\frac{\partial g_{l\beta}}{\partial x_\alpha} = \frac{\partial \gamma_{\alpha\beta}}{\partial x_l}\begin{bmatrix} \alpha & \beta \\ & i \end{bmatrix} + \frac{\partial \gamma_{\alpha\beta}}{\partial x_i}\begin{bmatrix} \alpha & \beta \\ & l \end{bmatrix} + \frac{\partial \gamma_{\alpha\beta}}{\partial x_i}\frac{\partial g_{\alpha\beta}}{\partial x_l} \tag{557}$$

Combining the expressions (547), (557), and (553)–(554), one arrives at

280 This is the first time in the notebook that Einstein introduced the Christoffel symbols to replace ordinary derivatives of the metric. Up to now the calculations in the notebook always proceeded by expanding the Christoffel symbols and by rewriting the tensor expressions in terms of simple derivatives of the metric instead of using the compact Christoffel symbols. Why Einstein proceeded the other way around here is not clear, but it is tempting to speculate that it reflects a first inkling on Einstein's part of the importance of the Christoffel symbols in his gravitational theory.

$$-2T^x_{il} = \gamma_{\alpha\beta}\frac{\partial^2 g_{il}}{\partial x_\alpha \partial x_\beta} - \frac{1}{2}\gamma_{\alpha k}\gamma_{\beta\lambda}\left(\frac{\partial g_{i\alpha}}{\partial x_\beta} - \frac{\partial g_{i\beta}}{\partial x_\alpha}\right)\left(\frac{\partial g_{lk}}{\partial x_\lambda} - \frac{\partial g_{l\lambda}}{\partial x_k}\right)$$
$$\frac{\partial \gamma_{\alpha\beta}}{\partial x_i}\begin{bmatrix}\alpha & \beta \\ & l\end{bmatrix} + \frac{\partial \gamma_{\alpha\beta}}{\partial x_l}\begin{bmatrix}\alpha & \beta \\ & i\end{bmatrix} + \frac{3}{2}\frac{\partial \gamma_{\alpha\beta}}{\partial x_i}\frac{\partial g_{\alpha\beta}}{\partial x_l} \tag{558}$$

for (minus 2 times) the reduced November tensor (i.e., the November tensor in unimodular coordinates satisfying the Hertz restriction). In the corresponding expression at the bottom of p. 22R in the notebook, the second line contains some errors: above the summation sign in front of the first two terms, Einstein added a factor $1/2$; and the third term has $1/4$ instead of $3/2$.

22L
5.5.3 *Non-autonomous Transformations Leaving the Hertz Restriction Invariant (22L)*

The field equations based on the reduced November tensor (see equation (558)) will be covariant under those unimodular transformations that preserve the Hertz restriction. On the top half of p. 22L, Einstein derived the condition for non-autonomous unimodular transformations leaving the Hertz restriction, and thus the corresponding field equations, invariant.[281] Note that this calculation precedes the calculations showing that the Hertz restriction can be used to eliminate unwanted second-order derivative terms from the November tensor on the bottom half of p. 22L and on p. 22R.

At the top of p. 22L, Einstein began by writing the Hertz restriction in primed coordinates

$$\sum \frac{\partial \gamma'_{\mu\nu}}{\partial x'_\nu} = 0\,. \tag{559}$$

Next to it, he wrote the determinant condition on the transformation matrix $p_{\mu\nu}$ for a unimodular transformation

$$|p_{\mu\nu}| = 1\,. \tag{560}$$

These two equations can be seen as a cryptic statement of the question being addressed on this page: given a metric field satisfying the Hertz restriction in some (unprimed) coordinate system, what are the unimodular coordinate transformations such that the Hertz restriction will be satisfied in the new (primed) coordinate system as well? The answer to this question takes the form of an equation for the transformation matrix $p_{\mu\nu}$ and its inverse $\pi_{\mu\nu}$. This equation involves the components of the metric field in the unprimed system. The transformations preserving the Hertz restriction, the solutions of this equation, will thus depend on which metric field one starts from in the

281 Earlier in the notebook and in a different context (see p. 10L ff.), Einstein had already investigated this question for infinitesimal transformations (see sec. 4.5.1). The calculation on p. 22L closely follows the one on p. 10L.

unprimed system. In other words, these transformations are examples of “non-autonomous transformations.”[282]

To find the equation for these non-autonomous transformations, Einstein transformed equation (559) from x'_μ- to x_μ-coordinates[283]

$$\sum \pi_{\nu i} \frac{\partial}{\partial x_i} \{ p_{\mu\alpha} p_{\nu\beta} \gamma_{\alpha\beta} \} = 0 . \tag{561}$$

Using the relation $\pi_{\nu i} p_{\nu\beta} = \delta_{i\beta}$, he rewrote the left-hand side as

$$p_{\mu\alpha} \frac{\partial \gamma_{\alpha i}}{\partial x_i} + \gamma_{\alpha\beta} \pi_{\nu i} \frac{\partial p_{\mu\alpha} p_{\nu\beta}}{\partial x_i} . \tag{562}$$

The first term vanishes on account of the assumption that forms the starting point of this calculation, viz. that the metric field satisfies the Hertz restriction in the unprimed coordinates. Einstein rewrote the second term as

$$\sum \gamma_{\alpha\beta} \pi_{\underline{\nu} i} \left\{ p_{\mu\alpha} \frac{\partial p_{\nu\beta}}{\partial x_i} + p_{\underline{\nu}\beta} \frac{\partial p_{\mu\alpha}}{\partial x_i} \right\} , \tag{563}$$

and then, on the next line, as

$$\sum \gamma_{\alpha i} \frac{\partial p_{\mu\alpha}}{\partial x_i} + \sum \gamma_{\alpha\beta} \pi_{\nu i} p_{\mu\alpha} \frac{\partial p_{\nu\beta}}{\partial x_i} . \tag{564}$$

The first term in this last expression was familiar to Einstein from the analogous calculation on p. 10L (cf. the second term in equation (234)), which may be why Einstein underlined it. On p. 10L he had found that the vanishing of the first term is the condition for *infinitesimal* unimodular non-autonomous transformations preserving the Hertz restriction. In the infinitesimal case, the second term vanishes (see footnote 148). This is probably why, on p. 22L, he wrote next to the second term in expression (564): “vanishes if funct[ional] det[erminant] = 1.” (“verschwindet, wenn Funkt. Det. = 1.”). For finite transformations, however, this term does not vanish, as Einstein presumably realized, for he included it in an attempt to further simplify the expression in expression (564) on the next line.

The first term in expression (564) can be rewritten as

$$\frac{\partial}{\partial x_i} (\gamma_{\alpha i} p_{\mu\alpha}) - p_{\mu\alpha} \frac{\partial \gamma_{\alpha i}}{\partial x_i} ; \tag{565}$$

the second term as

282 See the discussion in sec. 4.3. Other examples of conditions for “non-autonomous transformations” can be found on pp. 7L, 8R, 10L, and 23R.

283 The transformation matrices are defined as $p_{\alpha\beta} = \partial x'_\alpha / \partial x_\beta$ and $\pi_{\alpha\beta} = \partial x_\beta / \partial x'_\alpha$ (see equations (119)–(120)).

$$\frac{\partial}{\partial x_i}(\gamma_{\alpha\beta}\pi_{\nu i}p_{\mu\alpha}p_{\nu\beta}) - \frac{\partial}{\partial x_i}(\gamma_{\alpha\beta}\pi_{\nu i}p_{\mu\alpha})p_{\nu\beta}\,, \tag{566}$$

which in turn can be rewritten as

$$\frac{\partial}{\partial x_i}(\gamma_{\alpha i}p_{\mu\alpha}) - \gamma_{\alpha\beta}p_{\mu\alpha}p_{\nu\beta}\frac{\partial\pi_{\nu i}}{\partial x_i} - \frac{\partial}{\partial x_i}(\gamma_{\alpha i}p_{\mu\alpha})\,. \tag{567}$$

Adding expressions (565) and (567), one arrives at the expression given on the next line in the notebook:

$$\sum\frac{\partial}{\partial x_i}(\gamma_{\alpha i}p_{\mu\alpha}) - p_{\mu\alpha}\frac{\partial\gamma_{\alpha i}}{\partial x_i} + \frac{\partial}{\partial x_i}(\gamma_{\alpha i}p_{\mu\alpha}) - \gamma_{\alpha\beta}p_{\mu\alpha}p_{\nu\beta}\frac{\partial\pi_{\nu i}}{\partial x_i} - \frac{\partial}{\partial x_i}(\gamma_{\alpha i}p_{\mu\alpha})\,. \tag{568}$$

As Einstein noted, the second term vanishes (because of the Hertz restriction) and the third term cancels with the fifth. What is left is an expression that has basically the same structure as expression (564). On the next line, Einstein reverted to the latter. The upshot then was that the matrix $p_{\mu\nu}$ (and its inverse $\pi_{\mu\nu}$) for some unimodular coordinate transformation from coordinates x_μ to x'_μ must satisfy

$$\gamma_{\alpha i}\frac{\partial p_{\mu\alpha}}{\partial x_i} + \gamma_{\alpha\beta}\pi_{\nu i}p_{\mu\alpha}\frac{\partial p_{\nu\beta}}{\partial x_i} = 0 \tag{569}$$

to ensure that a metric field that satisfies the Hertz restriction in the x_μ-coordinates satisfies the Hertz restriction in the x'_μ-coordinates as well.

This is a rather complex condition. Examining the simpler version for infinitesimal transformations in which case the second term in equation (569) vanishes automatically, Einstein had found that it does not allow a transformation to uniformly accelerated frames of reference in the important special case of Minkowski spacetime (see pp. 10L–11L).[284] He nonetheless continued to use the Hertz restriction (see p. 23L).

23L–R

5.5.4 *Extracting Field Equations from the November Tensor Using the ϑ-Restriction (23L–R)*

At the bottom of p. 22R, Einstein had arrived at a candidate for the left-hand side of the field equations extracted from the November tensor by imposing the Hertz restriction (see eq. (558)). This candidate contains numerous terms with products of first-order derivatives of the metric. Most of these terms come from the product of Chri-
23L stoffel symbols in the second term of the November tensor (544). On p. 23L, Einstein returned to the expression for the November tensor in terms of Christoffel symbols and added a new coordinate restriction to the Hertz condition with which he could eliminate two of the three first-order derivative terms in every Christoffel symbol. In this way he could eliminate most of terms quadratic in first-order derivative terms

284 On p. 11L, Einstein had convinced himself that the Hertz restriction does allow rotations. Rotation, however, is also ruled out by the Hertz restriction (see sec. 4.5.2, especially equations (306)–(310) and notes 163–164)

found at the bottom of p. 22R. This additional coordinate restriction limits the range of allowed coordinate transformations to those unimodular transformations under which a quantity denoted by $\vartheta_{ik\lambda}$, which is essentially the fully symmetrized version of $g_{ik\lambda}$, transforms as a tensor. We call this quantity the "ϑ-expression," the unimodular transformations under which it transforms as a tensor "ϑ-transformations," and the restriction to such transformations the "ϑ-restriction."[285] Einstein's basic strategy on p. 23L was to subtract terms from the November tensor that by themselves transform as tensors under ϑ-transformations. What was left served as Einstein's new candidate for the left-hand side of the field equations. In the course of the calculations on p. 23L, Einstein came to realize that with the ϑ-restriction he did not need the Hertz restriction anymore to eliminate terms with unwanted second-order derivatives of the metric. The ϑ-restriction took care of those terms all by itself. Einstein therefore abandoned the Hertz restriction and focused on the ϑ-restriction.

On the third line of p. 23L, separated from the first two[286] by a horizontal line, Einstein stated the assumption that forms the starting point of the argument on this page, viz. that the quantity[287]

$$\frac{1}{2}\left(\frac{\partial g_{ik}}{\partial x_\lambda}+\frac{\partial g_{k\lambda}}{\partial x_i}+\frac{\partial g_{\lambda i}}{\partial x_k}\right) \tag{570}$$

"be a tensor $\vartheta_{ik\lambda}$" ("sei Tensor $\vartheta_{ik\lambda}$"). In other words, Einstein was interested in transformations under which this ϑ-expression would transform according to the transformation law for a fully covariant third-rank tensor. The next line explicitly gives this transformation law[288]

$$\vartheta'_{ikl} = \sum_{\alpha\beta\gamma}\vartheta_{\alpha\beta\gamma}\frac{\partial x_\alpha}{\partial x'_i}\frac{\partial x_\beta}{\partial x'_k}\frac{\partial x_\gamma}{\partial x'_l}. \tag{571}$$

Einstein used the ϑ-expression to rewrite the Christoffel symbols of the first and second kind as

---

285 The Hertz expression, it turns out, transforms as a vector under ϑ-transformations (see the discussion following eq. (585)).

286 At the top of p. 23L, Einstein began an explicit transformation of the ordinary derivative of the metric: $\frac{\partial g'_{ik}}{\partial x_\lambda} = \pi_{\lambda\alpha}\frac{\partial}{\partial x_\alpha}(\pi_{i\sigma}\pi_{k\tau}g_{\sigma\tau})$ (x_λ on the left-hand side should be x'_λ). The calculation breaks off almost immediately but is taken up again on the facing page, p. 23R, in order to find the condition for "non-autonomous transformations" under which the ϑ-expression is invariant (see sec. 5.5.5).

287 In the notebook, Einstein wrote "-2" above the last plus sign in expression (570). Above many of the plus and minus signs in subsequent expressions on this page the opposite has been written (and, in some cases, has been deleted again). These sign changes are related to Einstein's consideration on p. 25R of a variant of the ϑ-restriction, which we shall call the $\hat{\vartheta}$-restriction (see sec. 5.6.3 for discussion).

288 Note that Einstein deviates from the notation $p_{\alpha\beta}$ and $\pi_{\alpha\beta}$ typically used in the notebook for the transformation matrices $\partial x'_\alpha/\partial x_\beta$ and $\partial x_\beta/\partial x'_\alpha$.

$$\begin{bmatrix} i & l \\ & k \end{bmatrix} = \vartheta_{ilk} - \frac{\partial g_{il}}{\partial x_k} \tag{572}$$

and

$$\begin{Bmatrix} i & l \\ & k \end{Bmatrix} = \gamma_{k\alpha}\left(\vartheta_{il\alpha} - \frac{\partial g_{il}}{\partial x_\alpha}\right), \tag{573}$$

respectively. He substituted the latter expression into the November tensor (cf. equation (544)):

$$T_{il}^{x} = \frac{\partial}{\partial x_k}\left(\gamma_{k\alpha}\left(\vartheta_{il\alpha} - \frac{\partial g_{il}}{\partial x_\alpha}\right)\right) - \gamma_{\lambda\alpha}\gamma_{k\beta}\left(\vartheta_{ik\alpha} - \frac{\partial g_{ik}}{\partial x_\alpha}\right)\left(\vartheta_{l\lambda\beta} - \frac{\partial g_{l\lambda}}{\partial x_\beta}\right). \tag{574}$$

The rationale behind the ϑ-restriction now becomes clear. Any (combination of) term(s) in equation (574) that transforms as a tensor under ϑ-transformations can be subtracted from the November tensor and what is left will still transform as a tensor under ϑ-transformations. Originally, Einstein probably only meant to eliminate terms quadratic in first-order derivatives of the metric from the November tensor. He initially expected that the Hertz restriction would still be needed to eliminate terms with unwanted second-order derivatives. On the next line he explicitly stated the requirement that the Hertz expression,

$$\sum \frac{\partial \gamma_{k\alpha}}{\partial x_k}, \tag{575}$$

"be $= 0$" ("sei $= 0$"). He subsequently added the clause "is not necessary" ("ist nicht nötig").[289]

The end result of the calculation on p. 23L is:

$$\sum \frac{\partial}{\partial x_k}\left(\gamma_{k\alpha}\frac{\partial g_{il}}{\partial x_\alpha}\right) + \sum \gamma_{\rho\alpha}\gamma_{k\beta}\frac{\partial g_{ik}}{\partial x_\alpha}\frac{\partial g_{l\rho}}{\partial x_\beta}. \tag{576}$$

It turns out, as Einstein himself came to recognize, that there is a short-cut to get from eq. (574), which gives the November tensor in terms of the ϑ-expression, to eq. (576). If one sets $\vartheta_{ik\lambda} = 0$ in eq. (574), which amounts to replacing the Christoffel symbols by the truncated Christoffel symbols $-\gamma_{k\alpha}(\partial g_{il}/\partial x_\alpha)$ (see eq. (573)), one arrives at:

289 The evidence for our assumption that this clause was indeed added later is twofold. First, the awkwardness of the syntax of the sentence "$\sum \partial\gamma_{k\alpha}/\partial x_k$ *be* $= 0$ *is* not necessary" (our emphasis) disappears under the assumption. Second, Einstein uses the Hertz restriction in the equation immediately following this sentence.

$$-\frac{\partial}{\partial x_k}\left(\gamma_{k\alpha}\frac{\partial g_{il}}{\partial x_\alpha}\right)-\gamma_{\lambda\alpha}\gamma_{k\beta}\frac{\partial g_{ik}}{\partial x_\alpha}\frac{\partial g_{l\lambda}}{\partial x_\beta}, \tag{577}$$

which is just (minus) Einstein's expression (576).[290] This shows that the ϑ-restriction takes care of the terms with unwanted second-order derivatives in the November tensor as well. Once Einstein recognized this, he presumably added the clause "is not necessary" to his statement (575) of the Hertz restriction and made the necessary corrections to the derivation of eq. (576).

In the original derivation, Einstein used the Hertz restriction in the very first step. He introduced the quantity

$$T_{il}^{xx} = -\gamma_{k\alpha}\frac{\partial^2 g_{il}}{\partial x_k \partial x_\alpha}+\gamma_{k\alpha}\frac{\partial\vartheta_{il\alpha}}{\partial x_k}-\gamma_{\lambda\alpha}\gamma_{k\beta}\frac{\partial g_{ik}}{\partial x_\alpha}\frac{\partial g_{l\lambda}}{\partial x_\beta}+\gamma_{\lambda\alpha}\gamma_{k\beta}\left(\vartheta_{ik\alpha}\frac{\partial g_{l\lambda}}{\partial x_\beta}+\vartheta_{l\lambda\beta}\frac{\partial g_{ik}}{\partial x_\alpha}\right), \tag{578}$$

and remarked that this "is also a tensor" ("ist ebenfalls ein Tensor"). What he meant no doubt was: "transforms as a tensor under ϑ-transformations." The new tensor T_{il}^{xx} differs from T_{il}^{x} by two terms: $\gamma_{\lambda\alpha}\gamma_{k\beta}\vartheta_{ik\alpha}\vartheta_{l\lambda\beta}$ and

$$\frac{\partial\gamma_{k\alpha}}{\partial x_k}\left(-\frac{\partial g_{il}}{\partial x_\alpha}+\vartheta_{il\alpha}\right). \tag{579}$$

290 It is tempting to speculate that these considerations are related to a remark Einstein made in 1915 when he published field equations based on the November tensor. Einstein wrote that earlier he had looked upon the quantities "$\frac{1}{2}\sum_\mu g^{\tau\mu}\frac{\partial g_{\mu\nu}}{\partial x_\nu}$ as the natural expression for the components of the gravitational field although in the light of the formulae of the differential calculus, it is more natural to introduce the Christoffel symbols $\left\{\begin{matrix}\nu\sigma\\ \tau\end{matrix}\right\}$ instead of those quantities. This was a fateful prejudice." ("… als den natürlichen Ausdruck für die Komponenten des Gravitationsfeldes, obwohl es im Hinblick auf die Formeln des absoluten Differentialkalküls näher liegt, die Christoffelschen Symbole … statt jener Größen einzuführen. Dies war ein verhängnisvolles Vorurteil." Einstein 1915a, 782). This point is also emphasized in Einstein's letter to Arnold Sommerfeld of 28 November 1915: "What gave me the key to this solution [the Einstein field equations in their final form] was the insight that not $\sum_\mu g^{l\alpha}\frac{\partial g_{ai}}{\partial x_m}$ but the related Christoffel symbols $\left\{\begin{matrix}\nu\sigma\\ \tau\end{matrix}\right\}$ should be looked upon as the natural expression for the "components" of the gravitational field" ("Den Schlüssel zu dieser Lösung lieferte mir die Erkenntnis, dass nicht …sondern die damit verwandten Christoffel'schen Symbole … als natürlichen Ausdruck für die "Komponente" des Gravitationsfeld anzusehen ist." CPAE 8, Doc. 153). For further discussion, see sec. 5 of "Untying the Knot …" (in this volume).

The first one transforms as a tensor under ϑ-transformations itself. This means that T_{il}^{x} minus this term will still be a tensor under ϑ-transformations. The second one, expression (579), vanishes on account of the Hertz restriction. When Einstein subsequently retracted the Hertz condition, he indicated that $\gamma_{k\alpha}$ should be included within the scope of the derivative $\partial/\partial x_k$ in the first term on the right-hand side of equation (578). He neglected to do so for the second term. This may simply have been an oversight but the fragmentary calculation at the top of p. 23R suggests that Einstein may have realized that $(\partial\gamma_{k\alpha}/\partial x_k)\vartheta_{il\alpha}$ itself transforms as a tensor under ϑ-transformations. So this term can also be subtracted from T_{il}^{x} and the result will still be a tensor under ϑ-transformations.

On the next line, Einstein wrote down the covariant derivative of the ϑ-expression contracted with $\gamma_{k\alpha}$

$$\gamma_{k\alpha}\frac{\partial\vartheta_{il\alpha}}{\partial x_k} - \sum\left(\begin{Bmatrix} k & i \\ & \rho \end{Bmatrix}\vartheta_{\rho l\alpha} + \begin{Bmatrix} k & l \\ & \rho \end{Bmatrix}\vartheta_{i\rho\alpha} + \begin{Bmatrix} k & \alpha \\ & \rho \end{Bmatrix}\vartheta_{il\rho}\right)\gamma_{k\alpha}\,. \tag{580}$$

This expression will also transform as a tensor under ϑ-transformations, as Einstein noted: “Likewise [a tensor]” (“Ebenso”). He substituted equation (573) for the Christoffel symbols into expression (580) and subtracted terms with products of the ϑ-expression and the metric which are themselves tensors under ϑ-transformations. In this way, he arrived at

$$\gamma_{k\alpha}\frac{\partial\vartheta_{il\alpha}}{\partial x_k} + \sum\gamma_{k\alpha}\gamma_{\rho\beta}\left(\frac{\partial g_{ik}}{\partial x_\beta}\vartheta_{\rho\lambda\alpha} + \frac{\partial g_{kl}}{\partial x_\beta}\vartheta_{i\rho\alpha} + \frac{\partial g_{k\alpha}}{\partial x_\beta}\vartheta_{il\rho}\right), \tag{581}$$

and noted that this would “therefore also” (“also auch”) be “a tensor” (“ein Tensor”) under ϑ-transformations.

In modern notation, the last term in expression (581) can be rewritten as

$$g^{k\alpha}g^{\rho\beta}g_{k\alpha,\beta}\vartheta_{il\rho} = g^{\rho\beta}(\log{-g})_{,\beta}\vartheta_{il\rho}\,. \tag{582}$$

The determinant of the metric transforms as a scalar under all unimodular transformation and hence under ϑ-transformations. The derivative of its logarithm therefore transforms as a vector under ϑ-transformations.[291] As Einstein noted, the last term in expression (581) is therefore “itself a tensor” (“an sich ein Tensor”) under ϑ-transformations. It follows that the remaining three terms in expression (581) also form a tensor under ϑ-transformations. Precisely these three terms occur in the expression for T_{il}^{xx}, as Einstein verified by underlining them, both in equation (578) and in expression (581), and by relabeling some of the indices in equation (578). After “subtraction” (“Subtraktion”) of these three terms from the right-hand side of equation (578)

291 On p. 22R, it was explicitly noted that this quantity transforms as a vector under unimodular transformations.

and changing the minus signs in front of the two remaining terms to plus signs, he arrived at

$$\sum \gamma_{k\alpha} \frac{\partial^2 g_{il}}{\partial x_k \partial x_\alpha} + \sum \gamma_{\rho\alpha} \gamma_{k\beta} \frac{\partial g_{ik}}{\partial x_\alpha} \frac{\partial g_{l\rho}}{\partial x_\beta} \tag{583}$$

and noted that this still "is a tensor" ("ist Tensor") under ϑ-transformations. The combination of the Hertz restriction and the ϑ-restriction thus allowed Einstein to extract from the November tensor a candidate for the left-hand side of the field equations that has a remarkably simple form. It is the sum of a core-operator term and a term with a product of first-order derivatives.

It is probably at this point that Einstein noticed the short-cut from eq. (574) to expression (576). Expression (583) only differs from the latter in that $\gamma_{k\alpha}$ in the first term is not included within the scope of $\partial / \partial x_k$. Einstein marked this term to indicate that $\gamma_{k\alpha}$ should be included within the scope of $\partial / \partial x_k$. He did the same with the first term on the right-hand side of eq. (578), which is the starting point of the argument that got him from the November tensor to the new candidate for the left-hand side of the field equations (583). This amounts to rescinding the Hertz restriction, which explains why Einstein wrote "is not necessary" next to statement (575) of this restriction. The unwanted second-order derivatives that he had eliminated on p. 22R by imposing the Hertz restriction were absorbed into expression (580) for the covariant derivative of the ϑ-expression and subtracted from the November tensor. By including $\gamma_{k\alpha}$ within the scope of $\partial / \partial x_k$ in eq. (583), Einstein had thus extracted the following candidate for the left-hand side of the field equations by imposing the ϑ-restriction alone:

$$\sum \frac{\partial}{\partial x_k}\left(\gamma_{k\alpha} \frac{\partial g_{il}}{\partial x_\alpha}\right) + \sum \gamma_{\rho\alpha} \gamma_{k\beta} \frac{\partial g_{ik}}{\partial x_\alpha} \frac{\partial g_{l\rho}}{\partial x_\beta}\,. \tag{584}$$

Without the Hertz restriction, expression (584) inherits one more term involving the ϑ-expression from eq. (574), namely $(\partial \gamma_{k\alpha} / \partial x_k)\ \vartheta_{il\alpha}$ (cf. the discussion following expression (579)). It turns out, however, that this term is a tensor under ϑ-transformations itself, so that expression (584) without this term is still a tensor under ϑ-transformations.[292]

That $\partial \gamma_{k\alpha} / \partial x_k$ —and consequently $(\partial \gamma_{k\alpha} / \partial x_k)\ \vartheta_{il\alpha}$ —are indeed tensors under ϑ-transformations can be seen as follows. Consider the contraction of g^{ik} and ϑ_{ikl}:

292 Note that this correction term would only give rise to terms with products of first-order derivatives of the metric anyway, and hence would not affect the result that the ϑ-restriction can be used to eliminate unwanted second-order derivative terms.

$$
\begin{aligned}
g^{ik}\vartheta_{ikl} &= \frac{1}{2}g^{ik}(g_{ik,\,l}+g_{kl,\,i}+g_{li,\,k}) \\
&= g^{ik}\left(\frac{1}{2}g_{ik,l}+g_{kl,\,i}\right) \\
&= \frac{1}{2}(\log-g)_{,l}-g^{ik}{}_{,i}g_{kl}.
\end{aligned} \tag{585}
$$

Hence, $g^{ik}{}_{,i}g_{kl}$ is the difference between two expressions, $g^{ik}\vartheta_{ikl}$ and $\frac{1}{2}(\log-g)_{,l}$, that are both tensors under ϑ-transformations. It is therefore a tensor under ϑ-transformations itself. It follows that $g^{ik}{}_{,i}$ is a tensor under ϑ-transformations as well, which is the result that we wanted to prove.

Einstein may in fact have gone through a similar calculation. This is suggested by
23R the calculation at the top of p. 23R, where he wrote down the right-hand side of the first line of equation (585),

$$
\gamma_{ik}\left(\frac{\partial g_{ik}}{\partial x_l}+\frac{\partial g_{kl}}{\partial x_i}+\frac{\partial g_{li}}{\partial x_k}\right), \tag{586}
$$

as well as the terms $\sum\gamma_{ik}\frac{\partial g_{li}}{\partial x_k}$, $-\sum g_{li}\frac{\partial\gamma_{ik}}{\partial x_k}$, and $\sum\frac{\partial\gamma_{ik}}{\partial x_k}$, which are all involved in the calculation given in equation (585).[293]

23L To conclude our discussion of p. 23L, we want to emphasize that the calculation on this page nicely illustrates the usage of coordinate restrictions as opposed to coordinate conditions in the notebook. Thinking in terms of coordinate conditions, one would look upon expression (584) as representing the left-hand side of some candidate field equations of broad covariance in coordinates satisfying the ϑ-restriction chosen to facilitate comparison with Newtonian theory. It is not at all clear, however, whether expression (584) actually is the representation of some tensor of broader covariance in the class of coordinate systems determined by the ϑ-restriction. It can be seen as the November tensor in coordinates such that the ϑ-expression vanishes, as follows from the short-cut from equation (574) to equation (576). But the ϑ-restriction only requires that the ϑ-expression transform as a tensor, not that it vanish.

Given that Einstein conceived of the restriction to ϑ-transformations as an essential feature of the theory and not as a feature of a particular representation of the theory, it was of no real interest to him to find the tensor of broader covariance corresponding to expression (584). The relation between expression (584) and tensors of broader covariance, such as the November tensor or the Ricci tensor, is important only in that it allowed Einstein to construct field equations that are invariant under a precisely defined class of coordinate transformations. It is important to keep in mind

293 Another possibility is that this calculation was related to the investigation of a modified ϑ-expression on p. 43LA (cf. footnote 313).

that for Einstein this construction gave field equations, or candidate field equations, in their most general form, not just an expression of field equations of broader covariance in some restricted class of coordinate systems.

5.5.5 *Non-autonomous Transformations Leaving the ϑ-Expression Invariant (23R)* 23R

With the exception of the first few lines, which we tentatively identified as a fragmentary version of the calculation given in equation (585), the purpose of the calculations on p. 23R is to derive the condition for infinitesimal "non-autonomous transformations" leaving the ϑ-expression and thereby the field equations based on expression (584) invariant. As in the case of the corresponding condition for the Hertz restriction (see p. 22L and the discussion in sec. 5.5.3), this condition takes the form of an equation for the transformation matrices $p_{\mu\nu}$ and $\pi_{\mu\nu}$ involving the components of the metric field in the original coordinate system.[294]

After drawing a horizontal line under the fragmentary calculation at the top of the page, Einstein first derived the condition for non-autonomous transformations under which derivatives of the metric transform as tensors.[295] Adding three such equations with cyclically permuted indices, Einstein then derived the corresponding condition for the ϑ-expression.

He began by writing down how derivatives of the metric transform under an arbitrary transformation from unprimed to primed coordinates:[296]

$$\left(\frac{\partial g_{ik}}{\partial x_l}\right)' = \pi_{l\lambda}\frac{\partial}{\partial x_\lambda}(\pi_{i\iota}\pi_{k\kappa}g_{\iota\kappa}) . \tag{587}$$

If $(\partial g_{ik}/\partial x_l)'$ is transformed back to the unprimed system on the assumption that the coordinate transformation relating the two coordinate systems is no longer arbitrary but one under which derivatives of the metric actually transform as tensors, one finds

$$\begin{aligned}\frac{\partial g_{\rho\sigma}}{\partial x_\tau} &= p_{i\rho}p_{k\sigma}p_{l\tau}\frac{\partial g'_{ik}}{\partial x'_l} \\ &= p_{i\rho}p_{k\sigma}p_{l\tau}\pi_{l\lambda}\frac{\partial}{\partial x_\lambda}(\pi_{i\iota}\pi_{k\kappa}g_{\iota\kappa}),\end{aligned} \tag{588}$$

where in the second step equation (587) was used, which is valid for arbitrary transformations. Einstein's own description of the transition from equation (587) to the

294 Although Einstein presumably was interested only in unimodular transformations preserving the ϑ-expression, he did not explicitly impose the condition $|p_{\mu\nu}| = |\pi_{\mu\nu}| = 1$ on the determinant of the transformation matrices, as he had on p. 22L in the case of the Hertz restriction (see equation (560)).

295 This condition can also be found on p. 8R (see equation (169)).

296 The transformation matrices are defined as $p_{\alpha\beta} = \partial x_\alpha'/\partial x_\beta$ and $\pi_{\alpha\beta} = \partial x_\beta/\partial x_\alpha'$ (see equations (119)–(120)). Essentially the same equation had been written down on top of p. 23L, which suggests that Einstein began to examine the transformation properties of the ϑ-expression before he extracted eq (583) from the November tensor with the help of the ϑ-restriction.

expression on second line of equation (588) is more cryptic: "transformed back $p_{i\rho}$ $p_{k\sigma}$ $p_{l\tau}$" ("zurück transformiert $p_{i\rho}$ $p_{k\sigma}$ $p_{l\tau}$").

On the next line, following the comment "in detail" ("ausführlich"), the expression on the second line of equation (588) is expanded. The relation $p_{ij}\pi_{ik} = \delta_{jk}$ is used to simplify the resulting three terms

$$\frac{\partial g_{\rho\sigma}}{\partial x_\tau} + p_{i\rho}\frac{\partial \pi_{i\iota}}{\partial x_\tau}g_{\iota\sigma} + p_{k\sigma}\frac{\partial \pi_{k\kappa}}{\partial x_\tau}g_{\rho\kappa} \quad . \tag{589}$$

"For infinitesimal transformations" ("Für infinitesimale Transformationen"), one can replace p_{ij} in the last two terms by δ_{ij}. If in addition the summation indices are relabeled, expression (589) reduces to

$$\frac{\partial g_{\rho\sigma}}{\partial x_\tau} + \frac{\partial \pi_{\rho\alpha}}{\partial x_\tau}g_{\sigma\alpha} + \frac{\partial \pi_{\sigma\alpha}}{\partial x_\tau}g_{\rho\alpha} \,. \tag{590}$$

Substituting this result into equation (588), one arrives at

$$\frac{\partial g_{\rho\sigma}}{\partial x_\tau} = \frac{\partial g_{\rho\sigma}}{\partial x_\tau} + \frac{\partial \pi_{\rho\alpha}}{\partial x_\tau}g_{\sigma\alpha} + \frac{\partial \pi_{\sigma\alpha}}{\partial x_\tau}g_{\rho\alpha} \,. \tag{591}$$

It follows that derivatives of the metric transform as tensors under infinitesimal non-autonomous transformations if and only if the last two terms in equation (591) vanish, i.e., if and only if the matrix π_{ij} for such a transformation satisfies the equation

$$\frac{\partial \pi_{\rho\alpha}}{\partial x_\tau}g_{\sigma\alpha} + \frac{\partial \pi_{\sigma\alpha}}{\partial x_\tau}g_{\rho\alpha} = 0 \tag{592}$$

for the metric field under consideration.

Since $\vartheta_{\rho\sigma\tau}$ is essentially the symmetrized version of $\partial g_{\rho\sigma}/\partial x_\tau$, one can derive the condition for infinitesimal non-autonomous transformations under which $\vartheta_{\rho\sigma\tau}$ transforms as a tensor by adding equation (591) and the two equations obtained from it through cyclic permutation of its indices,

$$\begin{aligned} \frac{\partial g_{\sigma\tau}}{\partial x_\rho} &= \frac{\partial g_{\sigma\tau}}{\partial x_\rho} + \frac{\partial \pi_{\sigma\alpha}}{\partial x_\rho}g_{\tau\alpha} + \frac{\partial \pi_{\tau\alpha}}{\partial x_\rho}g_{\sigma\alpha} \\ \frac{\partial g_{\tau\rho}}{\partial x_\sigma} &= \frac{\partial g_{\tau\rho}}{\partial x_\sigma} + \frac{\partial \pi_{\tau\alpha}}{\partial x_\sigma}g_{\rho\alpha} + \frac{\partial \pi_{\rho\alpha}}{\partial x_\sigma}g_{\tau\alpha} . \end{aligned} \tag{593}$$

As Einstein put it: "Through addition [of the various terms coming] from all three terms [of $\vartheta_{\rho\sigma\tau}$] one obtains" ("Durch Addition aus allen drei Termen erhält man")

$$\vartheta_{\rho\sigma\tau} = \vartheta_{\rho\sigma\tau} + \frac{1}{2}\left[g_{\rho\alpha}\left(\frac{\partial\pi_{\tau\alpha}}{\partial x_\sigma} + \frac{\partial\pi_{\sigma\alpha}}{\partial x_\tau}\right) + g_{\sigma\alpha}\left(\frac{\partial\pi_{\rho\alpha}}{\partial x_\tau} + \frac{\partial\pi_{\tau\alpha}}{\partial x_\rho}\right) + g_{\tau\alpha}\left(\frac{\partial\pi_{\sigma\alpha}}{\partial x_\rho} + \frac{\partial\pi_{\rho\alpha}}{\partial x_\sigma}\right)\right] \tag{594}$$

Einstein only wrote down the right-hand side of this equation, left out the factor $1/2$ in front of the expression in square brackets, and only indicated the last two terms in this expression by a dot. For infinitesimal transformations one can use that[297]

$$\frac{\partial\pi_{\sigma\alpha}}{\partial x_\tau} = \frac{\partial\pi_{\tau\alpha}}{\partial x_\sigma}, \tag{595}$$

in which case the right-hand side of equation (594) reduces to

$$\vartheta_{\rho\sigma\tau} + \left[g_{\rho\alpha}\frac{\partial\pi_{\sigma\alpha}}{\partial x_\tau} + g_{\sigma\alpha}\frac{\partial\pi_{\tau\alpha}}{\partial x_\rho} + g_{\tau\alpha}\frac{\partial\pi_{\rho\alpha}}{\partial x_\sigma}\right]. \tag{596}$$

In the notebook, because of the factor $1/2$ omitted in equation (594), there is an extra factor 2 in front of the expression in square brackets.

It follows that the condition for infinitesimal non-autonomous transformations under which the ϑ-expression transforms as a tensor is that "the [expression in square] brackets [in expression (596)] should vanish for all combinations of $\rho\sigma\tau$" ("Die Klammer soll für alle Kombinationen von $\rho\sigma\tau$ verschwinden"). In other words, the matrix π_{ij} for (the inverse of) such transformations should satisfy the equation

$$g_{\rho\alpha}\frac{\partial\pi_{\sigma\alpha}}{\partial x_\tau} + g_{\sigma\alpha}\frac{\partial\pi_{\tau\alpha}}{\partial x_\rho} + g_{\tau\alpha}\frac{\partial\pi_{\rho\alpha}}{\partial x_\sigma} = 0 \tag{597}$$

for the metric field under consideration. The field equations extracted from the November tensor with the help of the ϑ-restriction will be invariant under all (non-autonomous) coordinate transformations that satisfy this condition.

5.5.6 *Solving the ϑ-Equation (42L–R)* 42L–R

No attempt is made in the notebook to find solutions of the condition for infinitesimal non-autonomous ϑ-transformations given in equation (597). Einstein adopted a

297 For infinitesimal transformations, one has $\partial\pi_{\sigma\alpha}/\partial x_\tau = -\partial p_{\alpha\sigma}/\partial x_\tau$. Using the definition of $p_{\alpha\sigma}$ and changing the order of differentiation, one can then write:

$$\frac{\partial\pi_{\sigma\alpha}}{\partial x_\tau} = -\frac{\partial p_{\alpha\sigma}}{\partial x_\tau} = -\frac{\partial}{\partial x_\tau}\left(\frac{\partial x_\alpha'}{\partial x_\sigma}\right) = -\frac{\partial}{\partial x_\sigma}\left(\frac{\partial x_\alpha'}{\partial x_\tau}\right) = -\frac{\partial p_{\alpha\tau}}{\partial x_\sigma} = \frac{\partial\pi_{\tau\alpha}}{\partial x_\sigma}.$$

Earlier in the notebook, Einstein had already (implicitly) used a very similar relation (see p. 10L and footnote 148).

somewhat different approach to get a sense of the range of transformations allowed by the ϑ-restriction. It is clear upon inspection of its definition (570) that the ϑ-expression vanishes for the Minkowski metric in its standard diagonal form. Any ϑ-transformation will preserve the vanishing of the ϑ-expression. Hence, if transformations to accelerated frames of reference in Minkowski spacetime are ϑ-transformations, the ϑ-expression should also vanish for the Minkowski metric expressed in the coordinates of such accelerated frames. On p. 42L, toward the end of the part starting from the other end of the notebook, Einstein set himself the task of finding the most general form of the metric field $g_{\mu\nu}$ satisfying the equation $\vartheta_{\rho\sigma\tau} = 0$, imposing the additional constraints that the metric be time-independent and that its determinant be equal to unity and suppressing one spatial dimension. Given Einstein's heuristic principles, the general solution should include the Minkowski metric in (uniformly) rotating coordinates. Unfortunately, this turns out not to be the case. What Einstein discovered instead (at the top of p. 42R), is that one does obtain a solution if one simply interchanges the covariant and contravariant components of the rotation metric. We shall call this solution the "ϑ-metric."

Einstein tried to come to terms with this tantalizing result in various ways. One approach was to derive the equations of motion for a particle moving in a gravitational field described by the ϑ-metric and to identify the various components of the ϑ-metric occurring in these equations in terms of inertial forces in rotating frames of reference just as one would for the ordinary Minkowski metric in rotating coordinates. Einstein made two attempts along these lines. In one case, he derived the equations of motion as the Euler-Lagrange equations for the Lagrangian of a particle moving in a metric field (calculations at the bottom of pp. 42R and 43LA; discussed in sec. 5.5.7). In the other, he derived the equations of motion from the energy-momentum balance between matter and gravitational field (p. 24L; discussed in sec. 5.5.9). Neither calculation produced a satisfactory result. Neither did a somewhat different approach which Einstein tried at the top of p. 43LA (discussed in sec. 5.5.8). He replaced the covariant components of the metric in the ϑ-expression by contravariant ones, ensuring that the new expression vanishes for the rotation metric without interchanging its co- and contravariant components. Einstein discovered, however, that the modified ϑ-expression could not be used to eliminate terms with unwanted second-order derivatives of the metric from the November tensor. Moreover, the new expression is mathematically ill-defined. In the end, Einstein was thus forced to give up the ϑ-restriction and the promising candidate (584) for the left-hand side of the field equations constructed with the help of it.

The precise temporal order of the calculations on pp. 23L–24L at one end of the notebook and on pp. 42L–43L at the other remains unclear. There are various indications that Einstein switched back and forth between these two sets of pages. As we already noted, for instance, attempts to interpret the ϑ-metric in terms of inertial forces in rotating frames of reference can be found both on p. 42R–43L and on p. 24L.[298] Several questions remain. Had Einstein already encountered the ϑ-expression in some other context when he used it on p. 23L to extract field equations from the November tensor? If so, it would explain why Einstein's solution of the equation $\vartheta_{\rho\sigma\tau} = 0$ occurs in a different part of the notebook. It would, however, also raise the question why Einstein originally got interested in the ϑ-expression. If the calculation on p. 23L preceded the one on p. 42R,[299] this question does not arise, but in that case it is unclear why Einstein turned over the notebook for the calculation on p. 42R instead of simply continuing his entries on p. 24L ff. Given these uncertainties, it is important to emphasize that the order in which we present these calculations may not fully reflect their temporal order.

After these introductory and cautionary remarks, we turn to the actual calculation on p. 42L. At the top of the page,[300] Einstein wrote down the components of the metric 42L
suppressing one spatial dimension[301]

$$\begin{matrix} g_{11} & g_{12} & g_{14} \\ g_{21} & g_{22} & g_{24} \\ g_{41} & g_{42} & g_{44} \end{matrix} \tag{598}$$

Einstein then set out to find the most general time-independent solution of a set of linear first-order coupled differential equations for this metric field. In modern notation, this set of equations can be written as

$$g_{(\mu\nu,\lambda)} = g_{\mu\nu,\lambda} + g_{\lambda\mu,\nu} + g_{\nu\lambda,\mu} = 0\,, \tag{599}$$

where the indices can take on the values 1, 2, and 4. Although the ϑ-expression is not explicitly mentioned on pp. 42L–43L, equation (599) can also be written as (cf. expression (570))

$$\vartheta_{\mu\nu\lambda} = 0\,. \tag{600}$$

On p. 42R, Einstein explicitly imposed the additional constraint that the determinant of the metric be equal to unity. This condition was probably part of the original prob-

---

298 A possible further connection between these two attempts to interpret the ϑ-metric in terms of inertial forces in rotating frames of reference is the derivation on p. 43LB of the equation of motion for a point mass moving in a metric field from the vanishing of the covariant divergence of the stress-energy tensor for pressureless dust (see sec. 5.5.10)

299 The occurrence in a calculation on p. 43LA of Christoffel symbols, which are otherwise absent from the part that starts from the back of the notebook, suggests that at least this calculation is later than pp. 23L–R (cf. footnote 317).

300 In the top-left corner, Einstein wrote the six independent components of this metric field arranging them in a somewhat peculiar way. The reason behind this is unclear.

301 Underneath the 44-component, he wrote "0," which suggests that Einstein assumed the g_{44}-component to vanish. This is puzzling but would be consistent with the expression written right next to the metric (598) for the determinant of the metric. Perhaps, the components of (598) refer to small deviations of the metric from the flat Minkowski metric $\mathrm{diag}(-1,-1,1)$. This interpretation would be consistent with most of the rest of the calculation on this page, though not with the calculation of the determinant of the metric right next to (598).

lem that Einstein set himself on p. 42L, as is suggested by the fact that in the top right corner there is an expression for the determinant of the metric (598),[302]

$$\begin{aligned} & g_{14}(g_{12}g_{24} - g_{22}g_{14}) \\ & + g_{24}(g_{14}g_{12} - g_{24}g_{12}). \end{aligned} \tag{601}$$

Einstein then listed the index-combinations for all independent components of the set of differential equations in equation (599)

$$\begin{array}{cccccc} \underline{111} & \underline{112} & \underline{114} & \underline{122} & \underline{124} & \underline{144} \\ \underline{222} & \underline{224} & \cancel{233} & \cancel{234} & \underline{244} & , \\ \underline{444}. & & & & & \end{array} \tag{602}$$

He explicitly wrote down the equations corresponding to these index-combinations, with the exception of the last one, a clear indication that Einstein was interested only in time-independent solutions. The equations are grouped together as follows. First, Einstein gave the 111 and 222 components of equation (599), i.e., in modern notation,

$$g_{11,\,1} = 0\,, \tag{603}$$

$$g_{22,\,2} = 0\,. \tag{604}$$

He then gave the components with two indices equal to 4, followed by the ones with one index equal to 4. In the first four of these equations, he included but then deleted terms with derivatives with respect to x_4. In the fifth, he omitted this term altogether. Einstein thus arrived at

$$g_{44,\,1} = 0\,, \tag{605}$$

$$g_{44,\,2} = 0\,, \tag{606}$$

$$2g_{14,\,1} = 0\,, \tag{607}$$

$$2g_{24,\,2} = 0\,, \tag{608}$$

$$g_{14,\,2} + g_{24,\,1} = 0\,. \tag{609}$$

Finally, he wrote down the 112 and 122 components of equation (599),

$$2g_{12,\,1} + g_{11,\,2} = 0\,, \tag{610}$$

$$2g_{12,\,2} + g_{22,\,1} = 0\,. \tag{611}$$

From equations (603) and (604) it follows that g_{11} can only be a function of x_2 and that g_{22} can only be a function of x_1. Einstein thus wrote

302 Einstein either did not finish the calculation or he set g_{44} equal to zero as he indicated in (598) (cf. footnote 301).

$$g_{11} = \varphi(x_2), \tag{612}$$

$$g_{22} = \psi(x_1). \tag{613}$$

Substituting these expressions for g_{11} and g_{22} into equations (610) and (611), he obtained

$$2g_{12,1} = -\varphi'(x_2), \tag{614}$$

$$2g_{12,2} = -\psi'(x_1). \tag{615}$$

From these last two equations it follows that g_{12} has to be linear both in x_1 and x_2. Terms in g_{12} of second-order or higher in x_1 would make $g_{12,1}$ dependent on x_1, which is contrary to equation (614); terms of second-order or higher in x_2 would likewise make $g_{12,2}$ dependent on x_2, which is contrary to equation (615). Hence, g_{12} has to be of the form

$$g_{12} = c_0 + c_1x_1 + c_2x_2 + \alpha x_1x_2, \tag{616}$$

where c_0, c_1, c_2, and α are arbitrary constants. Inserting equation (616) into equations (614) and (615), Einstein found

$$\varphi'(x_2) = -2(c_1 + \alpha x_2), \tag{617}$$

$$\psi'(x_1) = -2(c_2 + \alpha x_1) \tag{618}$$

Integrating these equations and substituting the results into equations (612) and (613), Einstein found[303]

$$g_{11} = \varphi(x_2) = -2\left(c_1x_2 + \frac{\alpha}{2}x_2^2 + \kappa''\right), \tag{619}$$

$$g_{22} = \psi(x_1) = -2\left(c_2x_1 + \frac{\alpha}{2}x_1^2 + \kappa'''\right). \tag{620}$$

Expressions for the components g_{14} and g_{24} can be found in a similar way. From equations (607) and (608) it follows that g_{14} and g_{24} can be written as

$$g_{14} = \hat{\varphi}(x_2), \tag{621}$$

$$g_{24} = \hat{\psi}(x_1), \tag{622}$$

where we introduced the notation $\hat{\varphi}$ and $\hat{\psi}$ to distinguish these functions from the functions φ and ψ above. In the notebook, no such distinction is made. Inserting equations (621) and (622) into equation (609), one finds that

$$\hat{\varphi}'(x_2) + \hat{\psi}'(x_1) = 0, \tag{623}$$

303 The constants κ'' and κ''' in these equations probably only got these designations after the introduction of the constants κ and κ' in the expressions for g_{14} and g_{24}, which occur immediately below equations (619)–(620) in the notebook (cf. equations (626)–(627) below).

which allows the separation

$$\hat{\varphi}'(x_2) = \beta\,, \tag{624}$$

$$\hat{\psi}'(x_1) = -\beta\,, \tag{625}$$

where β is a constant. The notebook has α rather than β at this point, but in subsequent equations Einstein renamed this constant β, presumably to avoid confusion with the constant α introduced in equation (616). Integrating equations (624) and (625) and substituting the results into equations (621) and (622), one finds

$$g_{14} = \hat{\varphi}(x_2) = \beta x_2 + \kappa\,, \tag{626}$$

$$g_{24} = \hat{\psi}(x_1) = -\beta x_1 + \kappa'\,. \tag{627}$$

An expression for g_{44} was not explicitly given in the notebook, but from equations (605) and (606) it immediately follows that g_{44} has to be a constant.

The most general solution of the equations $g_{(\mu\nu,\lambda)} = 0$, under the additional constraint that $g_{\mu\nu,4} = 0$, is thus given by (cf. equations (616), (619)–(620), and (626)–(627))[304]

$$g_{\mu\nu} = \begin{pmatrix} -2\left(c_1 x_2 + \frac{\alpha}{2}x_2^2 + \kappa''\right) & c_0 + c_1 x_1 + c_2 x_2 + \alpha x_1 x_2 & \beta x_2 + \kappa \\ c_0 + c_1 x_1 + c_2 x_2 + \alpha x_1 x_2 & -2\left(c_2 x_1 + \frac{\alpha}{2}x_1^2 + \kappa'''\right) & -\beta x_1 + \kappa' \\ \beta x_2 + \kappa & -\beta x_1 + \kappa' & g_{44} \end{pmatrix} \tag{628}$$

At the bottom of p. 42L Einstein drew a figure indicating rotation around the z-axis. The relation between the calculation on p. 42L and rotation becomes clear on the next page.

42R At the top of page 42R Einstein wrote down a metric—or a "g-system" ("g-Schema") as he called it here—which is obtained from equation (628) by setting the integration constants c_0, c_1, c_2, κ, and κ' equal to zero, $\kappa'' = \kappa''' = 1/2$ and $g_{44} = 1$[305]

$$g_{\mu\nu} = \begin{pmatrix} -(1+\alpha x_2^2) & \alpha x_1 x_2 & \beta x_2 \\ \alpha x_1 x_2 & -(1+\alpha x_1^2) & -\beta x_1 \\ \beta x_2 & -\beta x_1 & 1 \end{pmatrix}. \tag{629}$$

304 The general solution for the 2+1-dimensional case can trivially be turned into a particular solution for the 3+1-dimensional case by adding $g_{\mu 3} = g_{3\mu} = (0, 0, -1, 0)$

305 If $g_{\mu\nu}$ represents small deviations from a flat Minkowski metric (cf. footnote 301), one needs to set $\kappa'' = \kappa''' = 0$ and $g_{44} = 0$ instead.

This is the solution of the equation $\vartheta_{\mu\nu\lambda} = 0$ that we shall refer to as the ϑ-metric. Einstein now imposed the condition that the determinant G of this metric be equal to unity. For G he wrote[306]

$$\begin{aligned} G &= (1+\alpha x_1^2)(1+\alpha x_2^2) - \alpha\beta^2 x_1^2 x_2^2 - \alpha\beta^2 x_1^2 x_2^2 \\ &\quad + (1+\alpha x_2^2)\beta^2 x_1^2 - \alpha^2 x_1^2 x_2^2 + (1+\alpha x_1^2)\beta^2 x_2^2 \\ &= 1 + (\alpha+\beta^2)x_1^2 + (\alpha+\beta^2)x_2^2 \\ &\quad + (\alpha^2 - 2\alpha\beta^2 + \alpha\beta^2 + \alpha\beta^2 - \alpha^2)x_1^2 x_2^2. \end{aligned} \tag{630}$$

Next to this equation, Einstein wrote down the condition that this determinant be equal to unity,

$$\alpha + \beta^2 = 0\,. \tag{631}$$

Using this relation between α and β, he inverted the metric in equation (629) and wrote down the result,

$$\gamma_{\mu\nu} = \begin{pmatrix} -1 & 0 & \beta x_2 \\ 0 & -1 & -\beta x_1 \\ \beta x_2 & -\beta x_1 & 1+\alpha(x_1^2+x_2^2) \end{pmatrix}. \tag{632}$$

which he denoted by γ, next to "g-system" at the top of the page. As Einstein noted in a comment that he wrote next to these expressions of the ϑ-metric in its covariant and contravariant form, "the system of the γ's for a rotating body [is] identical to the g-system given here" ("Schema der γ für rotierenden Körper mit nebenstehendem g-Schema identisch"). This is an intriguing result. The ϑ-restriction does not allow the Minkowski metric in rotating coordinates, but it does allow the ϑ-metric, which is closely related to it. Is that enough to satisfy Einstein's heuristic requirements? Can the ϑ-restriction be modified in such a way that it does allow the rotation metric without the need to switch its co- and contravariant components? These are the questions that are behind the calculations on pp. 42R, 43LA, and 24L that will be discussed in secs. 5.5.7–5.5.10.

Before Einstein began his closer examination of the ϑ-metric, he briefly considered another special case of equation (628). He drew a horizontal line and wrote the non-vanishing components of the metric

306 With the help of the fully anti-symmetric Levi-Civita tensor ε_{ijk} —which is equal to 1 for every even permutation of 1, 2, 3 (or, rather, 4 in this case), equal to -1 for every odd permutation, and equal to 0 otherwise—the determinant G can be written as $G = \varepsilon_{ijk} g_{1i} g_{2j} g_{4k}$. The three terms on the first line of equation (630) correspond to the even permutations 124, 421, and 241; the three terms on the second line to the odd permutations 142, 214, 421. This way of evaluating the determinant is known as Sarrus' rule.

$$\begin{matrix} -1-2c_1x_2 & c_1x_1+c_2x_2 & 0 \\ c_1x_1+c_2x_2 & -1-2c_2x_1 & 0 \\ 0 & 0 & 1. \end{matrix} \tag{633}$$

This metric is obtained from equation (628) by setting the integration constants c_0, α, β, κ, and κ' equal to zero, setting $\kappa'' = \kappa''' = 1/2$ and $g_{44} = 1$.[307] This solution, however, does not satisfy one of Einstein's additional constraints. As he wrote next to the metric (633), "the determinant [of this metric field] is not 1" ("Determinante ist nicht 1"). Einstein then drew another horizontal line and began his closer examination of the tantalizing ϑ-metric of equations (629) and (632).

5.5.7 *Reconciling the ϑ-Metric and Rotation (I): Identifying Coriolis and Centrifugal Forces in the Geodesic Equation (42R, 43LA)*

42R On the bottom half of p. 42R, under the second horizontal line, Einstein first gave a short derivation, similar to the one he gave on pp. 12L–R (see sec. 4.5.6), of the Minkowski metric in rotating coordinates. As in the derivation of the ϑ-metric, with which he wanted to compare the rotation metric, he suppressed one of the spatial dimensions.

Consider a Cartesian coordinate system $x_\mu = (x, y, t)$ in $2+1$-dimensional Minkowski spacetime which is rotating clockwise with angular frequency ω with respect to another Cartesian coordinate system $x_\mu{}' = (x', y', t')$ in which the metric has its usual diagonal form:

$$ds^2 = \eta_{\mu\nu}dx'_\mu dx'_\nu = (\mathbf{v}'^2 - 1)dt'^2. \tag{634}$$

In this equation, $\eta_{\mu\nu} = \text{diag}(1, 1, -1)$, $c = 1$ (in accordance with Einstein's conventions at this point), and the components of $\mathbf{v}'$ are $v'_i = dx'_i/dt' = \dot{x}'_i$ ($i = 1, 2$). The relation between the velocity with respect to the non-rotating frame and the velocity $\mathbf{v}$ with respect to the rotating frame is given by (cf. equation (352)):

$$\mathbf{v}' = \mathbf{v} + \vec{\omega} \times \mathbf{r} = (\dot{x} + \omega y, \dot{y} - \omega x), \tag{635}$$

where $\vec{\omega} \times \mathbf{r}$ is the cross-product of the vectors $\vec{\omega} = (0, 0, -\omega)$ ($-\omega$ because the rotation is clockwise) and $\mathbf{r} = (x, y, 0)$ in three dimensions. Taking the square of equation (635), one finds that

$$\begin{aligned} \mathbf{v}'^2 &= (\dot{x} + \omega y)^2 + (\dot{y} - \omega x)^2 \\ &= \dot{x}^2 + \dot{y}^2 + 2\omega y\dot{x} - 2\omega x\dot{y} + \omega^2 r^2. \end{aligned} \tag{636}$$

The expressions following the equality signs in equation (636) are actually the only ones Einstein explicitly wrote down before giving the matrix for the rotation metric.

307 Again, this metric could be obtained by setting $\kappa'' = \kappa''' = 0$ and $g_{44} = 0$ in the case that the $g_{\mu\nu}$ were deviations. Taken together both special cases exhaust the general solution (up to constants).

As we mentioned above, he had gone through essentially the same derivation on pp. 12L–R. Inserting equation (636) along with $dt' = dt$ into equation (634), one finds that the Minkowski line element in rotating coordinates is given by

$$ds^2 = dx^2 + dy^2 + 2\omega y dx dt - 2\omega x dy dt - (1 - \omega^2 r^2) dt^2 , \tag{637}$$

from which one can read off the components of the metric,

$$g_{\mu\nu} = \begin{pmatrix} 1 & 0 & \omega y \\ 0 & 1 & -\omega x \\ \omega y & -\omega x & -1 + \omega^2 r^2 \end{pmatrix} . \tag{638}$$

Einstein read off the components of the metric from the last line of equation (636), which is probably why he omitted the term -1 in g_{44}. As on pp. 12L–R, he also ended up with additional factors of 2 in g_{14} and g_{24}.[308] When the constant β in equation (629) is set equal to the angular frequency ω (in which case $\alpha = -\omega^2$ according to equation (631)), the contravariant form of the ϑ-metric does indeed turn into the covariant form of the rotation metric in equation (638), as was noted in the top-right corner of p. 42R.[309]

As he had done earlier when he was faced with the conflict between the modified weak-field equations of p. 21R and the static metric (see the discussion in sec. 5.4.4), Einstein turned to particle dynamics to see whether his mathematical results could be given a physically sensible interpretation. In this case, he apparently wanted to check whether the components of the ϑ-metric and their derivatives can be given the same sort of physical interpretation in terms of inertial forces as the components of the usual rotation metric and their derivatives. Although the relevant calculations—at the bottom of p. 42R and again at the bottom of p. 43LA—break off after just a few lines, it seems to be clear that this was their purpose.

Einstein inserted the ϑ-metric in its covariant form (see equation (629)) into the Lagrangian

$$H = \sqrt{\frac{ds^2}{dt^2}} = \sqrt{g_{\mu\nu}\dot{x}_\mu \dot{x}_\nu} \tag{639}$$

for a point mass moving in a given metric field. The result is that

$$H = \sqrt{-\dot{x}^2 - \dot{y}^2 + 2\beta y\dot{x} - 2\beta x\dot{y} \underbrace{- \alpha y^2\dot{x}^2 - \alpha x^2\dot{y}^2 + 2\alpha xy\dot{x}\dot{y}}_{-\alpha(x\dot{y} - y\dot{x})^2} + 1} . \tag{640}$$

Einstein then wrote down the variational principle

308 Einstein made the same mistake on pp. 12L–R (see footnote 176) and in the Einstein-Besso manuscript (CPAE 4, Doc. 14, pp. [41–42]).

309 The two equations are still not completely identical because Einstein wrote the flat Minkowski metric as $\mathrm{diag}(-1, -1, 1)$ in equation (629) and as $\mathrm{diag}(1, 1, -1)$ in equation (638).

$$\delta\int H dt = 0 , \tag{641}$$

and went through a quick derivation of the x-component of the corresponding Euler-Lagrange equations, writing

$$\int\left(\frac{\partial H}{\partial \dot{x}}\delta\dot{x} + \frac{\partial H}{\partial x}\delta x\right)dt = 0 ,\text{[310]} \tag{642}$$

which, upon partial integration, gives

$$-\frac{d}{dt}\left(\frac{\partial H}{\partial \dot{x}}\right) + \frac{\partial H}{\partial x} = 0 . \tag{643}$$

Einstein now began to compute the first term of this Euler-Lagrange equation for the Lagrangian in equation (640). The derivative of H with respect to $\dot{x}$ is given by

$$\frac{\partial H}{\partial \dot{x}} = \frac{-\dot{x} + \beta y + \alpha(x\dot{y} - y\dot{x})y}{\sqrt{-\dot{x}^2 - \dot{y}^2 + 2\beta y\dot{x} - 2\beta x\dot{y} - \alpha(x\dot{y} - y\dot{x})^2 + 1}} . \tag{644}$$

In the notebook, there is a deleted factor of 2 in front of all three terms in the numerator and the denominator was indicated only by a square root sign. On the next line, Einstein took the time derivative of equation (644) in its uncorrected form

$$\frac{d}{dt}\left(\frac{\partial H}{\partial \dot{x}}\right) = -2\ddot{x} + \beta\dot{y} + 2\alpha y(x\ddot{y} - y\ddot{x}) + 2\alpha\dot{y}(x\dot{y} - y\dot{x}) , \tag{645}$$

i.e., with the extra factors of 2 and without the denominator. On the last line of p. 42R, Einstein wrote down the second term of the Euler-Lagrange equation (643) as well, but did not actually evaluate it for the Lagrangian under consideration. The inclusion of the denominator in equation (644), which Einstein had originally omitted, considerably complicates the Euler-Lagrange equations, which may well be why Einstein gave up on this calculation, at least for the time being.

At the top of the next page, p. 43LA, he tried to resolve the apparent conflict between the ϑ-metric and the rotation metric in a different manner (to be discussed in sec. 5.5.8). This new approach, however, turned out not to be viable, and at the bottom of p. 43LA Einstein briefly returned to the approach he had abandoned at the bottom of p. 42R.

43LA Under the heading "dynamics in a symmetric static rotational field" ("Dynamik im symmetrischen statischen Rotationsfeld"), Einstein once again considered the motion of a point mass in the ϑ-metric. The reason he explicitly referred to this case as "symmetric" and "static" may have been that the ϑ-metric is a time-independent solution of an equation in which the fully symmetrized derivative of the metric is set equal to zero (see equation (599)). Using the relation $\alpha + \beta^2 = 0$, which ensures that the determinant of the ϑ-metric is equal to unity (see equation (631)), Einstein was able to write the Lagrangian H of equation (640) more compactly as[311]

310 In the notebook the differential dt was omitted.

$$H = \sqrt{1 - \dot{x}^2 - \dot{y}^2 - 2\beta(x\dot{y} - y\dot{x}) + \beta^2(x\dot{y} - y\dot{x})^2}$$
$$= \sqrt{[1 - \beta(x\dot{y} - y\dot{x})]^2 - \dot{x}^2 - \dot{y}^2}. \tag{646}$$

However, he only wrote down one term of the Euler-Lagrange equations,

$$\frac{\partial H}{\partial x} = \frac{1}{H} \cdot -\beta(x\dot{y} - y\dot{x})\dot{y}\,, \tag{647}$$

before once again breaking off this calculation. This attempt to give physical meaning to the ϑ-metric thus remained inconclusive. On p. 24L, Einstein made another attempt along these lines (see sec. 5.5.9), but first we shall discuss the calculation at the top of p. 43LA.

5.5.8 *Reconciling the ϑ-Metric and Rotation (II): Trying to Construct a Contravariant Version of the ϑ-Expression (43LA)* 43LA

At the top of p. 43LA,[312] Einstein introduced a variant of the ϑ-expression introduced on p. 23L, replacing covariant components of the metric by contravariant ones.[313] Since the original ϑ-expression vanishes for the ϑ-metric, i.e., the rotation metric with its co- and contravariant components switched, this modified ϑ-expression vanishes for the rotation metric itself. The modified ϑ-restriction, i.e., the restriction to unimodular transformations under which the modified ϑ-expression transforms as a tensor, thus allows transformations to rotating coordinates in the important special case of Minkowski spacetime. However, Einstein found that, unlike the original ϑ-restriction, the modified ϑ-restriction could not be used to eliminate terms with unwanted second-order derivatives of the metric from the November tensor. At that point, he abandoned the modified ϑ-expression. He may also have come to realize that the expression is mathematically ill-defined (because of the way in which it mixes co- and contravariant components).

At the top of the page Einstein wrote down the expression

$$g_{ik}\left(\frac{\partial \gamma_{ik}}{\partial x_l} + \frac{\partial \gamma_{kl}}{\partial x_i} + \frac{\partial \gamma_{li}}{\partial x_k}\right). \tag{648}$$

---

311 In the notebook, the last term of the first line has $\dot{x}\dot{y}$ instead of $x\dot{y}$.

312 This is the last page of the part that starts from the back of the notebook (pp. 32L–43L). It contains entries starting from the top and from the bottom of the page (transcribed as 43LA and 43LB respectively), thus suggesting that this is the place where the two parts of the notebook meet. There are, however, six blank pages between p. 43L and p. 31L, the last page of the part that starts from the front of the notebook. The calculation that starts from the bottom of p. 43L is also not a continuation of p. 31L. In this brief calculation, Einstein derived the equation of motion of a particle in a metric field by integrating the energy-momentum balance between matter and gravitational field over the volume of the particle (for discussion see sec. 5.5.10).

313 The fragmentary calculation at the top of p. 23R could be related to the investigation of the modified ϑ-expression on this page (cf. the discussion following equation (585) in sec. 5.5.4 for a different interpretation of this calculation).

The expression in parentheses is obtained (down to the labeling of the indices) by substituting $\gamma_{\mu\nu}$ for $g_{\mu\nu}$ in definition (570) of the ϑ-expression. Farther down on the page, Einstein introduced the quantity $t_{\lambda\alpha\kappa}$ which appears to be defined as ($1/2$ times) this expression

$$t_{\lambda\alpha\kappa} \equiv \frac{1}{2}\left(\frac{\partial\gamma_{\lambda\alpha}}{\partial x_\kappa} + \frac{\partial\gamma_{\alpha\kappa}}{\partial x_\lambda} + \frac{\partial\gamma_{\kappa\lambda}}{\partial x_\alpha}\right). \tag{649}$$

Given the convention in the notebook of using Greek and Latin characters to denote covariant and contravariant quantities, respectively,[314] the notation indicates that $t_{\lambda\alpha\kappa}$ is a contravariant version of ϑ_{ikl}. Solutions of the equation $t_{\lambda\alpha\kappa} = 0$ are obtained simply by interchanging co- and contravariant components in solutions of $\vartheta_{ikl} = 0$. Since the ϑ-metric is a solution of $\vartheta_{ikl} = 0$ (see equations (629) and (632)), the rotating metric is a solution of $t_{\lambda\alpha\kappa} = 0$. The quantity $t_{\lambda\alpha\kappa}$, however, is mathematically ill-defined. Covariant indices in one term occur as contravariant ones in another, and there are summations over pairs of covariant indices. Next to expression (648), Einstein nonetheless wrote "Tensor," presumably to indicate that he wanted to consider a variant of the ϑ-restriction, i.e., a restriction to (unimodular) transformations under which $t_{\lambda\alpha\kappa}$ transforms as a tensor.[315]

At the beginning of the calculation on p. 23L, the ϑ-expression was used to rewrite the Christoffel symbols (see equations (572)–(573)). If $t_{\lambda\alpha\kappa}$, the "contravariant" version of the ϑ-expression, is going to be used in a similar manner, one first needs to lower its indices. Einstein indeed wrote down the expression

$$g_{i\alpha}g_{k\beta}\left(\frac{\partial\gamma_{ik}}{\partial x_l} + \frac{\partial\gamma_{kl}}{\partial x_i} + \frac{\partial\gamma_{li}}{\partial x_k}\right), \tag{650}$$

which again is mathematically ill-defined, and wrote next to it that this is a covariant or "plane tensor" ("Ebenentensor"[316]). Einstein rewrote this expression in such a way that, just as the Christoffel symbols, it contains only derivatives of the covariant components of the metric

---

314 Cf., e.g., on p. 24L where T_{ik} and Θ_{ik} represent the stress-energy tensor for matter in its contravariant and covariant form, respectively. Note that this convention is just the opposite of the one adopted in (Einstein and Grossmann 1913), where all contravariant quantities are indicated by Greek and all covariant ones by Latin characters.

315 Unfortunately, this interpretation does not explain why, as Einstein noted on the next line, (minus) the ill-defined second term of (648), $\sum\gamma_{kl}\frac{\partial g_{ik}}{\partial x_i}$, should be a vector. Maybe Einstein meant (minus) the unproblematic first term, $\sum\gamma_{ik}\frac{\partial g_{ik}}{\partial x_l}$, which is equal to $\frac{\partial\log(-g)}{\partial x_l}$ and hence a vector under all unimodular transformations.

316 The term "Ebenentensor" also appears on pp. 17L, 17R, 19L, and 24R. The term "Ebenenvektor" appears on p. 13L. Cf. the discussion of this terminology following equation (453) above.

$$-\frac{\partial g_{\alpha\beta}}{\partial x_l} - \gamma_{\kappa l} g_{i\alpha} \frac{\partial g_{\kappa\beta}}{\partial x_i} - \gamma_{li} g_{\kappa\beta} \frac{\partial g_{i\alpha}}{\partial x_\kappa} . \tag{651}$$

He then checked whether this modified ϑ-expression can be used to eliminate terms with unwanted second-order derivatives of the metric from the November tensor. He began by writing down the term in the November tensor containing second-order derivatives[317] (cf. equation (544))[318]

$$\frac{\partial}{\partial x_l}\left(\gamma_{l\alpha}\begin{bmatrix} i & \kappa \\ & \alpha \end{bmatrix}\right) . \tag{652}$$

Using the definition of the Christoffel symbol, he rewrote this term as[319]

$$\frac{\partial \gamma_{l\alpha}}{\partial x_l}\begin{bmatrix} i & \kappa \\ & \alpha \end{bmatrix} + \frac{1}{2}\gamma_{l\alpha}\left(\frac{\partial^2 g_{i\alpha}}{\partial x_l \partial x_\kappa} + \frac{\partial^2 g_{\kappa\alpha}}{\partial x_l \partial x_i} - \frac{\partial^2 g_{i\kappa}}{\partial x_l \partial x_\alpha}\right) . \tag{653}$$

Einstein deleted this expression and made a fresh start on the next line, writing the expression in parentheses in expression (652) as[320]

$$\frac{1}{2}\gamma_{\lambda\alpha}\left(\frac{\partial g_{i\alpha}}{\partial x_\kappa} + \frac{\partial g_{\kappa\alpha}}{\partial x_i} - \frac{\partial g_{i\kappa}}{\partial x_\alpha}\right) . \tag{654}$$

If this expression is inserted into expression (652), the first two terms give rise to unwanted second-order derivatives of the metric. Following the strategy on p. 23L, one could try to eliminate these terms by absorbing them into a quantity involving $t_{\lambda\alpha\kappa}$, transforming as a tensor under transformations under which $t_{\lambda\alpha\kappa}$ itself transforms as a tensor. One can then subtract this quantity from the November tensor without losing invariance under this restricted class of transformations. This appears to be the rationale behind the last two lines of this calculation. First, Einstein underlined the first two terms in expression (654) and rewrote them as

317 This is the first and only occurrence of the Christoffel symbol in the part that starts from the end of the notebook. In the part that starts from the beginning of the notebook, the Christoffel symbols are first introduced on p. 14L.

318 The notebook originally had x_α instead of x_l.

319 In the notebook, the Christoffel symbol is indicated only by square brackets and is multiplied by $\frac{\partial \gamma_{\lambda\alpha}}{\partial x_\alpha}$ instead of $\frac{\partial \gamma_{l\alpha}}{\partial x_l}$. The first two occurrences of the index l in the second term were originally λ's.

320 Einstein at this point changed his notation for one of the summation indices in expression (652) from l to λ.

$$-\frac{1}{2}\left(g_{i\alpha}\frac{\partial\gamma_{\lambda\alpha}}{\partial x_\kappa}+g_{\kappa\alpha}\frac{\partial\gamma_{\lambda\alpha}}{\partial x_i}\right). \qquad (655)$$

Underneath this expression he wrote[321]

$$t_{\lambda\alpha\kappa}-\frac{\partial\gamma_{\kappa\alpha}}{\partial x_\lambda}-\frac{\partial\gamma_{\lambda\kappa}}{\partial x_\alpha}, \qquad (656)$$

On the assumption that Einstein did indeed define $t_{\lambda\alpha\kappa}$ as in equation (649), this last expression is equal to $\partial\gamma_{\lambda\alpha}/\partial x_\kappa$, which is part of the first term in expression (655). At this point, it apparently became clear to Einstein that the modified ϑ-expression could not be used to eliminate unwanted second-order derivative terms from the November tensor. Einstein may also have come to realize that $t_{\lambda\alpha\kappa}$ is mathematically ill-defined. In any case, Einstein abandoned this attempt to modify the ϑ-restriction to ensure that it would include transformations to rotating coordinates in Minkowski spacetime. As was already discussed at the end of sec. 5.5.7, he briefly returned to the approach he had tried on p. 42R before giving up on that approach as well.

24L 5.5.9 *Reconciling the ϑ-Metric and Rotation (III): Identifying the Centrifugal Force in the Energy-Momentum Balance (24L)*

On p. 24L, Einstein made yet another attempt to give physical meaning to the components of the ϑ-metric. He checked whether the force on a particle at rest in the gravitational field described by the ϑ-metric can be interpreted as the centrifugal force on a particle at rest in a rotating frame of reference. This strategy is similar to the one behind the aborted calculations at the bottom of pp. 42R and 43LA (see sec. 5.5.7). The new element is that in order to find the equations of motion and the expression for the force on the particle Einstein now substituted the ϑ-metric into the energy-momentum balance between matter and gravitational field rather than into the Lagrangian for a point particle in a metric field.

The energy-momentum balance can be written as the vanishing of the covariant divergence of either the contravariant or the covariant stress-energy tensor. If the Minkowski metric in rotating coordinates is inserted into the equation which sets the covariant divergence of the *contravariant* stress-energy tensor equal to zero, one readily establishes that a term, which can be interpreted as the force on a particle at rest in this coordinate system, is equal to the usual centrifugal force (plus correction terms of higher order in the angular frequency of the rotating coordinate system). What Einstein tried to do on p. 24L was to check whether one can establish an analogous result if the ϑ-metric, i.e., the rotation metric with its co- and contravariant components switched, is inserted into the equation which sets the covariant divergence of the *covariant* stress-energy tensor equal to zero. The result of Einstein's calculation

321 In the notebook, the quantity $t_{\lambda\alpha\kappa}$ was actually written as $t^c_{\lambda\alpha\kappa}$. It is unclear what the letter 'c' stands for.

on p. 24L suggests that one can. The calculation, however, is in error. If the errors are corrected, one sees that one cannot.

At the top of p. 24L, Einstein wrote down an equation which expresses the vanishing of the contravariant stress-energy tensor, denoted by $T_{\mu\nu}$,[322] in the special case that the determinant of the metric is equal to unity,

$$\sum \frac{\partial}{\partial x_n}(g_{m\nu}T_{\nu n}) - \frac{1}{2}\frac{\partial g_{\mu\nu}}{\partial x_m}T_{\mu\nu} = 0 . \tag{657}$$

This equation is obtained (down to the numbering of the indices) from equation (71) derived on p. 5R by setting $\sqrt{G} = 1$. On p. 5R, as in several subsequent publications,[323] this equation was derived as a generalization of the equations of motion that follow from the variational principle $\delta\int(ds/dt)dt = 0$ for one particle to a cloud of pressureless dust described by the stress-energy tensor

$$T_{\mu\nu} = \rho_0 \frac{dx_\mu}{d\tau}\frac{dx_\nu}{d\tau} , \tag{658}$$

which is explicitly given at the top of p. 24L as well. On p. 43LB, Einstein went through this derivation in reverse, showing that the equation of motion for one particle can be obtained by integrating equation (657), with $T_{\mu\nu}$ given by equation (658), over the volume of the particle. This calculation will be discussed in the next subsection (sec. 5.5.10). Einstein could thus look upon equation (657) as giving the equation of motion of a particle in a metric field with a determinant equal to unity. In particular, he could look upon the second term,

$$-\frac{1}{2}\frac{\partial g_{\mu\nu}}{\partial x_m}T_{\mu\nu} , \tag{659}$$

as giving the (density of the) force experienced by a particle in a metric field.[324]

Consider a particle at rest with respect to a rotating coordinate system in Minkowski spacetime. In that case, the stress-energy tensor in equation (658) reduces to

$$T_{\mu\nu} = \mathrm{diag}(0, 0, 0, \rho_0/g_{44}) . \tag{660}$$

For counterclockwise rotation around the z-axis with angular frequency ω, the metric is given by

---

322 We want to remind the reader that the convention used here to distinguish covariant and contravariant quantities is the opposite of the one adopted in (Einstein and Grossmann 1913), where all contravariant quantities are indicated by Greek and all covariant ones by Latin characters.

323 (Einstein and Grossmann 1913, sec. 4), (Einstein 1913, sec. 5).

324 The general interpretation of the second term in equation (657) is that it is "an expression for the effects which are transferred from the gravitational field to the material process [as described by the stress-energy tensor]" ("ein Ausdruck für die Wirkungen, welche vom Schwerefelde auf den materiellen Vorgang übertragen werden;" Einstein and Grossmann 1913, 11). See also the discussion of p. 19R in sec. 5.4.2 (especially expression (482) and equation (483)).

$$g_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & \omega y \\ 0 & -1 & 0 & -\omega x \\ 0 & 0 & -1 & 0 \\ \omega y & -\omega x & 0 & 1-\omega^2(x^2+y^2) \end{pmatrix}, \tag{661}$$

Inserting equations (660) and (661) into the $m = 1$ component of expression (659) for the force on the particle, one arrives at

$$-\frac{1}{2}g_{44,\,1}T_{44} = \frac{\rho_0}{g_{44}}\omega^2 x = \rho_0\omega^2 x + O(\omega^4), \tag{662}$$

which, when terms of order ω^4 are neglected, is the x-component of the centrifugal force in ordinary Newtonian theory.

On p. 24L, Einstein performed a variant of this calculation starting from the vanishing of the covariant divergence of the covariant rather than the contravariant stress-energy tensor. Einstein wrote the contravariant $T_{\mu\nu}$ in terms of the covariant $\Theta_{\mu\nu}$,

$$T_{\mu\nu} = \gamma_{\mu\alpha}\gamma_{\nu\beta}\Theta_{\alpha\beta}, \tag{663}$$

and substituted this expression into to equation (657), which then turns into

$$\sum\frac{\partial}{\partial x_n}(\gamma_{n\alpha}\Theta_{\alpha m}) + \frac{1}{2}\frac{\partial\gamma_{\alpha\beta}}{\partial x_m}\Theta_{\alpha\beta} = 0. \tag{664}$$

For the second term Einstein used the result of an auxiliary calculation which appears next to equation (664),

$$\frac{\partial g_{\mu\nu}}{\partial x_m}\gamma_{\mu\alpha}\gamma_{\nu\beta}\Theta_{\alpha\beta} = -g_{\mu\nu}\frac{\partial\gamma_{\mu\alpha}}{\partial x_m}\gamma_{\nu\beta}\Theta_{\alpha\beta} = -\frac{\partial\gamma_{\alpha\beta}}{\partial x_m}\Theta_{\alpha\beta}. \tag{665}$$

He then wrote down the covariant version of the stress-energy tensor for pressureless dust in equation (658)

$$\Theta_{\alpha\beta} = \rho_0 g_{\mu\alpha}g_{\nu\beta}\frac{dx_\mu}{d\tau}\frac{dx_\nu}{d\tau}. \tag{666}$$

Under the heading "Force acting on material point at rest $n = 1$" ("Kraft auf ruhenden materiellen Punkt $n = 1$") he then wrote[325]

---

325 In the notebook only the 1β - and the 44 -components of $\Theta_{\alpha\beta}$ were written down.

$$\begin{array}{ccc} \dfrac{\partial\gamma_{\alpha\beta}}{\partial x_1} & \Theta_{\alpha\beta} & \dfrac{dx}{d\tau} \\[2ex] \begin{matrix} 0 & 0 & 0 \\ 0 & 0 & \beta \\ 0 & \beta & 2\alpha x \end{matrix} & \begin{matrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \rho_0 \end{matrix} & 0\ 0\ 0\ \dfrac{1}{\sqrt{g_{44}}} \end{array} \tag{667}$$

The calculation for the ϑ-metric that Einstein indicated in this manner is completely analogous to the calculation for the rotation metric in equations (659)–(662). Einstein interpreted the second term in equation (664),

$$\frac{1}{2}\frac{\partial\gamma_{\alpha\beta}}{\partial x_m}\Theta_{\alpha\beta}\,, \tag{668}$$

with $\Theta_{\mu\nu}$ given by equation (666), as the (density of the) force experienced by a particle in a metric field. Whereas expression (659) for the force density was a contraction of the *covariant* metric and the contravariant stress-energy tensor, expression (668) is a contraction of the *contravariant* metric and the covariant stress-energy tensor. Since the contravariant components of the ϑ-metric are equal to the covariant components of the rotation metric, expression (668) would give the same result for a particle at rest in the field of the ϑ-metric as expression (659) for a particle at rest in a rotating frame in Minkowski spacetime, *if only* the components of $\Theta_{\mu\nu}$ in the former case were equal to those of $T_{\mu\nu}$ in the latter. Contrary to what is suggested by the expression for $\Theta_{\mu\nu}$ in the expressions (667), however, this last condition does not hold.

For a particle at rest with respect to a given coordinate system, as Einstein explicitly wrote down (see the expressions (667)),

$$\frac{dx_\mu}{d\tau} = \left(0, 0, 0, \frac{1}{\sqrt{g_{44}}}\right), \tag{669}$$

and equation (666) reduces to

$$\Theta_{\alpha\beta} = \rho_0\frac{g_{4\alpha}g_{4\beta}}{g_{44}}\,. \tag{670}$$

The covariant components of the ϑ-metric are given by (cf. equation (629))[326]

326 In the notebook, both on p. 42R and on p. 24L, the ϑ-metric is given for the $2+1$-dimensional case. The generalization to the $3+1$-dimensional case is trivial (cf. footnote 304)

$$g_{\mu\nu} = \begin{pmatrix} -(1+\alpha x_2^2) & \alpha x_1 x_2 & 0 & \beta x_2 \\ \alpha x_1 x_2 & -(1+\alpha x_1^2) & 0 & -\beta x_1 \\ 0 & 0 & -1 & 0 \\ \beta x_2 & -\beta x_1 & 0 & 1 \end{pmatrix}, \tag{671}$$

the contravariant ones by (cf. equation (632))

$$\gamma_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & \beta x_2 \\ 0 & -1 & 0 & -\beta x_1 \\ 0 & 0 & -1 & 0 \\ \beta x_2 & -\beta x_1 & 0 & 1+\alpha(x_1^2+x_2^2) \end{pmatrix}, \tag{672}$$

where $\alpha = -\beta^2$ (see equation (631)). Taking the derivative of equation (672) with respect to x_1, one arrives at

$$\frac{\partial\gamma_{\alpha\beta}}{\partial x_1} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\beta \\ 0 & 0 & 0 & 0 \\ 0 & -\beta & 0 & 2\alpha x_1 \end{pmatrix}. \tag{673}$$

The notebook has β instead of $-\beta$ (see the expressions (667)), but this is only a minor discrepancy. Inserting equation (671) into equation (670), however, one immediately notices that $\Theta_{\mu\nu}$ will be considerably more complicated in this case than $\mathrm{diag}(0, 0, 0, \rho_0)$ given in the expressions (667) in the notebook. Inserting the simple expression $\Theta_{\mu\nu} = \mathrm{diag}(0, 0, 0, \rho_0)$ and equation (673) into the $m = 1$ component[327] of expression (668), one arrives at

$$\frac{1}{2}\frac{\partial\gamma_{\alpha\beta}}{\partial x_1}\Theta_{\alpha\beta} = \frac{1}{2}\frac{\partial\gamma_{44}}{\partial x_1}\Theta_{44} = \rho_0\alpha x_1. \tag{674}$$

which is equal to the centrifugal force in Newtonian theory if the identification $\alpha = \omega^2$ is made. This would mean that the ϑ-metric can be interpreted in terms of centrifugal forces just as the rotation metric (cf. equations (659)–(662)). However, when Einstein's expression for $\Theta_{\alpha\beta}$ is corrected, there will be additional terms which spoil this physical interpretation of the ϑ-metric.

It is not entirely clear what conclusion Einstein drew from his calculation on p. 24L. It would seem that initially he felt that he could recover the correct expression for the centrifugal force with the ϑ-metric as well as with the rotation metric. On the

327 In the notebook, the free index in the second term in equation (664) was originally n instead of m. This explain why the header above the expressions (667) in the notebook has "$n = 1$" instead of $m = 1$.

next page, however, he made a fresh start, abandoning the idea of the ϑ-restriction that had led him to consider the ϑ-metric in the first place. This suggests that eventually Einstein came to realize that the calculation on p. 24L was in error and that his third attempt at reconciling the ϑ-metric with his heuristic requirements concerning rotation, like the first two (on pp. 42R–43LA), had failed.

The November tensor, like the Ricci tensor from which it was extracted under the restriction to unimodular coordinate transformations, thus failed to yield acceptable candidates for the left-hand side of the field equations. Einstein had found two different coordinate restrictions, the Hertz restriction and the ϑ-restriction, with which terms containing unwanted second-order derivatives of the metric can be eliminated from the November tensor. He had also discovered, however, that both coordinate restrictions rule out transformations to accelerated frames of reference in the important special case of Minkowski spacetime.

5.5.10 *Relating Attempts (I) and (III) to Reconcile the ϑ-Metric and Rotation: from the Energy-Momentum Balance to the Geodesic Equation (43LB)* 43LB

In a brief calculation starting on p. 43LB, Einstein showed that the vanishing of the covariant divergence of the stress-energy tensor for pressureless dust implies the equations of motion for a point particle in a metric field. On p. 5R, Einstein had proved the converse of this implication. The derivation on p. 43LB essentially goes through this earlier derivation in reverse.[328] The calculation may be connected to Einstein's attempts to interpret the components of the ϑ-metric (see equation (629)) in terms of the inertial forces of rotation. On p. 42R and p. 43LA he tried to do so starting from the equations of motion (see sec. 5.5.7), whereas on p. 24R he started from the energy-momentum balance between matter and gravitational field (see sec. 5.5.9). The point of Einstein's calculation on p. 43LB may have been to reassure himself that these two approaches are equivalent.

The calculation starts from equation (71) of p. 5R for the special case that $\sqrt{G} = 1$:

$$\frac{\partial}{\partial x_n}(g_{m\nu}T_{\nu n}) - \frac{1}{2}\sum\frac{\partial g_{\mu\nu}}{\partial x_m}T_{\mu\nu} = 0 \tag{675}$$

(cf. equation (657) on p. 24L). Next to this equation, Einstein drew a line and wrote $\int dxdydz$ to indicate that he wanted to integrate this equation over three-dimensional space. When the first term of the expression on the left-hand side of equation (675) is integrated over all of space, the first three terms of the summation over n vanish on account of Gauss' theorem (and suitable assumptions about $T_{\mu\nu}$), while the time

328 The same pattern can be found in (Einstein and Grossmann 1913, sec. 4). First, Einstein derives the energy-momentum balance between matter and gravitational field from the equations of motion, closely following the calculation on p 5R; then he mentions that the latter can be recovered from the former by "integration over the filament of the flow" ("Integration über Stromfaden").

derivative in the $n = 4$ term commutes with taking the integral. Using x_t and $T_{\nu t}$ for x_4 and $T_{\nu 4}$, respectively, Einstein could thus write the integral as:

$$\frac{\partial}{\partial x_t}\left(\int g_{\mu\nu} T_{\nu t} dxdydz\right) - \frac{1}{2}\int \frac{\partial g_{\mu\nu}}{\partial x_m} T_{\mu\nu} dxdydz\,. \tag{676}$$

He then substituted

$$T_{\mu\nu} = \rho \frac{dx_\mu}{ds}\frac{dx_\nu}{ds}\,, \tag{677}$$

the stress-energy tensor for pressureless dust, and the relation

$$dxdydz = \frac{V}{\sqrt{G}}\frac{ds}{dt} \tag{678}$$

between the volume $dxdydz$ in the coordinates used and the rest volume $V = d\xi d\eta d\zeta$ of the particles described by $T_{\mu\nu}$. This relation follows from the relation $\sqrt{G} \cdot dxdydzdt = d\xi d\eta d\zeta ds$ written to the far right of equations (675)–(676) in the notebook. Using furthermore that the density ρ is non-vanishing only inside the filament representing the flow of matter, Einstein arrived at

$$\frac{d}{dt}\left(g_{\mu\nu}\rho \frac{dx_\nu}{ds}\frac{dt}{ds} \cdot \frac{ds}{dt}\frac{V}{\sqrt{G}}\right) - \frac{1}{2}\int \frac{\partial g_{\mu\nu}}{\partial x_m}\rho \frac{dx_\mu}{ds}\frac{dx_\nu}{ds} \cdot \frac{V}{\sqrt{G}}\frac{ds}{dt}\,, \tag{679}$$

as is indicated in the line following expression (676) in the notebook.[329] The factors $\sqrt{G}$ should be set equal to 1: if Einstein had not set $\sqrt{G} = 1$ in going from equation (71) to equation (675), the factors $\sqrt{G}$ would simply cancel at this point. Dividing by ρV, Einstein rewrote expression (679) as:

$$\frac{d}{dt}\left(\frac{1}{\sqrt{G}} g_{m\nu}\frac{dx_\nu}{ds}\right) - \frac{1}{2}\frac{1}{\sqrt{G}}\frac{\partial g_{\mu\nu}}{\partial x_m}\frac{dx_\mu}{dt}\frac{dx_\nu}{dt}\frac{dt}{ds}\,. \tag{680}$$

Setting $\dot{x}_\mu \equiv dx_\mu/dt$ and $w \equiv ds/dt$,[330] he rewrote this expression as

$$\frac{d}{dt}\left(\frac{1}{\sqrt{G}}\frac{g_{m\nu}\dot{x}_\nu}{w}\right) - \frac{1}{2}\frac{1}{\sqrt{G}}\frac{\frac{\partial g_{\mu\nu}}{\partial x_m}\dot{x}_\mu \dot{x}_\nu}{w}\,, \tag{681}$$

and, finally, as

$$\frac{d}{dt}\left(\frac{1}{\sqrt{G}}\frac{\partial w}{\partial \dot{x}_m}\right) - \left(\frac{1}{\sqrt{G}}\frac{\partial w}{\partial x_m}\right). \tag{682}$$

---

329 In the second term, Einstein wrote m instead of x_m.

330 The same notation is used in the Einstein-Besso manuscript on the perihelion advance of Mercury (CPAE 4, Doc. 14, e.g., [p. 15], [eq. 105]).

Setting this expression (with $\sqrt{G} = 1$) equal to zero, one recovers the Euler-Lagrange equations for the Lagrangian $H = w = ds/dt$ (see p. 5R and equation (55)). This shows that the equations of motion of a test particle in a metric field can indeed be derived from the energy-momentum balance between matter and gravitational field.

5.6 *Transition to the Entwurf Strategy (24R–25R)*

24R–25R

On pp. 19L–23L, Einstein had extracted various candidates for the left-hand side of the field equations from the Ricci tensor and the November tensor by imposing suitable coordinate restrictions. These coordinate restrictions should (*a*) make it possible to eliminate all unwanted terms with second-order derivatives of the metric, (*b*) be compatible with energy-momentum conservation at least in linear approximation, and (*c*) minimally allow transformations to accelerated frames of reference in Minkowski spacetime. In this way, Einstein's heuristic requirements (the correspondence principle, the conservation principle, the equivalence principle, and the relativity principle) would all at least to some extent be satisfied. All coordinate restrictions he had considered, however, failed on one count or another.

On p. 24R, Einstein proposed yet another candidate for the left-hand side of the field equations. However, while the various candidates proposed on pp. 19L–23L had been products of Einstein's mathematical strategy, the new candidate was a product of the physical strategy.[331] Instead of extracting candidate field equations from the Ricci tensor with various coordinate restrictions, Einstein on p. 24R generated field equations starting from the requirement of energy-momentum conservation. That does not mean that Einstein had now given up on the mathematical strategy altogether. The weak-field equations he started from and the restriction to unimodular transformations in all calculations on p. 24R strongly suggest that Einstein hoped to connect the field equations found through considerations of energy-momentum conservation to the November tensor. On p. 25L Einstein explicitly tried find a coordinate restriction with which he could recover field equations found along the lines of the argument on p. 24R from the November tensor.

At the top of p. 24R, Einstein wrote down an expression that can be identified as the divergence of a quantity representing gravitational stress-energy density. Although the derivation of this expression is not in the notebook, there are enough clues to give a plausible reconstruction of how Einstein arrived at it. As he had done for other linearized field equations on pp. 19R, 20L, and 21L, Einstein used the linearized version of field equations extracted from the November tensor to rewrite, in linear approximation, the gravitational force density as the divergence of the gravitational stress-energy pseudo-tensor. The expression at the top of p. 24R is the result of this calculation.

---

331 See sec. 1.2 for a discussion of these two strategies.

The gravitational stress-energy pseudo-tensor that one finds in this linear approximation looks like a plausible candidate for the exact expression for this quantity. It is understandable therefore that Einstein proceeded to look for terms that would need to be added to the weak-field equations to make sure that the expression at the top of p. 24R becomes *exactly* equal to the gravitational force density (see sec. 5.6.1). Essentially the same method would give him the *Entwurf* field equations on pp. 26L–R and in (Einstein and Grossmann 1913).[332]

Field equations constructed in this manner automatically satisfy both the correspondence principle and the conservation principle. The problem is to determine whether they are covariant under a wide enough class of coordinate transformations to meet the requirements of the relativity and equivalence principles as well. In the case of field equations extracted from the Ricci tensor or the November tensor with the help of coordinate restrictions, this question can, at least in principle, be settled by examining the transformation properties of expressions much simpler than the left-hand side of the field equations, such as the Hertz expression or the ϑ-expression. In the case of the field equations introduced on p. 24R, the construction of the equations is of no help in determining their covariance properties. The construction only guarantees that the equations will be covariant under unimodular linear transformations (unimodular because the determinant of the metric is set equal to unity in all calculations on p. 24R).

At the bottom of p. 24R, Einstein checked whether the rotation metric is a solution of the new field equations. According to the first entry on p. 24R, the expression from which they were derived vanishes for the rotation metric, a necessary condition for the rotation metric to be a solution of the vacuum field equations. This was an encouraging result—itself in error, it turns out—but Einstein discovered that the rotation metric is in fact not a solution of his new field equations. He also discovered, however, that he had erroneously cancelled two terms in his derivation of these equations. There would consequently be additional terms quadratic in first-order derivatives of the metric in the field equations. This in turn opened up the possibility that, once the equations were corrected, the rotation metric would be a solution after all.

Rather than making these corrections, Einstein (on pp. 25L–R) tried to find a coordinate restriction with the help of which (the corrected version of) these new field equations could be recovered from the November tensor (see sec. 5.6.2). In other words, he examined whether the field equations found following the physical strategy could also be found following the mathematical strategy. At the bottom of p. 25L he indicated which terms in the November tensor would have to be eliminated by imposing a coordinate restriction and which ones preserved. At the top of p. 25R, Einstein considered an ingenious modification of the ϑ-restriction, which we shall call the $\hat{\vartheta}$-restriction (see sec. 5.6.3), although it is not clear whether the purpose of the $\hat{\vartheta}$-

332 Einstein had in effect used this method before to derive the final version of the field equations for his theory for static gravitational fields in 1912. In that case, he had also added terms to the original field equations of the theory to make sure that they be compatible with energy-momentum conservation (see Einstein 1912b, 455–456).

restriction was to extract (the corrected version of) the candidate field equations of p 24R from the November tensor. In any case, Einstein failed to connect the November tensor to the physically motivated field equations of p. 24R, a connection that would have helped clarify the covariance properties of the latter.

It is at this point in the notebook that Einstein abandoned the mathematical strategy completely. On the remainder of p. 25R, Einstein started tinkering with the expression found on p. 24R for the left-hand side of field equations to make sure that it vanishes for the rotation metric (see sec. 5.6.4). Undeterred by the fact that the resulting expression is mathematically ill-defined, he checked whether this modified expression still allowed him to write the gravitational force density as the divergence of gravitational stress-energy density. He found that it did not, at least not exactly (see sec. 5.6.5). It may have been because the resulting conflict with energy-momentum conservation already ruled out this ill-defined expression as a candidate for the left-hand side of the field equations, but Einstein made no attempt to connect the modified expression with the November tensor. On p. 25L and at the top of p. 25R, Einstein had used the mathematical strategy to complement the physical strategy that he had used on p. 24R. On the remainder of p. 25R, however, Einstein apparently decided to go exclusively with the physical strategy, which on the very next page gave him the *Entwurf* equations.

5.6.1 *Constructing Field Equations from Energy-Momentum Conservation and Checking Them for Rotation (24R)* 24R

At the top of p. 24R, Einstein wrote down the expression[333]

$$\frac{\partial}{\partial x_i}\left(\gamma_{i\varepsilon}\frac{\partial g_{\alpha\beta}}{\partial x_\varepsilon}\frac{\partial \gamma_{\alpha\beta}}{\partial x_\sigma}\right)-\frac{1}{2}\frac{\partial}{\partial x_\sigma}\left(\gamma_{i\varepsilon}\frac{\partial g_{\alpha\beta}}{\partial x_\varepsilon}\frac{\partial \gamma_{\alpha\beta}}{\partial x_i}\right), \tag{683}$$

and noted that it vanishes for a metric field that he wrote down as

$$\begin{array}{ccc} & g & \\ -1 & 0 & \omega y \\ 0 & -1 & -\omega x \\ -y & -\omega x & 1-\omega^2(x^2+y^2) \end{array} \qquad \begin{array}{ccc} & \gamma & \\ -1+\omega^2 y^2 & \omega x y & \omega y \\ \omega x y & -1+\omega^2 x^2 & -\omega x \\ \omega y & -\omega x & 1. \end{array} \tag{684}$$

He explicitly wrote: “The expression [683] vanishes for the system [684]” (“Der Ausdruck … verschwindet für das System …”). The metric in equation (684) is easily recognized, despite some discrepancies, which are probably due to slips on Einstein’s part, as the rotation metric with one spatial dimension suppressed. It is harder to see what expression (683) represents. This seems to be one of the few places in the notebook where the calculations are not self-contained. Fortunately, a convincing case can be made for the following reconstruction of how Einstein arrived at expression (683).

333 In the notebook, the partial derivative with respect to x_i in the first term in equation (683) is written as an ordinary derivative.

In the field equations extracted from the November tensor with the help of the ϑ-restriction, the term with second-order derivatives of the metric is written as (see, e.g., expression (584) [p. 23L] and expressions (652) and (654) [p. 43LA])[334]

$$\frac{\partial}{\partial x_i}\left(\gamma_{i\varepsilon}\frac{\partial g_{\alpha\beta}}{\partial x_\varepsilon}\right), \tag{685}$$

where the labeling of the indices is chosen with a view to expression (683), the derivation of which we want to reconstruct. In linear approximation this is the only non-negligible term on the left-hand side of these field equations. The linearized version of these equation thus becomes

$$\frac{\partial}{\partial x_i}\left(\gamma_{i\varepsilon}\frac{\partial g_{\alpha\beta}}{\partial x_\varepsilon}\right) = \kappa\Theta_{\alpha\beta}, \tag{686}$$

where $\Theta_{\alpha\beta}$ is the covariant stress-energy tensor in the notation that Einstein used in the notebook.[335] Energy-momentum conservation required that these linearized field equations can be used to rewrite the gravitational force density as the divergence of a quantity representing gravitational stress-energy density.[336] Einstein had checked this for the linearized field equations considered on p. 19R (see equation (481)) and on p. 20L (see equation (500)). A natural explanation of how he arrived at the expression at the top of p. 24R is that he did the same for the linearized field equations (686).

The expression for the force density can be read off from the energy-momentum balance between matter and gravitational field in either its covariant or its contravariant form. On p. 24L, the page immediately preceding the one under consideration here, Einstein had actually considered both possibilities in his attempt to find a physical interpretation for the ϑ-metric (see sec. 5.5.9). These two alternative expressions for the force density are

$$-\frac{1}{2}\frac{\partial g_{\alpha\beta}}{\partial x_\sigma}T_{\alpha\beta}, \tag{687}$$

where $T_{\alpha\beta}$ is the contravariant stress-energy tensor, and

$$\frac{1}{2}\frac{\partial \gamma_{\alpha\beta}}{\partial x_\sigma}\Theta_{\alpha\beta} \tag{688}$$

---

334 In the equations extracted from the November tensor with the Hertz restriction (see equation (558) [p. 22R]) and in those extracted from the Ricci tensor with the harmonic restriction (see equation (477) [p. 19L]), the term with second-order derivatives is written as $\gamma_{\alpha\beta}\frac{\partial^2 g_{il}}{\partial x_\alpha \partial x_\beta}$. The two expressions only differ, of course, by a term quadratic in first-order derivatives of the metric.

335 See, e.g., equation (663). In our reconstruction of the derivation of equation (683), we follow the notation of the notebook.

336 Cf. the discussion following equation (481) in sec. 5.4.2.

(cf. expressions (659) and (668), respectively). Note that these expressions are correct only in linear approximation. In the exact version there will be another factor $\sqrt{-g}$ (cf. expression (482)), which in linear approximation can be set equal to unity.

On pp. 19R, 20L, and 21L, Einstein had used expression (687) for the force density but had substituted the left-hand side of the *covariant* linearized field equations for the *contravariant* stress-energy tensor (cf. footnote 241). Using expression (688) instead and eliminating the stress-energy tensor with the help of the linearized field equations (686), one arrives at

$$\frac{1}{2\kappa}\frac{\partial\gamma_{\alpha\beta}}{\partial x_\sigma}\frac{\partial}{\partial x_i}\left(\gamma_{i\varepsilon}\frac{\partial g_{\alpha\beta}}{\partial x_\varepsilon}\right). \tag{689}$$

Ignoring the factor 2κ (as Einstein did in the corresponding calculations on pp. 19R, 20L, and 21L), one can rewrite this expression as

$$\frac{\partial}{\partial x_i}\left(\gamma_{i\varepsilon}\frac{\partial g_{\alpha\beta}}{\partial x_\varepsilon}\frac{\partial\gamma_{\alpha\beta}}{\partial x_\sigma}\right)-\gamma_{i\varepsilon}\frac{\partial g_{\alpha\beta}}{\partial x_\varepsilon}\frac{\partial^2\gamma_{\alpha\beta}}{\partial x_i\partial x_\sigma}. \tag{690}$$

The first term is identical to the first term in expression (683) written at the top of p. 24R. The second term can be rewritten in the form of the second term in expression (683) if terms of third power in derivatives of the metric are neglected. Such third-power terms would correspond to quadratic terms in the field equations. Since this whole calculation is based on the linearized field equations, such terms can indeed be neglected. One can thus write the second term in expression (690) as

$$\gamma_{i\varepsilon}\frac{\partial g_{\alpha\beta}}{\partial x_\varepsilon}\frac{\partial^2\gamma_{\alpha\beta}}{\partial x_i\partial x_\sigma} = \frac{\partial}{\partial x_\sigma}\left(\gamma_{i\varepsilon}\frac{\partial g_{\alpha\beta}}{\partial x_\varepsilon}\frac{\partial\gamma_{\alpha\beta}}{\partial x_i}\right)-\gamma_{i\varepsilon}\frac{\partial^2 g_{\alpha\beta}}{\partial x_\sigma\partial x_\varepsilon}\frac{\partial\gamma_{\alpha\beta}}{\partial x_i}. \tag{691}$$

Once again neglecting terms of third power in derivatives of the metric, one easily establishes that the last term on the right-hand side is equal and opposite to the term on the left-hand side[337]

$$\gamma_{i\varepsilon}\frac{\partial^2 g_{\alpha\beta}}{\partial x_\sigma\partial x_\varepsilon}\frac{\partial\gamma_{\alpha\beta}}{\partial x_i} = \gamma_{i\varepsilon}\frac{\partial g_{\alpha\beta}}{\partial x_\varepsilon}\frac{\partial^2\gamma_{\alpha\beta}}{\partial x_i\partial x_\sigma}. \tag{692}$$

equation (691) can thus be rewritten as

337 Farther down on p. 24R, Einstein initially cancelled these two terms with one another in a calculation that is supposed to be exact. He later rescinded these calculations. We prove the approximate equality of these two terms in modern notation. Using that $0 = (g^{\mu\rho}g_{\rho\nu})_{,\alpha} = g^{\mu\rho}{}_{,\alpha}g_{\rho\nu}+g^{\mu\rho}g_{\rho\nu,\alpha}$, one can write $g_{\alpha\beta,\varepsilon\sigma}g^{\alpha\beta}{}_{,i} = (-g_{\alpha\mu}g_{\beta\nu}g^{\mu\nu}{}_{,\varepsilon})_{,\sigma}(-g^{\alpha\kappa}g^{\beta\lambda}g_{\kappa\lambda,i})$. If terms of third power in derivatives of the metric are neglected, this reduces to $g_{\alpha\beta,\varepsilon\sigma}g^{\alpha\beta}{}_{,i} = g^{\alpha\beta}{}_{,\varepsilon\sigma}g_{\alpha\beta,i}$. Contracting both sides with $g^{i\varepsilon}$ and switching i and ε on the right-hand side, one arrives at $g^{i\varepsilon}g_{\alpha\beta,\varepsilon\sigma}g^{\alpha\beta}{}_{,i} = g^{i\varepsilon}g^{\alpha\beta}{}_{,i\sigma}g_{\alpha\beta,\varepsilon}$, which is what we set out to prove.

$$\gamma_{i\varepsilon}\frac{\partial g_{\alpha\beta}}{\partial x_\varepsilon}\frac{\partial^2\gamma_{\alpha\beta}}{\partial x_i \partial x_\sigma} = \frac{1}{2}\frac{\partial}{\partial x_\sigma}\left(\gamma_{i\varepsilon}\frac{\partial g_{\alpha\beta}}{\partial x_\varepsilon}\frac{\partial \gamma_{\alpha\beta}}{\partial x_i}\right). \qquad (693)$$

Substituting this expression for the second term in expression (690), one recovers expression (683) written at the top of p. 24R.

The reconstruction given here of how Einstein arrived at this expression also seems to fit with the deleted phrase on the very first line of p. 24R: "divergence of a plane tensor Θ_{ik}" ("Divergenz eines Ebenentensors Θ_{ik}").[338] The expression in equation (683) is, in fact, the divergence of (-2κ times) the covariant gravitational stress-energy pseudo-tensor of the *Entwurf* theory.[339] The notation Θ_{ik} suggests that Einstein was referring to the stress-energy tensor of matter rather than to the stress-energy pseudo-tensor of the gravitational field, but energy-momentum conservation requires the divergence of the latter to be equal and opposite to the divergence of the former. The deleted phrase at the top of p. 24R thus seems to provide additional support for our reconstruction of the derivation of expression (683).

If there is no matter, energy-momentum conservation requires that the divergence of the stress-energy pseudo-tensor of the gravitational field vanishes. It follows that expression (683) should vanish for vacuum solutions of the field equations, at least in linear approximation. As becomes clear on the bottom half of p. 24R, Einstein tried to find the exact field equations corresponding to the linearized field equations (686) on the assumption that expression (683) is exactly equal to the gravitational force density. In that case, the expression should vanish exactly for vacuum solutions of the field equations. We can thus understand why it was important for Einstein to check whether the expression vanish exactly, for instance, for the rotation metric, which should be a vacuum solution of any acceptable candidate field equations according to Einstein's heuristic principles.

Contrary to Einstein's claim at the top of p. 24R, expression (683) does not vanish for the rotation metric. In the $2+1$-dimensional case, the covariant components of the rotation metric are given by

338 The term "plane tensor" ("Ebenentensor") is used in the notebook for a covariant tensor (cf. p. 17L and the discussion in sec. 5.3.3).

339 Relabeling indices to make it easier to compare the expression in the notebook with the corresponding expressions in the *Entwurf* paper, we can rewrite expression (683) as:

$$\frac{\partial}{\partial x_\nu}\left(\gamma_{\mu\nu}\left[\frac{\partial g_{\tau\rho}}{\partial x_\mu}\frac{\partial \gamma_{\tau\rho}}{\partial x_\sigma} - \frac{1}{2}g_{\mu\sigma}\gamma_{\alpha\beta}\frac{\partial g_{\tau\rho}}{\partial x_\alpha}\frac{\partial \gamma_{\tau\rho}}{\partial x_\beta}\right]\right).$$

The quantity in square brackets is equal to $-2\kappa t_{\mu\sigma}$ as defined in (Einstein and Grossmann 1913, p. 16, equation (14)). Substituting this definition into the expression above and dividing by -2, one arrives at the left-hand side of equation (12b) of the same paper for the special case that $\sqrt{-g} = 1$. Equation (12b) is obtained from equation (12), the left-hand side of which reduces to expression (683) if one sets $\sqrt{-g} = 1$.

$$g_{\mu\nu} = \begin{pmatrix} -1 & 0 & \omega y \\ 0 & -1 & -\omega x \\ \omega y & -\omega x & 1-\omega^2(x^2+y^2) \end{pmatrix}, \tag{694}$$

and the contravariant components by

$$\gamma_{\mu\nu} = \begin{pmatrix} -1+\omega^2 y^2 & -\omega^2 xy & \omega y \\ -\omega^2 xy & -1+\omega^2 x^2 & -\omega x \\ \omega y & -\omega x & 1 \end{pmatrix}. \tag{695}$$

These expressions differ slightly, but significantly as it turns out, from the expressions (684) given in the notebook.

Consider the $\sigma = 1$ component of expression (683) for the metric in equations (694)–(695). One easily verifies that the first term only contributes[340]

$$2\frac{\partial}{\partial x_2}\left(\gamma_{21}\frac{\partial g_{24}}{\partial x_1}\frac{\partial \gamma_{24}}{\partial x_1}\right) = 2\frac{\partial \gamma_{21}}{\partial x_2}\omega^2 = -2\omega^4 x, \tag{696}$$

whereas the second term only contributes[341]

$$-\frac{\partial}{\partial x_1}\left(\gamma_{22}\frac{\partial g_{14}}{\partial x_2}\frac{\partial \gamma_{14}}{\partial x_2}\right) = -\frac{\partial \gamma_{22}}{\partial x_1}\omega^2 = -2\omega^4 x. \tag{697}$$

The $\sigma = 1$ component of expression (683) thus gives $-4\omega^4 x$ for the metric in equations (694)–(695). The $\sigma = 2$ component likewise gives $-4\omega^4 y$.[342] Only the $\sigma = 4$ component vanishes, since the metric is time-independent.

A comparison between the correct expressions for the components of the metric in equations (694)–(695) and Einstein's faulty expression (684) suggests that it was because of the sign errors in γ_{12} and γ_{21} that Einstein came to believe that expression (683) vanishes for the rotation metric. Inserting $\omega^2 xy$ instead of $-\omega^2 xy$ for γ_{21} in equation (696), the contribution coming from equation (696) would cancel the contribution coming from equation (697) and the $\sigma = 1$ component of equation (683) would vanish for the rotation metric.[343]

---

340 There will be identical contributions for $(\alpha = 2, \beta = 4)$ and $(\alpha = 4, \beta = 2)$.

341 There will be identical contributions for $(\alpha = 1, \beta = 4)$ and $(\alpha = 4, \beta = 1)$.

342 The non-vanishing contributions to the $\sigma = 2$ component of expression (683) for this metric are

$$2\frac{\partial}{\partial x_1}\left(\gamma_{12}\frac{\partial g_{14}}{\partial x_2}\frac{\partial \gamma_{14}}{\partial x_2}\right) - \frac{\partial}{\partial x_2}\left(\gamma_{11}\frac{\partial g_{24}}{\partial x_1}\frac{\partial \gamma_{24}}{\partial x_1}\right) = 2\frac{\partial \gamma_{12}}{\partial x_1}\omega^2 - \frac{\partial \gamma_{11}}{\partial x_2}\omega^2 = -4\omega^4 y.$$

343 A similar cancellation would occur in the $\sigma = 2$ component of expression (683) (see the preceding note).

It is interesting to note in this context that the sign error in γ_{12} is the only one of the errors in expression (684) that Einstein repeated on p. 25R where he once again wrote down the components of the rotation metric. What may have happened is that Einstein read off the contravariant components of the rotation metric from the expression for the covariant components of the ϑ-metric on p. 42R (see equation (629)), which would give $\gamma_{12} = \alpha xy$, and then set α equal to ω rather than to $-\omega^2$.

Whatever happened, Einstein somehow convinced himself that expression (683) vanishes exactly for the important special case of the rotation metric.[344] It now made sense for him to look upon equation (683), which he presumably found as the result of a calculation in linear approximation, as the exact expression for the divergence of the gravitational stress-energy density,[345] at least for metric fields with a determinant equal to unity.

It is not entirely clear whether Einstein was aware of this last complication at this point. Perhaps he erroneously continued to use expression (688) for the gravitational force density in linear approximation, even though the calculation on the bottom half of p. 24R was supposed to be exact. It is also possible that he consciously set $\sqrt{-g} = 1$ to facilitate comparison of the result of his calculations with the November tensor, which is a tensor only under unimodular transformations. Einstein would include the factors $\sqrt{-g}$ that were omitted on p. 24R in the derivation of the *Entwurf* equations on p. 26L. Once again it is not clear whether that was because he realized that he should have used the exact expression for the gravitational force density in a calculation that is supposed to be exact or because he was no longer interested in trying to recover his new candidate field equations from the November tensor.

If expression (683) is exactly equal to the divergence of the gravitational stress-energy density and the determinant of the metric is set equal to unity, one can find the exact field equations by going through the derivation of expression (683) given in equations (686)–(693) in reverse, this time without neglecting any terms. More specifically, one can rewrite expression (683) in the form

---

344 In passing we note that Einstein thus missed an early opportunity to discover that the rotation metric is not a solution of the *Entwurf* equations. As was pointed out in footnote 339, expression (683) is the divergence of the gravitational stress-energy pseudo-tensor of the *Entwurf* theory in the special case that $\sqrt{-g} = 1$. The rotation metric has a determinant equal to unity. The vanishing of expression (683) for this metric is therefore a necessary condition for it to be a solution of the *Entwurf* field equations. It expresses energy-momentum conservation in this case. For further discussion of Einstein's struggles with rotation, see (Janssen 1999; 2005, 68–71), and “What Did Einstein Know …” sec. 3 (in this volume).

345 Note the close structural similarity between the expression for gravitational stress-energy density one reads off from expression (683) (see the expression in square brackets in footnote 339) and the gravitational stress tensor of Einstein's 1912 static theory, the 11-component of which Einstein had tried to translate into a gravitational stress-energy (pseudo-)tensor in his metric theory at the bottom of p. 21R (see sec. 5.4.5). This similarity may have been another factor leading Einstein to adopt expression (683) as the *exact* expression for the divergence of the gravitational stress-energy density.

$$\frac{\partial\gamma_{\alpha\beta}}{\partial x_\sigma}(\qquad)_{\alpha\beta} \tag{698}$$

and take the as yet unknown expression in parentheses as the left-hand side of the field equations. After all, substituting the right-hand side of these candidate field equations—the stress-energy tensor $\Theta_{\alpha\beta}$ —for the as yet unknown expression in parentheses above, one recovers (except for an immaterial factor of $\frac{1}{2}$) expression (688) for the gravitational force density

$$\frac{\partial\gamma_{\alpha\beta}}{\partial x_\sigma}\Theta_{\alpha\beta}\,. \tag{699}$$

These new field equations thus automatically and exactly satisfy energy-momentum conservation. They guarantee that the gravitational force density is equal to the divergence of the gravitational stress-energy density, in which case the energy-momentum balance between matter and gravitational field can be written as the vanishing of the divergence of the total stress-energy density.

This is exactly Einstein's line of reasoning on p. 24R. With the comment "the above expression yields" ("obiger Ausdruck liefert"), he began to rewrite expression (683) in the form (698):

$$\begin{aligned}&\frac{\partial\gamma_{\alpha\beta}}{\partial x_\sigma}\frac{d}{dx_i}\left(\gamma_{i\varepsilon}\frac{\partial g_{\alpha\beta}}{\partial x_\varepsilon}\right)+\gamma_{i\varepsilon}\frac{\partial g_{\alpha\beta}}{\partial x_\varepsilon}\frac{\partial^2\gamma_{\alpha\beta}}{\partial x_i\partial x_\sigma}\\ &-\gamma_{i\varepsilon}\frac{\partial^2 g_{\alpha\beta}}{\partial x_\varepsilon\partial x_\sigma}\frac{\partial\gamma_{\alpha\beta}}{\partial x_i}-\frac{1}{2}\frac{\partial\gamma_{i\varepsilon}}{\partial x_\sigma}\frac{\partial g_{\alpha\beta}}{\partial x_\varepsilon}\frac{\partial\gamma_{\alpha\beta}}{\partial x_i}\end{aligned} \tag{700}$$

In the notebook, Einstein initially (and erroneously) cancelled the second term in this expression against the third.[346] Relabeling the summation indices in the last term, he read off the left-hand side of the field equations from the remaining two terms. "This suggests" ("Hierdurch nahe gelegt"), he wrote,

$$\frac{\partial}{\partial x_i}\left(\gamma_{i\varepsilon}\frac{\partial g_{\alpha\beta}}{\partial x_\varepsilon}\right)-\frac{1}{2}\frac{\partial g_{i\varepsilon}}{\partial x_\alpha}\frac{\partial\gamma_{i\varepsilon}}{\partial x_\beta}\,. \tag{701}$$

Einstein's next step was to check whether expression (701) vanishes for the rotation metric as it should if this metric is to be a vacuum solution of these new field equations. It is not, as Einstein noted on the last two lines of p. 24R: "Tried for the case of a rotating body[.] $\alpha = \beta = 1$ gives $-\omega^2$" ("Probiert am Fall des rot[ierenden] Körpers $\alpha = \beta = 1$ liefert $-\omega^2$").

346 These two terms cancel only if terms of third power in derivatives of the metric are neglected (see equation (692) and footnote 337).

Inserting either our equations (694)–(695) or Einstein's expressions (684)[347] for the components of the rotation metric into the 11 -component of expression (701), one readily verifies this result. The only non-vanishing contributions come from the second term in expression (701)

$$-\frac{\partial g_{24}}{\partial x_1}\frac{\partial \gamma_{24}}{\partial x_1} = -\omega^2 . \tag{702}$$

Einstein's new candidate field equations are therefore unacceptable as they stand. However, Einstein also discovered that the two terms that he had cancelled with one another in his derivation of expression (701) for the left-hand side of his new field equations do in fact not cancel. To the right of expression (700) in the notebook, Einstein did a short calculation to check whether these two terms are equal to one another. He started with the fairly self-evident relation[348]

$$\frac{\partial g_{\alpha\beta}}{\partial x_\sigma}\frac{\partial \gamma_{\alpha\beta}}{\partial x_i} = \frac{\partial \gamma_{\alpha\beta}}{\partial x_\sigma}\frac{\partial g_{\alpha\beta}}{\partial x_i} . \tag{703}$$

This equation expresses that one can simply switch co- and contravariant components of the metric in contractions of two first-order derivatives of this form. One might expect that this is also true for similar contractions of a first-order derivative and a second-order derivative. Differentiating equation (703) with respect to x_ε, however, as Einstein did on the next line,

$$\frac{\partial^2 g_{\alpha\beta}}{\partial x_\varepsilon \partial x_\sigma}\frac{\partial \gamma_{\alpha\beta}}{\partial x_i} + \frac{\partial g_{\alpha\beta}}{\partial x_\sigma}\frac{\partial^2 \gamma_{\alpha\beta}}{\partial x_\varepsilon \partial x_i} = \frac{\partial^2 \gamma_{\alpha\beta}}{\partial x_\varepsilon \partial x_\sigma}\frac{\partial g_{\alpha\beta}}{\partial x_i} + \frac{\partial \gamma_{\alpha\beta}}{\partial x_\sigma}\frac{\partial^2 g_{\alpha\beta}}{\partial x_\varepsilon \partial x_i} , \tag{704}$$

one sees that this is not the case. At this point, Einstein probably rescinded his cancellations in expression (700) with the proof readers' stet mark that he typically used for this purpose.

On the face of it, these two terms in expression (700) would contribute additional terms with unwanted second-order derivatives of the metric to the field equations in equation (701). However, as we already noted in our reconstruction of the derivation of equation (683), the two terms only differ by an expression of third power in first-order derivatives of the metric (see footnote 337). They thus only give rise to another term quadratic in first-order derivatives of the metric in the field equations.

Einstein did not add such a term to expression (701), but he probably realized that his new candidate field equations only needed to be corrected by terms quadratic in first-order derivatives rather than by second-order derivative terms. Otherwise it becomes hard to understand his calculations on the next page. On p. 25L he looked for a new coordinate restriction to extract field equations from the November tensor,

347 The errors in (684) do not matter in this calculation.

348 In modern notation, the proof runs as follows (cf. footnote 337):

$$g_{\alpha\beta,\sigma}g^{\alpha\beta}{}_{,i} = (-g_{\alpha\mu}g_{\beta\nu}g^{\mu\nu}{}_{,\sigma})(-g^{\alpha\kappa}g^{\beta\lambda}g_{\kappa\lambda,i}) = \delta^\kappa_\mu\delta^\lambda_\nu g^{\mu\nu}{}_{,\sigma}g_{\kappa\lambda,i} = g^{\kappa\lambda}{}_{,\sigma}g_{\kappa\lambda,i} .$$

using as his guide that these equations should at least include the two terms in expression (701) and no additional terms with second-order derivatives of the metric.

All in all, Einstein had made important progress on p. 24R. He had found a method to construct exact field equations out of linearized ones by demanding exact compliance with energy-momentum conservation. The first result of this method, however, was problematic. He found that the rotation metric is not a solution of his new field equations. Einstein also realized, however, that he had made an error along the way. It was thus at least conceivable that the correct application of his new method would yield field equations that do allow the rotation metric as a solution.

5.6.2 *Trying to Recover the Physically Motivated Field Equations from the November Tensor (25L)* 25L

On p. 25L, Einstein tried to recover the field equations he had found on p. 24R from the November tensor, not just the two terms explicitly given at the bottom of p. 24R (see expression (701)) but also the additional terms coming from the erroneously cancelled terms in expression (700) for the gravitational force density from which he had read off these new field equations.

At the top of p. 25L, Einstein wrote down the November tensor (see equation (544))

$$\frac{\partial}{\partial x_k}\begin{Bmatrix}\alpha & \beta \\ & k\end{Bmatrix} - \begin{Bmatrix}\alpha & \lambda \\ & \mu\end{Bmatrix}\begin{Bmatrix}\beta & \mu \\ & \lambda\end{Bmatrix}. \tag{705}$$

A first indication that his purpose was to recover his new candidate field equations from this object is that he changed the free indices i and l in the original expression for the November tensor to α and β, the free indices in expression (701) for the left-hand side of the new candidate field equations. Using the definitions of the Christoffel symbols, Einstein rewrote (2 times) expression (705) as

$$\begin{gathered}\frac{\partial}{\partial x_k}\gamma_{k\lambda}\left[\frac{\partial g_{\alpha\lambda}}{\partial x_\beta} + \frac{\partial g_{\beta\lambda}}{\partial x_\alpha} - \underline{\frac{\partial g_{\alpha\beta}}{\partial x_\lambda}}\right] - \frac{1}{2}\gamma_{\mu\sigma}\gamma_{\lambda\tau}\left[\frac{\partial g_{\alpha\sigma}}{\partial x_\lambda} - \frac{\partial g_{\alpha\lambda}}{\partial x_\sigma}\right]\left[\frac{\partial g_{\beta\tau}}{\partial x_\mu} - \frac{\partial g_{\beta\mu}}{\partial x_\tau}\right] \\ -\frac{1}{2}\underline{\gamma_{\mu\sigma}\gamma_{\lambda\tau}\frac{\partial g_{\lambda\sigma}}{\partial x_\alpha}\frac{\partial g_{\mu\tau}}{\partial x_\beta}}\end{gathered} \tag{706}$$

From the way the product of Christoffel symbols in expression (705) was rewritten in expression (706), it is clear that Einstein once again used the symmetry argument that he had already used several times before in this calculation (on pp. 17R, 19L, and 22R; see the discussion following expression (463)). The two terms that Einstein underlined in expression (706) are easily recognized as (minus) the two terms of the new candidate field equations in equation (701). For the first term this is just a matter of relabeling indices. For the second term it is shown by the calculation immediately following expression (706) in the notebook. Using that

$$\gamma_{\lambda\tau}\frac{\partial g_{\lambda\sigma}}{\partial x_\alpha} = -\frac{\partial\gamma_{\lambda\tau}}{\partial x_\alpha}g_{\lambda\sigma} \tag{707}$$

and that

$$\gamma_{\mu\sigma}g_{\lambda\sigma}\frac{\partial g_{\mu\tau}}{\partial x_\beta} = \frac{\partial g_{\lambda\tau}}{\partial x_\beta}, \tag{708}$$

Einstein rewrote (-2 times) the last term in expression (706) as

$$\gamma_{\mu\sigma}\gamma_{\lambda\tau}\frac{\partial g_{\lambda\sigma}}{\partial x_\alpha}\frac{\partial g_{\mu\tau}}{\partial x_\beta} = -\frac{\partial\gamma_{\lambda\tau}}{\partial x_\alpha}\frac{\partial g_{\lambda\tau}}{\partial x_\beta} = -\frac{\partial\gamma_{\lambda\tau}}{\partial x_\beta}\frac{\partial g_{\lambda\tau}}{\partial x_\alpha}. \tag{709}$$

In the notebook, the relations (707) and (708) are given underneath the relevant terms on the left-hand side of equation (709). Dividing the right-hand side of equation (709) by 2, one recovers the second term in expression (701).

This was a promising start. The next task would be to identify those terms in the November tensor that still need to be added to the field equations based on expression (701) because of the erroneous cancellation of two terms in expression (700) for the gravitational force density. These terms would have to come from the second of the three terms in expression (706). This is the term that Einstein turned to next. It can be rewritten as

$$-\frac{1}{2}\gamma_{\mu\sigma}\gamma_{\lambda\tau}\frac{\partial g_{\alpha\sigma}}{\partial x_\lambda}\left(\frac{\partial g_{\beta\tau}}{\partial x_\mu}-\frac{\partial g_{\beta\mu}}{\partial x_\tau}\right)+\frac{1}{2}\gamma_{\mu\sigma}\gamma_{\lambda\tau}\frac{\partial g_{\alpha\lambda}}{\partial x_\sigma}\left(\frac{\partial g_{\beta\tau}}{\partial x_\mu}-\frac{\partial g_{\beta\mu}}{\partial x_\tau}\right). \tag{710}$$

The second term in this expression is equal and opposite to the first, as one easily verifies by relabeling indices ($\lambda\sigma\mu\tau \rightarrow \sigma\lambda\tau\mu$). Expression (710) can thus be rewritten as

$$-\gamma_{\mu\sigma}\gamma_{\lambda\tau}\frac{\partial g_{\alpha\sigma}}{\partial x_\lambda}\frac{\partial g_{\beta\tau}}{\partial x_\mu}+\gamma_{\mu\sigma}\gamma_{\lambda\tau}\frac{\partial g_{\alpha\sigma}}{\partial x_\lambda}\frac{\partial g_{\beta\mu}}{\partial x_\tau}, \tag{711}$$

which is the expression written on the next line in the notebook. To facilitate comparison with the two cancelled terms in expression (700), Einstein contracted expression (711) with $\partial\gamma_{\alpha\beta}/\partial x_i$, which appears next to it separated by a vertical line. If the terms in expression (711) were part of the left-hand side of the field equations, this contraction would give their contribution to the gravitational force density. Einstein rewrote the first term in expression (711) as

$$-\frac{\partial\gamma_{\mu\sigma}}{\partial x_\lambda}\frac{\partial\gamma_{\lambda\tau}}{\partial x_\mu}g_{\alpha\sigma}g_{\beta\tau}, \tag{712}$$

which, upon contraction with $\dfrac{\partial\gamma_{\alpha\beta}}{\partial x_i}$ gives

$$+\frac{\partial\gamma_{\mu\sigma}}{\partial x_\lambda}\frac{\partial\gamma_{\lambda\alpha}}{\partial x_\mu}\frac{\partial g_{\alpha\sigma}}{\partial x_i}. \tag{713}$$

The calculation was not pursued any further.

Einstein drew a horizontal line and schematically rewrote (parts of) the three terms in expression (706) for the November tensor at the top of the page[349]

$$\begin{array}{ll} \frac{\partial}{\partial x_k}\left(\gamma_{k\lambda}\frac{\partial g_{\alpha\beta}}{\partial x_\lambda}\right) & -\frac{1}{2}\;\underline{\qquad\qquad} \\ & -\frac{1}{2}\;\underline{\qquad\qquad} \end{array} \tag{714}$$

The first term in equation (714) is the last of the three terms with second-order derivatives of the metric in equation (706) and the only one that occurs in the candidate field equations on p. 24R (see equation (701)). The problem is to find a coordinate restriction with the help of which the other two second-order derivative terms in the November tensor can be eliminated. Application of this coordinate restriction will also eliminate some of the terms quadratic in first-order derivatives that are schematically indicated in the last two terms in expression (714). A suitable coordinate restriction, however, should preserve the last term in its entirety (i.e., the second underlined term in expression (706)) as well as the part of the second term corresponding to the erroneously cancelled terms in expression (700). In this way the field equations of p. 24R could be extracted from the November tensor.

Labeling the three terms in expression (706) for the November tensor a, b, and c, respectively, and using a prime to distinguish parts that should be preserved from parts that should be eliminated, one can schematically write expression (706) as[350]

$$a - a' - b + b' + c'. \tag{715}$$

The signs of a' and c' reflect that the signs with which the corresponding expressions occur in the November tensor are the opposite of the signs with which they occur in the field equations of p. 24R. Stated in terms of expression (715), the problem is to find a coordinate restriction such that a and b can be eliminated and the left-hand side of the field equations becomes:

$$a' - b' - c'. \tag{716}$$

Since Einstein did not pursue the calculation on p. 25L any further it is hard to interpret the material at the bottom of the page, but the reasoning leading to equations

---

349 In the notebook, some of the indices in the first term in expression (714) are barely legible. They have been transcribed on the assumption that Einstein copied this term from the corresponding term in expression (706).

350 More explicitly, $a' \equiv \frac{\partial}{\partial x_k}\left(\gamma_{k\lambda}\frac{\partial g_{\alpha\beta}}{\partial x_\lambda}\right)$ and $c' \equiv \frac{1}{2}\frac{\partial\gamma_{\lambda\tau}}{\partial x_\beta}\frac{\partial g_{\lambda\tau}}{\partial x_\alpha}$.

(715) and (716) above at least provides a plausible interpretation of the two lines at the bottom the p. 25L,

$$\begin{array}{ccc} a - a' & -b + b' & +c' \\ a' & -b' & -c'. \end{array} \tag{717}$$

It remains unclear why next to this expression he wrote down the terms

$$\begin{gathered} 2\gamma_{\mu\sigma}\gamma_{\lambda\tau}\frac{\partial g_{\alpha\sigma}}{\partial x_\lambda}\frac{\partial g_{\mu\tau}}{\partial x_\beta} \\ +\frac{\partial g_{\beta\tau}}{\partial x_\mu}\frac{\partial g_{\lambda\sigma}}{\partial x_\alpha}. \end{gathered} \tag{718}$$

These terms can be identified as coming from the expression

$$-\frac{1}{2}\gamma_{\mu\sigma}\gamma_{\lambda\tau}\frac{\partial g_{\lambda\sigma}}{\partial x_\alpha}\left(\frac{\partial g_{\beta\tau}}{\partial x_\mu}-\frac{\partial g_{\beta\mu}}{\partial x_\tau}\right)+\frac{1}{2}\gamma_{\mu\sigma}\gamma_{\lambda\tau}\left(\frac{\partial g_{\alpha\sigma}}{\partial x_\lambda}-\frac{\partial g_{\alpha\lambda}}{\partial x_\sigma}\right)\frac{\partial g_{\mu\tau}}{\partial x_\beta}, \tag{719}$$

which Einstein neglected in his expansion of (twice) the product of Christoffel symbols in equation (706) on the basis of the symmetry argument that he had come to use routinely in this calculation. Both terms vanish identically, since they are contractions of a part that is symmetric and a part that is anti-symmetric in the same index pair. Perhaps Einstein wanted to include these terms because the application of the coordinate restriction he was looking for at this point would preserve some of the terms in equation (719) while eliminating others. Even on this interpretation, however, it remains unclear what special appeal the two terms in equation (718) had for Einstein or why they appear with a factor 2 rather than with a factor $1/2$ as in equation (717).

Despite these uncertainties, it seems clear that the purpose of Einstein's calculations on p. 25L was to find a way of extracting the physically motivated field equations of p. 24R from the November tensor.[351] The material on the bottom half of the page strongly suggests that Einstein hoped to achieve this goal by finding a coordinate restriction that would allow him to eliminate all terms from the November tensor that do not occur in these new field equations.

At the top of the next page, p. 25R, Einstein considered a variant of the ϑ-restriction, with the help of which he had eliminated unwanted terms from the November tensor on p. 23L. There is no indication, however, that Einstein specifically introduced or used this restriction to recover the field equations of p. 24R and its correction terms from the November tensor.

351 This is similar to Einstein's attempt on pp. 9L–9R to connect the physically motivated core operator to the mathematically well-defined second Beltrami invariant.

5.6.3 *The $\hat{\vartheta}$-Restriction (25R, 23L)* 25R, 23L

The fragmentary material at the top of p. 25R—various arrays of numbers, the components of the rotation metric, and several equations—can all be understood as part of a variant of the calculations on p. 23L and pp. 42L–R involving what we have called the ϑ-restriction. We shall call this variant the $\hat{\vartheta}$-restriction. This interpretation of the material on p. 25R also explains the alternate signs in many of the expressions on p. 23L. Einstein, it seems, went back to p. 23L and indicated what would need to be changed in his earlier calculation if the ϑ-restriction were replaced by the $\hat{\vartheta}$-restriction. Most of what is actually written down at the top of p. 25R is aimed at determining whether the $\hat{\vartheta}$-restriction allows transformations to rotating frames of reference in Minkowski spacetime. It does not, which is probably why the $\hat{\vartheta}$-restriction was quickly abandoned.

The basic idea of the ϑ-restriction (see sec. 5.5.4) was to absorb terms that need to be eliminated from the November tensor into the so-called ϑ-expression. The November tensor was then split into various parts that separately transform as tensors under ϑ-transformations, i.e., those unimodular transformations under which the ϑ-expression transforms as a tensor. Subtracting the parts containing the terms that need to be eliminated, one arrives at candidate field equations that are invariant under ϑ-transformations.

Looking at the first term of the November tensor, which Einstein had just reexamined on p. 25L (cf. expressions (705)–(706) on p. 25L),

$$\frac{\partial}{\partial x_k}\begin{Bmatrix}\alpha & \beta \\ & k\end{Bmatrix} = \frac{\partial}{\partial x_k}\left(\gamma_{k\lambda}\frac{1}{2}\left[\frac{\partial g_{\alpha\lambda}}{\partial x_\beta} + \frac{g_{\beta\lambda}}{\partial x_\alpha} - \frac{g_{\alpha\beta}}{\partial x_\lambda}\right]\right), \tag{720}$$

one sees that the unwanted terms with second-order derivatives of the metric come from the first two terms in the Christoffel symbol

$$\begin{bmatrix}\alpha & \beta \\ & \lambda\end{bmatrix} = \frac{1}{2}\left(\frac{\partial g_{\alpha\lambda}}{\partial x_\beta} + \frac{g_{\beta\lambda}}{\partial x_\alpha} - \frac{g_{\alpha\beta}}{\partial x_\lambda}\right). \tag{721}$$

These terms can be absorbed into the ϑ-expression, defined as (see expression (570))

$$\vartheta_{\alpha\beta\lambda} \equiv \frac{1}{2}\left(\frac{g_{\alpha\lambda}}{\partial x_\beta} + \frac{g_{\beta\lambda}}{\partial x_\alpha} + \frac{g_{\alpha\beta}}{\partial x_\lambda}\right). \tag{722}$$

With the help of the ϑ-expression, the Christoffel symbols of the first kind can be written as (see equation (572))

$$\begin{bmatrix}\alpha & \beta \\ & \lambda\end{bmatrix} = \vartheta_{\alpha\beta\lambda} - \frac{\partial g_{\alpha\beta}}{\partial x_\lambda}. \tag{723}$$

The two unwanted terms can also be absorbed into a slightly different expression, which we shall call the $\hat{\vartheta}$-expression, and which we define as

$$\hat{\vartheta}_{\alpha\beta\lambda} \equiv \frac{1}{2}\left(\frac{g_{\alpha\lambda}}{\partial x_\beta} + \frac{g_{\beta\lambda}}{\partial x_\alpha} - 2\frac{g_{\alpha\beta}}{\partial x_\lambda}\right) . \tag{724}$$

With the help of the $\hat{\vartheta}$-expression, the Christoffel symbols of the first kind can be written as

$$\begin{bmatrix} \alpha & \beta \\ & \lambda \end{bmatrix} = \hat{\vartheta}_{\alpha\beta\lambda} + \frac{1}{2}\frac{\partial g_{\alpha\beta}}{\partial x_\lambda} . \tag{725}$$

The $\hat{\vartheta}$-expression was not written down explicitly in the notebook. In the upper right
25R corner of p. 25R, however, we find the schematic array of numbers

$$\begin{array}{ccc} & & +1 \\ +1 & +1 & -1 \\ & & -2 \end{array} . \tag{726}$$

Note that this is essentially a matrix with three rows and three columns. The first two columns just have $+1$ on all three rows. Comparison of the matrix (726) with equations (721), (722), and (724) suggests the following interpretation of these numbers. The three rows may represent the coefficients of the three terms in the ϑ-expression, the Christoffel symbol, and the $\hat{\vartheta}$-expression, respectively.

One can use either the ϑ-expression or the $\hat{\vartheta}$-expression in combination with the corresponding coordinate restriction to eliminate unwanted terms from the November tensor. The pluses and minuses written above many of the signs in expressions on
23L p. 23L suggest that Einstein actually went back to p. 23L to see what would need to be changed in his earlier calculations if the ϑ-expression were replaced by the $\hat{\vartheta}$-expression. He began with expression (570), the definition of the ϑ-expression (see also equation (722)), where he wrote -2 above the last plus sign,

$$\frac{1}{2}\left(\frac{\partial g_{ik}}{\partial x_\lambda} + \frac{\partial g_{k\lambda}}{\partial x_i} \overset{-2}{+} \frac{\partial g_{\lambda i}}{\partial x_k}\right) . \tag{727}$$

In this way the expression turns into the definition of the $\hat{\vartheta}$-expression (see equation (724)).

Einstein then wrote plus signs above the minus signs in equations (572) and (573) to indicate that he wanted to express the Christoffel symbols of the first and the second kind in terms of this new $\hat{\vartheta}$-expression,

$$\begin{bmatrix} i & l \\ & k \end{bmatrix} = \vartheta_{ilk} \overset{+}{-} \frac{\partial g_{il}}{\partial x_k} , \tag{728}$$

$$\begin{Bmatrix} i & l \\ & k \end{Bmatrix} = \gamma_{k\alpha}\left(\vartheta_{il\alpha} \overset{+}{-} \frac{\partial g_{il}}{\partial x_\alpha}\right) . \tag{729}$$

Comparing equation (728) to equation (725), reading $\hat{\vartheta}$ for ϑ, one sees that there is a factor $1/2$ missing in front of the derivative of the metric in equation (728). Equation (729) inherits this error from equation (728). This may be why Einstein subsequently deleted the plus signs in these two equations. He went through the rest of the calculation on p. 23L, however, on the assumption that the Christoffel symbols are related to the $\hat{\vartheta}$-expression according to equations (728)-(729). He added pluses and minuses to equations (574) and (578), and to expression (581), making only one minor error. He neglected to change the minus sign in the first term in equation (578) to a plus sign. Because of these errors, he found that the field equations extracted from the November tensor with the help of the $\hat{\vartheta}$-restriction are exactly the same as those extracted with the help of the ϑ-restriction (see expression (584)).

Einstein had abandoned these field equations because of problems with the ϑ-restriction (see secs. 5.5.6–5.5.9). Perhaps these problems could be avoided with the $\hat{\vartheta}$-restriction. In particular, it would be interesting to know whether the $\hat{\vartheta}$-expression, unlike the ϑ-expression, vanishes for the rotation metric. In that case the $\hat{\vartheta}$-restriction, unlike the ϑ-restriction, would at least allow transformations to rotating coordinates in the important special case of Minkowski spacetime (cf. the discussion at the beginning of sec. 5.5.6).

At the top of p. 25R, Einstein once again wrote down the covariant and contravariant components of the rotation metric in the $2+1$-dimensional case (cf. equation (694)–(695)) 25R

$$
\begin{array}{ccc}
-1 & 0 & \omega y \\
0 & -1 & -\omega x \\
\omega y & -\omega x & 1-\omega^2(x^2+y^2) \\
\hline
-1+\omega^2 y^2 & \omega^2 xy & \omega y \\
\omega^2 xy & -1+\omega^2 x^2 & -\omega x \\
\omega y & -\omega x & -1 .
\end{array}
\tag{730}
$$

He also wrote down several components of an equation that can more compactly be written as

$$
-2\hat{\vartheta}_{\alpha\beta\lambda} = 2\frac{g_{\alpha\beta}}{\partial x_\lambda} - \frac{g_{\alpha\lambda}}{\partial x_\beta} - \frac{g_{\beta\lambda}}{\partial x_\alpha} = 0 . \tag{731}
$$

The choice of index combinations for which equation (731) was examined on p. 25R confirms that the purpose of Einstein's calculations at this point was to check whether the metric (730) satisfies equation (731), i.e., whether the $\hat{\vartheta}$-expression vanishes for the rotation metric. Notice that expression (730) for the components of the metric is still not completely accurate. Comparing expression (730) to expression (684) for the rotation metric, one sees that most of the errors on p. 24L have been corrected on p. 25R, but that γ_{12} and γ_{21} still have the wrong sign[352] (as does γ_{44} which was given correctly on p. 24L).

To the left of expression (730), Einstein schematically wrote down equation (731) for the index combinations 112 and 121 :

$$2\begin{matrix}1\;1\\2\end{matrix} - \begin{matrix}1\;2\\1\end{matrix} - \begin{matrix}1\;2\\1\end{matrix}\,, \tag{732}$$

$$2\begin{matrix}1\;2\\1\end{matrix} - \begin{matrix}1\;2\\1\end{matrix} - \begin{matrix}1\;1\\2\end{matrix}\,. \tag{733}$$

Both components give the same equation,

$$\frac{\partial g_{12}}{\partial x_1} - \frac{\partial g_{11}}{\partial x_2} = 0\,. \tag{734}$$

This equation is obviously satisfied by the metric (730) as are all components of equation (731) involving only the indices 1 and 2 . Only components in which at least one index is equal to 4 need to be examined. To the right of equation (730), Einstein wrote down equation (731) for three such components, corresponding to the index combinations 124 , 224 , and 441 .[353]

$$2\frac{\partial g_{12}}{\partial x_4} - \frac{\partial g_{14}}{\partial x_2} - \frac{\partial g_{24}}{\partial x_1} = 0 \tag{735}$$

$$\frac{\partial g_{22}}{\partial x_4} - \frac{\partial g_{24}}{\partial x_2} = 0\,, \tag{736}$$

$$\frac{\partial g_{44}}{\partial x_1} - \frac{\partial g_{14}}{\partial x_4} = 0\,. \tag{737}$$

The metric (730) satisfies the first and the second equation, but not the third. Equation (737), the last of the five components of equation (731) given in the notebook, gives

$$\frac{\partial g_{44}}{\partial x_1} - \frac{\partial g_{14}}{\partial x_4} = -2\omega^2 x\,. \tag{738}$$

The $\hat{\vartheta}$-expression therefore does not vanish for the rotation metric. At this point, Einstein seems to have abandoned the $\hat{\vartheta}$-expression and the corresponding $\hat{\vartheta}$-restriction.

In the calculation on the next two lines of p. 25R, the metric that we have called the ϑ-metric briefly resurfaces (for discussion, see sec. 5.6.4). Recall that Einstein

---

352 See the discussion following equation (697) in sec. 5.6.1 for a possible explanation of why Einstein made this particular error.

353 To the left of equation (735), Einstein wrote $\begin{matrix}12\\4\end{matrix}$ to indicate that he was considering the 124 -component of equation (731).

had found the ϑ-metric on p. 42R as a solution of the equation $\vartheta_{\alpha\beta\lambda} = 0$, the analogue of equation (731) for the original ϑ-restriction. The ϑ-metric is obtained by interchanging co- and contravariant components of the rotation metric (see sec. 5.5.6). Despite considerable effort (see pp. 42R, 43LA, 24L and the discussion in secs. 5.5.7 and 5.5.9), Einstein had been unable to find a satisfactory physical interpretation for this metric.

A possible explanation for the reoccurrence of the ϑ-metric on p. 25R is that Einstein checked whether the $\hat{\vartheta}$-expression, like the original ϑ-expression, vanishes for the ϑ-metric. Since $g_{44} = 1$ for the ϑ-metric, equation (737), the 441-component of the equation $\hat{\vartheta}_{\alpha\beta\lambda} = 0$ which was not satisfied by the rotation metric, is trivially satisfied by the ϑ-metric. The ϑ-metric, however, does not satisfy equation (734), the 121-component of the equation $\hat{\vartheta}_{\alpha\beta\lambda} = 0$,[354]

$$\frac{\partial g_{12}}{\partial x_1} - \frac{\partial g_{11}}{\partial x_2} = -\omega^2 y - 2\omega^2 y = -3\omega^2 y. \tag{739}$$

So the $\hat{\vartheta}$-expression does not vanish for the ϑ-metric either. There would thus be no reason for Einstein to resume his efforts to make sense of this peculiar metric. But if Einstein, as we conjectured, did return to the ϑ-metric in this context, it may have given him the idea for another calculation involving the ϑ-metric, which can be found on the next two lines of p. 25R and which we shall turn to below.

5.6.4 *Tinkering with the Field Equations to Make Sure That the Rotation Metric Is a Solution (25R)* 25R

Underneath the material at the top of p. 25R discussed in the preceding subsection, in the two lines above the first horizontal line on p. 25R, Einstein once again wrote down the candidate for the left-hand side of the field equations that he had found on p. 24R (see equation (701))

$$\frac{\partial}{\partial x_\varepsilon}\left(\gamma_{\varepsilon i}\frac{\partial g_{\alpha\beta}}{\partial x_i}\right) - \frac{1}{2}\frac{\partial \gamma_{\lambda\tau}}{\partial x_\alpha}\frac{\partial g_{\lambda\tau}}{\partial x_\beta}. \tag{740}$$

At the bottom of p. 24R he had noted that the 11-component of this expression does not vanish for the rotation metric (see equation (702)). He now inserted the ϑ-metric, i.e., the rotation metric with its co- and contravariant components switched, into the 11-component of equation (740).[355] This gives

$$\frac{\partial}{\partial x_\varepsilon}\left(\gamma_{\varepsilon 2}\frac{\partial g_{11}}{\partial x_2}\right) - \frac{\partial \gamma_{24}}{\partial x_1}\frac{\partial g_{24}}{\partial x_1}. \tag{741}$$

The second term gives $-\omega^2$ for the ϑ-metric as it does for the rotation metric. The first term vanishes for the rotation metric since $g_{11} = -1$, but it does not for the ϑ-

354 Einstein's sign error in g_{12} changes the last two steps in equation (739) to $\omega^2 y - 2\omega^2 y = -\omega^2 y$.
355 For a possible connection between this calculation and the calculations at the top of p. 25R, see the discussion at the end of sec. 5.6.3.

metric which has $g_{11} = -1 + \omega^2 y^2$. In the case of the ϑ-metric, the first term in equation (741) can thus be written as

$$\frac{\partial}{\partial x_\varepsilon}(\gamma_{\varepsilon 2}\ 2\omega^2 y) = \frac{\partial}{\partial x_1}(\gamma_{12}\ 2\omega^2 y) + \frac{\partial}{\partial x_2}(\gamma_{22}\ 2\omega^2 y)\,. \tag{742}$$

Einstein initially confused co- and contravariant components of the ϑ-metric and substituted $\omega^2 xy$ for γ_{12} and $-1 + \omega^2 x^2$ (or rather $-1 + \omega x^2$) for γ_{22}. He subsequently corrected these errors and substituted 0 and -1 for γ_{12} and γ_{22}, respectively. In the notebook, equation (742) was thus written as

$$\frac{\partial}{\partial x_\varepsilon}(\gamma_{\varepsilon 2}\ 2\omega y) = \frac{\partial}{\partial x_1}(\underset{0}{\cancel{\omega^2 xy}} \cdot\ 2\ \omega^2 y) + \frac{\partial}{\partial y}((-1 \cancel{+\omega x^2})\ \cdot\ 2\omega^2 y\,. \tag{743}$$

Even with Einstein's corrections, this equation still contains some minor errors.[356] Einstein also seems to have dropped the minus sign in the second term on the right-hand side at this point. This can be inferred from the fact that he clearly was under the impression that the 11-component of expression (740) would vanish for the ϑ-metric, if only the coefficient of the second term of the expression were changed from $-1/2$ to -1.[357] Inserting the ϑ-metric into this modified version of expression (740) and substituting $2\omega^2$ instead of $-2\omega^2$ for its first term, one arrives at

$$\frac{\partial}{\partial x_\varepsilon}\left(\gamma_{\varepsilon i}\frac{\partial g_{11}}{\partial x_i}\right) - \frac{\partial \gamma_{\lambda\tau}}{\partial x_1}\frac{\partial g_{\lambda\tau}}{\partial x_1} = 2\omega^2 - 2\omega^2 = 0\,. \tag{744}$$

Einstein's sign error can easily be corrected. Rather than changing the coefficient $-1/2$ of the second term in expression (740) to -1, Einstein should have changed it to 1.

If an expression vanishes for the ϑ-metric, it is only a matter of interchanging co- and contravariant components of the metric to obtain an expression that vanishes for the rotation metric.[358] On the next line of p. 25R, after drawing a horizontal line, Einstein therefore wrote down the expression

$$\frac{\partial}{\partial x_\varepsilon}\left(g_{\varepsilon i}\frac{\partial \gamma_{\alpha\beta}}{\partial x_i}\right) - \frac{\partial \gamma_{\lambda\tau}}{\partial x_\alpha}\frac{\partial g_{\lambda\tau}}{\partial x_\beta}\,, \tag{745}$$

which is obtained by interchanging co- and contravariant components of the metric in expression (740) and changing the factor $1/2$ in the second term to 1. Einstein assumed that the 11-component of this new expression vanishes for the rotation metric. This expression thus seemed to be a promising new candidate for the left-hand side of the field equations.[359] Mathematically, however, it is ill-defined: α and β appear as contravariant indices in the first term and as covariant ones in the second, and the

356 On the left-hand side, ω should be ω^2 and in the last term on the right-hand side, a closing bracket is missing.

357 This in turn can be inferred from his comment "$\alpha = 1\ \ \beta = 1$ correct" ("$\alpha = 1\ \ \beta = 1$ stimmt") farther down on p. 25R (see the paragraph following expression (748)).

358 Einstein had used this same insight to construct a modified ϑ-expression on p. 43LA (see sec. 5.5.8)

summations over i and ε are summations over pairs of covariant indices. Einstein had concocted expression (745) to meet the requirement that the rotation metric be a vacuum solution of the field equations and had done so at the expense of basic mathematical requirements.

Einstein still had to check whether all components of expression (745) vanish for the rotation metric. So far, he had only convinced himself that the 11-component does. In principle, one would have to check six index combinations: 11, 12, 14, 22, 24, and 44.[360] Upon inspection of expression (745), however, one easily sees that the 24-component will be equal to the 14-component for this particular metric, and that the 22-component will be equal to the 11-component. It thus suffices to check whether the remaining four components vanish.

Einstein first considered the index combination "$\alpha = 4 \;\; \beta = 1$." The second term in expression (745) vanishes because the metric is time-independent. The first term can only contribute for "$i = 2$," as Einstein noted, since otherwise $\partial\gamma_{41}/\partial x_i = 0$. Even for $i = 2$, however, there will be no contribution, as Einstein also noted,

$$\frac{\partial}{\partial x_\varepsilon}(g_{\varepsilon 2}\omega) = 0\,. \tag{746}$$

Einstein then turned to the index combination "$\alpha = 4 \;\; \beta = 4$." The second term in expression (745) again vanishes because the metric is time-independent. Consider the first term

$$\frac{\partial}{\partial x_\varepsilon}\left(g_{\varepsilon i}\frac{\partial\gamma_{44}}{\partial x_i}\right). \tag{747}$$

Since $\gamma_{44} = 1$, this term also vanishes. Einstein initially seems to have inserted $\gamma_{44} = 1 - \omega^2(x^2 + y^2)$ instead, which would explain the expressions

$$\begin{array}{ll} i = 1 & \quad i = 2 \\ -2\omega^2 x, & \end{array} \tag{748}$$

which he wrote down and deleted. He then simply wrote: "vanishes" ("verschwindet").

For the next index combination, "$\alpha = 1 \;\; \beta = 1$," which he wrote down after drawing another horizontal line, Einstein just wrote: "correct" ("stimmt"). After all, he had constructed expression (745) so that its 11-component would vanish for the

---

359 On the face of it, it may seem somewhat dubious to simply change the coefficient of the second term in expression (740) to make the expression vanish for the ϑ-metric. Einstein knew, however, that there should be additional terms quadratic in first-order derivatives of the metric in expression (740) coming from the terms that he had neglected in its derivation. Einstein's hope at this point may have been that these correction terms would only result in a change of the coefficient of the second-term in expression (740) as it stood.

360 In the $3 + 1$-dimensional case, the components 13, 23, 33 and 34 trivially vanish. All other components are equal to the corresponding ones in the $2 + 1$-dimensional case.

rotation metric (see equation (744)). As we already pointed out, the 11-component does in fact not vanish,

$$\frac{\partial}{\partial x_\varepsilon}\left(g_{\varepsilon i}\frac{\partial \gamma_{11}}{\partial x_i}\right) - \frac{\partial \gamma_{\lambda\tau}}{\partial x_1}\frac{\partial g_{\lambda\tau}}{\partial x_1} = g_{22}\frac{\partial^2 \gamma_{11}}{\partial x_2^2} - 2\frac{\partial \gamma_{24}}{\partial x_1}\frac{\partial g_{24}}{\partial x_1} = -4\omega^2 . \tag{749}$$

Finally, Einstein considered the index combination "$\alpha = 1 \;\; \beta = 2$." One easily sees that the second term in equation (745) once again vanishes. Given the sign error in the 12-component of $\gamma_{\mu\nu}$ in expression (730), Einstein wrote the possible contributions coming from the first term in expression (745) as

$$\begin{array}{cc} i = 1 & i = 2 \\ \dfrac{\partial}{\partial x_\varepsilon}(g_{\varepsilon 1}\ \omega^2 y) & \dfrac{\partial}{\partial x_\varepsilon}(g_{\varepsilon 2}\ \omega^2 x). \end{array} \tag{750}$$

Both these terms vanish. Einstein could thus concluded that the 12-component of equation (745) also "vanishes" ("verschwindet") for the rotation metric.

Einstein had now checked all independent components of equation (745) and concluded: "Equation satisfied" ("Gleichung erfüllt"). Making the right errors in the right places, he had convinced himself that the ill-defined expression in equation (745) finally gave him field equations with the rotation metric as a vacuum solution, a test that so many other promising candidates had failed.

25R 5.6.5 *Testing the Newly Concocted Field Equations for Compatibility with Energy-Momentum Conservation (25R)*

In the last four lines of p. 25R, Einstein checked whether the ill-defined new field equations based on expression (745), which looked promising from the point of view of the equivalence principle, would satisfy the conservation principle as well. As he had done for several other candidate field equations,[361] he checked whether these field equations could be used to write the gravitational force density as the divergence of gravitational stress-energy density. A few lines of calculation show that this can easily be done in linear approximation, but that the equality cannot hold exactly. Einstein thereupon abandoned these field equations. In the course of this short calculation, Einstein may also have come to realize that the equations were mathematically unacceptable to begin with.

In linear approximation, the gravitational force density can be written as the contraction of a derivative of the metric with the left-hand side of the linearized field equations. If one uses the covariant field equations, one has to use the contravariant metric; if one uses the contravariant field equations, one has to use the covariant metric (cf. expressions (687) and (688)). Einstein, however, wrote the force density as a contraction of the first term in expression (745), which has contravariant free indices, and the derivative of the contravariant metric,

---

361 See p. 19R, 20L, 21L, 24L and 25L, and the discussion following equation (481) in sec. 5.4.2.

$$\frac{\partial \gamma_{\alpha\beta}}{\partial x_\sigma} \frac{\partial}{\partial x_\varepsilon}\left(g_{\varepsilon i} \frac{\partial \gamma_{\alpha\beta}}{\partial x_i}\right). \tag{751}$$

He thereby further compounded the problem that his new field equations are ill-defined. In expression (751) the summations over α and β are over pairs of contravariant indices, while the summations of over ε and i are over pairs of covariant ones.

Oblivious to these problems, it seems, Einstein proceeded to rewrite expression (751) in the form of a divergence. First he rewrote it as[362]

$$\frac{\partial}{\partial x_\varepsilon}\left(\frac{\partial \gamma_{\alpha\beta}}{\partial x_\sigma} g_{\varepsilon i} \frac{\partial \gamma_{\alpha\beta}}{\partial x_i}\right) - \frac{\partial^2 \gamma_{\alpha\beta}}{\partial x_\varepsilon \partial x_\sigma} g_{\varepsilon i} \frac{\partial \gamma_{\alpha\beta}}{\partial x_i} \tag{752}$$

The first term has the required form of a divergence, so Einstein could focus on the second term, which he rewrote as

$$-\frac{\partial}{\partial x_\sigma}\left(\frac{\partial \gamma_{\alpha\beta}}{\partial x_\varepsilon} g_{\varepsilon i} \frac{\partial \gamma_{\alpha\beta}}{\partial x_i}\right) + \frac{\partial \gamma_{\alpha\beta}}{\partial x_\varepsilon} \frac{\partial \gamma_{\alpha\beta}}{\partial x_i} \frac{\partial g_{\varepsilon i}}{\partial x_\sigma} + \frac{\partial \gamma_{\alpha\beta}}{\partial x_\varepsilon} g_{\varepsilon i} \frac{\partial^2 \gamma_{\alpha\beta}}{\partial x_i \partial x_\sigma}. \tag{753}$$

The first term again has the required form of a divergence. The second term is of third power in derivatives of the metric and vanishes in linear approximation. Rather than deleting this term, however, Einstein underlined it, a clear indication that he wanted to make this seemingly approximative calculation exact. But first he turned his attention to the third term in expression (753), which he rewrote as[363]

$$\frac{\partial}{\partial x_i}\left(\frac{\partial \gamma_{\alpha\beta}}{\partial x_\varepsilon} g_{\varepsilon i} \frac{\partial \gamma_{\alpha\beta}}{\partial x_\sigma}\right) - \frac{\partial \gamma_{\alpha\beta}}{\partial x_\sigma} \frac{\partial}{\partial x_\varepsilon}\left(g_{\varepsilon i} \frac{\partial \gamma_{\alpha\beta}}{\partial x_\varepsilon}\right). \tag{754}$$

He underlined both terms. This is the end of the calculation. Einstein was clearly dissatisfied with the result, for in the lower left corner of p. 25R, to the left of equations (753) and (754), he wrote: “impossible” (“unmöglich”).

The reason for Einstein’s dissatisfaction must have been that the calculation in equations (751)–(754) convinced him that his new candidate field equations are incompatible with the *exact* equality of the gravitational force density and the divergence of gravitational stress-energy density. In linear approximation the calculation seems to show that the equations are compatible with this equality.[364] The last term in expression (754) is equal and opposite to expression (751) that Einstein started from. Bringing this term to the left-hand side and dividing both sides by a factor 2, Einstein would have succeeded in writing his ill-defined expression for the gravitational force density as a sum of three terms that all have the required form of a divergence (i.e., the first terms in expressions (752), (753), and (754)) and a term of third power in deriv-

362 In the notebook, the expressions in parentheses in equations (752) and (753) as well as the first expression in parentheses in equation (754) are indicated by pairs of parentheses only.

363 One of the derivatives $\partial/\partial x_\varepsilon$ in the last term should be $\partial/\partial x_i$.

364 In fact, the calculation shows nothing of the sort, since both the field equations and the expression that was used for the gravitational force density in this calculation are mathematically ill-defined. Einstein, however, appears to have been unaware, at least initially, of these problems.

atives of the metric that vanishes in linear approximation (i.e., the second term in expression (753)). The equality would hold exactly if this last term could be interpreted as (minus) the contraction of $\partial\gamma_{\alpha\beta}/\partial x_\sigma$ with the term quadratic in first-order derivatives of the metric in the field equations. Comparison of the second term in expression (753) to the second term in expression (745), the left-hand side of the candidate field equations under consideration, shows that the former cannot be interpreted in this way,[365]

$$\frac{\partial\gamma_{\alpha\beta}}{\partial x_\varepsilon}\frac{\partial\gamma_{\alpha\beta}}{\partial x_i}\frac{\partial g_{\varepsilon i}}{\partial x_\sigma} \neq \frac{\partial\gamma_{\alpha\beta}}{\partial x_\sigma}\left(\frac{\partial g_{\lambda\tau}}{\partial x_\alpha}\frac{\partial\gamma_{\lambda\tau}}{\partial x_\beta}\right). \tag{755}$$

This is probably why Einstein concluded that it was “impossible” (“unmöglich”) to write the gravitational force density as the divergence of gravitational stress-energy density.

Einstein’s remark may also refer, at least in part, to the more fundamental problem that expressions (751)–(754) as well as the new candidate field equations themselves are mathematically ill-defined. Whatever the case may be, Einstein at this point abandoned these ill-defined field equations and on the next page (p. 26L) made a fresh start with the method for generating field equations automatically satisfying energy-momentum conservation that he had introduced on p. 24R.

19L–25R 5.7 *Conclusion: Cutting the Gordian Knot (19L–25R)*

Of all attempts to find suitable gravitational field equations recorded in the notebook, the one on p. 25R was clearly the most desperate. As was shown in detail in secs. 5.6.4–5.6.5, the attempt is riddled with errors. One can, however, also look upon Einstein’s calculations on this page in a more positive way. They reveal very clearly which of the various requirements that had to be satisfied by putative field equations weighed most heavily for Einstein at this point. Looking at p. 25R from this perspective, one is struck by the fact that, even though many of the expressions considered by Einstein were not even well-defined mathematically, he continued to adhere strictly to

365 It may be interesting to note that such an interpretation would have been possible had Einstein contracted $\frac{\partial}{\partial x_\varepsilon}\left(g_{\varepsilon i}\frac{\partial\gamma_{\alpha\beta}}{\partial x_i}\right)$ with $\frac{\partial g_{\alpha\beta}}{\partial x_\sigma}$, as he should have done, rather than with $\frac{\partial\gamma_{\alpha\beta}}{\partial x_\sigma}$ (cf. the discussion in the paragraph with equation (751) above). In that case, the second term in equation (753) would change to $\frac{\partial g_{\alpha\beta}}{\partial x_\varepsilon}\frac{\partial\gamma_{\alpha\beta}}{\partial x_i}\frac{\partial g_{\varepsilon i}}{\partial x_\sigma} = \left(\frac{\partial g_{\lambda\tau}}{\partial x_\alpha}\frac{\partial\gamma_{\lambda\tau}}{\partial x_\beta}\right)\frac{\partial g_{\alpha\beta}}{\partial x_\sigma}$, where the term in parentheses is indeed equal to minus the second term in equation (745). With this modification of the calculation in equations (751)–(754), however, the last term in equation (754) would change to $\frac{\partial\gamma_{\alpha\beta}}{\partial x_\sigma}\frac{\partial}{\partial x_\varepsilon}\left(g_{\varepsilon i}\frac{\partial\gamma_{\alpha\beta}}{\partial x_i}\right)$ and would no longer be equal to the expression in equation (751). Consequently, the equality of gravitational force density and divergence of gravitational stress-energy density would not even hold in linear approximation. It seems unlikely that Einstein considered this alternative.

three requirements. First and foremost, there was the correspondence principle. The field equations should consist of a core operator, i.e., a term with second-order derivatives of the metric that reduces to the d'Alembertian acting on the metric in linear approximation, and terms quadratic in first-order derivatives of the metric that vanish in linear approximation. Then there was the conservation principle in the very specific form that the field equations should make it possible, not just in linear approximation but exactly, to write the gravitational force density as the divergence of gravitational stress-energy density. Finally, there was the demand that the rotation metric be a vacuum solution of the field equations. Einstein was willing, it seems, to weaken, if only temporarily, the much stronger demands of his relativity and equivalence principles that the field equations be invariant under (autonomous or non-autonomous) transformations to arbitrarily accelerated frames of reference. The focus on these three requirements, which constrain even the very problematic calculations on p. 25R, made the task of finding suitable field equations much more manageable.

The strategy that Einstein had used on pp. 19L–23L and again on p. 25L and at the top of p. 25R of extracting field equations from expressions of broad covariance by imposing coordinate restrictions had failed several times. With hindsight, one can see that most of the problems come from Einstein using coordinate restrictions rather than coordinate conditions in the modern sense to make sure that the field equations satisfy the correspondence principle. Einstein had found three different coordinate restrictions (the harmonic restriction, the Hertz restriction, and the ϑ-restriction) with the help of which the correspondence principle could be satisfied. The covariance properties of these coordinate restrictions, however, proved to be intractable. Since Einstein used coordinate restrictions rather than coordinate conditions, this meant that the covariance properties of the field equations themselves became intractable as well. It thus remained unclear whether these equations satisfy the demands of the equivalence and relativity principles. Moreover, Einstein had not been able to confirm that the coordinate restrictions and the associated field equations would be compatible with the exact validity of energy-momentum conservation.

In this situation something had to give. Einstein, it seems, cut the Gordian knot by weakening two of his heuristic principles. He weakened the equivalence principle to the requirement that the field equations at least allow the rotation metric. He weakened the relativity principle to the obvious minimal requirement that the field equations at least have well-defined transformation properties. On p. 24R Einstein had found candidate field equations imposing the correspondence and conservation principles and bracketing the problem of satisfying the remaining two principles. On pp. 26L–R, he turned this derivation into a powerful method for generating field equations that automatically meet the requirements of the correspondence and conservation principles. What probably made this option all the more appealing was that Einstein believed (mistakenly as was shown in sec. 5.6.1) that the expression that forms the starting point of the calculation on p. 24R vanishes exactly for the rotation metric, a necessary condition for the resulting field equations to satisfy the weakened version of the equivalence principle.

26L–R

6. DERIVATION OF THE *ENTWURF* EQUATIONS (26L–R)

On pp. 26L–R, Einstein derived the identity that is at the heart of the derivation of the *Entwurf* field equations in (Einstein and Grossmann 1913). His approach on these pages is very similar to his approach on p. 24R (cf. the discussion following expression (698) in sec. 5.6.1). He substituted the left-hand side of some linearized field equations for the stress-energy tensor $T_{\mu\nu}$ in the expression

$$\frac{\partial g_{\mu\nu}}{\partial x_m}T_{\mu\nu} \tag{756}$$

for the gravitational force density and tried to rewrite the resulting expression as the divergence of the gravitational stress-energy density. In addition to divergence terms, he found terms that are contractions of $\partial g_{\mu\nu}/x_m$ and terms quadratic in first-order derivatives of the metric. By adding the latter to the linearized field equations, Einstein arrived at exact field equations that guarantee that the gravitational force density is exactly equal to the divergence of the gravitational stress-energy density.

The calculation on pp. 26L–R is thus very similar to the calculation on p. 24R, but differs from it in two important respects. First, Einstein no longer set the determinant of the metric equal to unity, which can be seen as an indication that he had meanwhile given up hope to recover the new candidate field equations from the November tensor as he had still tried to do on p. 25L for the field equations of p. 24R. Second, Einstein did not start, as he had done on p. 24R (see expression (683)) from a specific expression for (the divergence of) the gravitational stress-energy pseudo-tensor. Einstein read off both the expression for the left-hand side of the field equations and the expression for the gravitational stress-energy pseudo-tensor from the identity obtained by rewriting expression (756) after substitution of the left-hand side of the linearized field equations for the stress-energy tensor.

At the top of p. 26L, Einstein began by writing down the energy-momentum balance between matter and gravitational field,[366]

$$\frac{\partial}{\partial x_n}(\sqrt{G}g_{\mu\nu}T_{\nu n}) - \frac{1}{2}\sqrt{G}\frac{\partial g_{\mu\nu}}{\partial x_m}T_{\mu\nu} = 0\,, \tag{757}$$

or, as he called it, the “system of the equations for matter” (“System der Gleichungen für Materie”) and, on the next line, the (contravariant) stress-energy tensor for pressureless dust,

$$T_{\mu\nu} = \rho\frac{dx_\mu}{d\tau}\frac{dx_\nu}{d\tau}\,. \tag{758}$$

Equation (757) shows that the exact expression for the gravitational force density is

$$-\frac{1}{2}\sqrt{G}\frac{\partial g_{\mu\nu}}{\partial x_m}T_{\mu\nu}\,, \tag{759}$$

366 In the first term of equation (757), $g_{\mu\nu}$ should be $g_{m\nu}$ (see equation (71)).

but in the calculations on the remainder of the page Einstein ignored the factor $-\frac{1}{2}\sqrt{G}$, a simplification that is easily corrected for at the end of the calculation.

Under the heading "derivation of the gravitation equations" ("Ableitung der Gravitationsgleichungen"), Einstein now substituted the left-hand side of linearized field equations based on a core operator,

$$\frac{\partial}{\partial x_\alpha}\left(\gamma_{\alpha\beta}\sqrt{G}\frac{\partial\gamma_{\mu\nu}}{\partial x_\beta}\right) = \kappa T_{\mu\nu}, \tag{760}$$

for the stress-energy tensor in the crude expression (756) for the force density,

$$\frac{\partial g_{\mu\nu}}{\partial x_m}\frac{\partial}{\partial x_\alpha}\left(\gamma_{\alpha\beta}\sqrt{G}\frac{\partial\gamma_{\mu\nu}}{\partial x_\beta}\right), \tag{761}$$

also ignoring, as he had done on several earlier occasions (see pp. 19R, 20L, 21L, 24R, and 25R), the gravitational constant κ. Einstein set out to rewrite expression (761) as a sum of terms of two kinds, divergence terms (marked '+' and '×' on pp. 26L–R) and terms that are contractions with $\partial g_{\mu\nu}/x_m$ (marked 'o' on pp. 26L–R[367]).

Einstein first rewrote expression (761) as

$$\left.\frac{\partial g_{\mu\nu}}{\partial x_m}\right|\frac{\partial}{\partial x_\alpha}\left(\gamma_{\alpha\beta}\sqrt{G}\frac{\partial\gamma_{\mu\nu}}{\partial x_\beta}\right) = \frac{\partial}{\partial x_\alpha}\left(\gamma_{\alpha\beta}\sqrt{G}\frac{\partial\gamma_{\mu\nu}}{\partial x_\beta}\frac{\partial g_{\mu\nu}}{\partial x_m}\right) - \sqrt{G}\gamma_{\alpha\beta}\frac{\partial\gamma_{\mu\nu}}{\partial x_\beta}\frac{\partial^2 g_{\mu\nu}}{\partial x_m\partial x_\alpha}. \tag{762}$$

The first term on the right-hand side is a divergence term. It is underlined and marked '+.' Einstein proceeded to rewrite the second term as

$$-\frac{\partial}{\partial x_m}\left(\gamma_{\alpha\beta}\sqrt{G}\frac{\partial\gamma_{\mu\nu}}{\partial x_\beta}\frac{\partial g_{\mu\nu}}{\partial x_\alpha}\right) + \frac{\partial g_{\mu\nu}}{\partial x_\alpha}\frac{\partial}{\partial x_m}\left(\sqrt{G}\gamma_{\alpha\beta}\frac{\partial\gamma_{\mu\nu}}{\partial x_\beta}\right). \tag{763}$$

Once again, the first term is a divergence term, which is underlined and marked '×.' The second term gives three terms:

$$\frac{1}{2}\frac{\partial g_{\sigma\tau}}{\partial x_m}\gamma_{\sigma\tau}\sqrt{G}\frac{\partial g_{\mu\nu}}{\partial x_\alpha}\frac{\partial\gamma_{\mu\nu}}{\partial x_\beta}\gamma_{\alpha\beta} + \frac{\partial\gamma_{\alpha\beta}}{\partial x_m}\sqrt{G}\frac{\partial g_{\mu\nu}}{\partial x_\alpha}\frac{\partial\gamma_{\mu\nu}}{\partial x_\beta} + \frac{\partial g_{\mu\nu}}{\partial x_\alpha}\sqrt{G}\gamma_{\alpha\beta}\frac{\partial^2\gamma_{\mu\nu}}{\partial x_m\partial x_\beta}. \tag{764}$$

This expression extends beyond the right margin of p. 26L and is continued on the facing page, p. 26R. The first term in expression (764) is obtained with the help of the relation

$$\frac{\partial\sqrt{G}}{\partial x_m} = \frac{1}{2\sqrt{G}}\frac{\partial G}{\partial x_m} = \frac{1}{2\sqrt{G}}\left(G\gamma_{\sigma\tau}\frac{\partial g_{\sigma\tau}}{\partial x_m}\right), \tag{765}$$

which Einstein had encountered several times before (see pp. 6L, 8R, and 9R, and equation (87)). By relabeling indices ($\sigma\tau\mu\nu \to \mu\nu\sigma\tau$), Einstein could show that this term can be written as a contraction of $\partial g_{\mu\nu}/x_m$ and a term quadratic in first-order derivatives of the metric,

367 Except in one case (see expression (766) below) which Einstein apparently forgot to mark.

$$\frac{1}{2}\frac{\partial g_{\mu\nu}}{\partial x_m}\gamma_{\mu\nu}\sqrt{G}\frac{\partial g_{\sigma\tau}}{\partial x_\alpha}\frac{\partial \gamma_{\sigma\tau}}{\partial x_\beta}\gamma_{\alpha\beta}\,. \tag{766}$$

The term is underlined, but Einstein for some reason neglected to mark it 'o' as he did with other terms of this kind.

The second term in expression (764) is of the same kind, although it takes a little more work to show this. Substituting $(\partial/\partial x_\alpha) = g_{\alpha\sigma}\gamma_{\alpha'\sigma}(\partial/\partial x_{\alpha'})$ and $(\partial/\partial x_\beta) = g_{\beta\tau}\gamma_{\beta'\tau}(\partial/\partial x_{\beta'})$, Einstein rewrote this term as

$$\frac{\partial \gamma_{\alpha\beta}}{\partial x_m}g_{\alpha\sigma}g_{\beta\tau}\;\big|\;\gamma_{\alpha'\sigma}\gamma_{\beta'\tau}\;\;\sqrt{G}\frac{\partial g_{\mu\nu}}{\partial x_{\alpha'}}\frac{\partial \gamma_{\mu\nu}}{\partial x_{\beta'}}\,, \tag{767}$$

which, with the help of the relation $(\partial\gamma_{\alpha\beta}/\partial x_m)g_{\alpha\sigma}g_{\beta\tau} = -(\partial g_{\sigma\tau}/\partial x_m)$, he then rewrote as

$$-\frac{\partial g_{\sigma\tau}}{\partial x_m}\sqrt{G}\gamma_{\alpha\sigma}\gamma_{\beta\tau}\frac{\partial g_{\mu\nu}}{\partial x_\alpha}\frac{\partial \gamma_{\mu\nu}}{\partial x_\beta}\,. \tag{768}$$

So, this term does indeed also have the form of a contraction of $\partial g_{\mu\nu}/\partial x_m$ and a term quadratic in first-order derivatives of the metric. It was underlined and marked 'o' accordingly.

At the bottom of p. 26R, Einstein turned to the third term in expression (764). This term contributes one '+'-term and two 'o'-terms. Einstein began by rewriting it as

$$\frac{\partial}{\partial x_\beta}\left(\frac{\partial \gamma_{\mu\nu}}{\partial x_m}\frac{\partial g_{\mu\nu}}{\partial x_\alpha}\sqrt{G}\gamma_{\alpha\beta}\right) - \frac{\partial \gamma_{\mu\nu}}{\partial x_m}\frac{\partial}{\partial x_\beta}\left(\sqrt{G}\gamma_{\alpha\beta}\frac{\partial g_{\mu\nu}}{\partial x_\alpha}\right)\,. \tag{769}$$

The first term is a divergence term. It is underlined and marked '+'. Rewriting it as

$$\frac{\partial}{\partial x_\alpha}\left(\frac{\partial g_{\mu\nu}}{\partial x_m}\frac{\partial \gamma_{\mu\nu}}{\partial x_\beta}\sqrt{G}\gamma_{\beta\alpha}\right)\,, \tag{770}$$

one sees that it is equal to this term in equation (762).

The second term in expression (769) is a contraction with $\partial\gamma_{\mu\nu}/x_m$ rather than with $\partial g_{\mu\nu}/x_m$. Einstein therefore rewrote it as (cf. expression (767))[368]

$$-\frac{\partial \gamma_{\mu\nu}}{\partial x_m}g_{\mu\sigma}g_{\nu\tau}\big|\gamma_{\mu'\sigma}\gamma_{\nu'\tau}\frac{\partial}{\partial x_\beta}\left(\sqrt{G}\gamma_{\alpha\beta}\frac{\partial g_{\mu\nu}}{\partial x_\alpha}\right)\,, \tag{771}$$

and then set $-(\partial\gamma_{\mu\nu}/\partial x_m)g_{\mu\sigma}g_{\nu\tau}$ equal to $\partial g_{\sigma\tau}/\partial x_m$. His next move was to bring $\gamma_{\mu'\sigma}$ and $\gamma_{\nu'\tau}$ within the scope of the differentiation $\partial/\partial x_\beta$ in order to turn the derivative of $g_{\mu\nu}$ into a derivative of $\gamma_{\mu\nu}$ (as in expression (761)). Expression (771) thus turns into[369]

368 In expression (771), $g_{\mu\nu}$ should be $g_{\mu'\nu'}$.

369 Expression (772) originally had $\gamma_{\mu'\sigma}\gamma_{\nu'\tau}$, but the primes were crossed out.

$$\frac{\partial g_{\sigma\tau}}{\partial x_m}\left|\frac{\partial}{\partial x_\beta}\left(\gamma_{\mu\sigma}\gamma_{\nu\tau}\sqrt{G}\gamma_{\alpha\beta}\frac{\partial g_{\mu\nu}}{\partial x_\alpha}\right)-\sqrt{G}\gamma_{\alpha\beta}\frac{\partial g_{\mu\nu}}{\partial x_\alpha}\frac{\partial \gamma_{\mu\sigma}\gamma_{\nu\tau}}{\partial x_\beta}\right. . \tag{772}$$

He then rewrote the two terms to the right of the vertical line as[370]

$$-\frac{\partial}{\partial x_\beta}\left(\frac{\partial \gamma_{\sigma\tau}}{\partial x_\alpha}\gamma_{\alpha\beta}\sqrt{G}\right)+2\sqrt{G}\gamma_{\alpha\beta}g_{\mu\nu}\frac{\partial \gamma_{\mu\sigma}}{\partial x_\alpha}\frac{\partial \gamma_{\nu\tau}}{\partial x_\beta} . \tag{773}$$

He underlined both terms and marked them 'o.' The first term is (minus) the core operator that formed the starting point of this whole calculation (cf. equations (760)–(761)).

Einstein now collected the 'o'-terms on the left-hand side and the '+/×'-terms on the right-hand side. The 'o'-terms come from the left-hand side of equation (762) (with a factor 2 because of the identical contribution from the first term in expression (773)), from expressions (766) and (768), and from the second term in expression (773) (contracted with $\partial g_{\sigma\tau}/\partial x_m$). These terms add up to

$$\begin{aligned}\frac{\partial g_{\mu\nu}}{\partial x_m}\Bigg[&2\frac{\partial}{\partial x_\alpha}\left(\gamma_{\alpha\beta}\sqrt{G}\frac{\partial \gamma_{\mu\nu}}{\partial x_\beta}\right)-\frac{1}{2}\sqrt{G}\gamma_{\mu\nu}\gamma_{\alpha\beta}\frac{\partial g_{\sigma\tau}}{\partial x_\alpha}\frac{\partial \gamma_{\sigma\tau}}{\partial x_\beta}\\&+\sqrt{G}\gamma_{\alpha\mu}\gamma_{\beta\nu}\frac{\partial g_{\sigma\tau}}{\partial x_\alpha}\frac{\partial \gamma_{\sigma\tau}}{\partial x_\beta}-2\sqrt{G}\gamma_{\alpha\beta}g_{\sigma\tau}\frac{\partial \gamma_{\mu\sigma}}{\partial x_\alpha}\frac{\partial \gamma_{\nu\tau}}{\partial x_\beta}\Bigg].\end{aligned} \tag{774}$$

The '+/×'-terms come from the first term on the right-hand side of equation (762) (with a factor 2 because of the identical contribution from expression (770)) and from the first term in expression (763):

$$2\frac{\partial}{\partial x_\alpha}\left(\sqrt{G}\gamma_{\alpha\beta}\frac{\partial \gamma_{\sigma\tau}}{\partial x_\beta}\frac{\partial g_{\sigma\tau}}{\partial x_m}\right)-\frac{\partial}{\partial x_m}\left(\sqrt{G}\gamma_{\alpha\beta}\frac{\partial g_{\sigma\tau}}{\partial x_\alpha}\frac{\partial \gamma_{\sigma\tau}}{\partial x_\beta}\right) . \tag{775}$$

Regrouping the 'o'-terms in expression (774),[371] setting them equal to the '+/×'-terms in expression (775), and dividing both sides by 2, one arrives at the identity

---

370 For the first term, he used the relation $\gamma_{\mu\sigma}\gamma_{\nu\tau}\frac{\partial g_{\mu\nu}}{\partial x_\alpha} = -\frac{\partial \gamma_{\sigma\tau}}{\partial x_\alpha}$. The second term can be rewritten as a sum of two identical terms

$$\begin{aligned}-\sqrt{G}\gamma_{\alpha\beta}\frac{\partial g_{\mu\nu}}{\partial x_\alpha}\frac{\partial(\gamma_{\mu\sigma}\gamma_{\nu\tau})}{\partial x_\beta} &= -\sqrt{G}\gamma_{\alpha\beta}\frac{\partial g_{\mu\nu}}{\partial x_\alpha}\frac{\partial \gamma_{\mu\sigma}}{\partial x_\beta}\gamma_{\nu\tau}-\sqrt{G}\gamma_{\alpha\beta}\frac{\partial g_{\mu\nu}}{\partial x_\alpha}\frac{\partial \gamma_{\nu\tau}}{\partial x_\beta}\gamma_{\mu\sigma}\\&= \sqrt{G}\gamma_{\alpha\beta}\frac{\partial \gamma_{\nu\tau}}{\partial x_\alpha}\frac{\partial \gamma_{\mu\sigma}}{\partial x_\beta}g_{\mu\nu}+\sqrt{G}\gamma_{\alpha\beta}\frac{\partial \gamma_{\mu\sigma}}{\partial x_\alpha}\frac{\partial \gamma_{\nu\tau}}{\partial x_\beta}g_{\mu\nu}.\end{aligned}$$

371 The physical reasoning behind this regrouping will become clear below.

$$\begin{aligned}&\frac{\partial g_{\mu\nu}}{\partial x_m}\Bigg[\left(\frac{\partial}{\partial x_\alpha}\left(\gamma_{\alpha\beta}\sqrt{G}\frac{\partial\gamma_{\mu\nu}}{\partial x_\beta}\right)-\sqrt{G}\gamma_{\alpha\beta}g_{\sigma\tau}\frac{\partial\gamma_{\mu\sigma}}{\partial x_\alpha}\frac{\partial\gamma_{\nu\tau}}{\partial x_\beta}\right)\\&+\frac{1}{2}\sqrt{G}\left(\gamma_{\alpha\mu}\gamma_{\beta\nu}\frac{\partial g_{\sigma\tau}}{\partial x_\alpha}\frac{\partial\gamma_{\sigma\tau}}{\partial x_\beta}-\frac{1}{2}\gamma_{\mu\nu}\gamma_{\alpha\beta}\frac{\partial g_{\sigma\tau}}{\partial x_\alpha}\frac{\partial\gamma_{\sigma\tau}}{\partial x_\beta}\right)\Bigg]\\&=\frac{\partial}{\partial x_\alpha}\left(\sqrt{G}\gamma_{\alpha\beta}\frac{\partial\gamma_{\sigma\tau}}{\partial x_\beta}\frac{\partial g_{\sigma\tau}}{\partial x_m}\right)-\frac{1}{2}\frac{\partial}{\partial x_m}\left(\sqrt{G}\gamma_{\alpha\beta}\frac{\partial g_{\sigma\tau}}{\partial x_\alpha}\frac{\partial\gamma_{\sigma\tau}}{\partial x_\beta}\right),\end{aligned}\tag{776}$$

which Einstein wrote down at the bottom of pp. 26L–R as the "summary" ("Zusammenfassung") of his calculations. Underneath this identity, he wrote: "This is the contra-form" ("Dies ist die Kontra-Form"). It is from this identity—given as equation (12) in Einstein's part of (Einstein and Grossmann 1913) and derived in sec. 4.3 of Grossmann's part—that both the contravariant form of the field equations and the contravariant form of the gravitational stress-energy pseudo-tensor of the *Entwurf* theory can be read off.

To identify the exact expressions for the field equations and the pseudo-tensor we need the exact relations between the left-hand side of the field equations, the gravitational force density, and the gravitational stress-energy pseudo-tensor. We can no longer afford to ignore factors $\sqrt{G}$. Consider once again equation (757), the vanishing of the covariant divergence of the mixed tensor density $\sqrt{G}g_{m\nu}T_{\nu n}$. This equation can be written as the *ordinary* divergence of the sum of the stress-energy density of matter and gravitational field if

$$-\frac{1}{2}\sqrt{G}\frac{\partial g_{\mu\nu}}{\partial x_m}T_{\mu\nu} = \frac{\partial}{\partial x_n}(\sqrt{G}g_{m\nu}t_{\nu n}),\tag{777}$$

where $t_{\mu\nu}$ is the contravariant form of the gravitational stress-energy pseudo-tensor. Multiplying both sides by -2κ, one arrives at

$$\sqrt{G}\frac{\partial g_{\mu\nu}}{\partial x_m}(\kappa T_{\mu\nu}) = \frac{\partial}{\partial x_n}(\sqrt{G}g_{m\nu}(-2\kappa t_{\nu n})).\tag{778}$$

The identity (776) guarantees that this equation holds if both the left-hand side of the field equations (to be substituted for $\kappa T_{\mu\nu}$ in equation (778)) and (-2κ times) the pseudo-tensor $t_{\mu\nu}$ are suitably chosen. Comparing the left-hand sides of equations (776) and (778), one sees that one must choose the expression

$$\begin{aligned}&\frac{1}{\sqrt{G}}\frac{\partial}{\partial x_\alpha}\left(\gamma_{\alpha\beta}\sqrt{G}\frac{\partial\gamma_{\mu\nu}}{\partial x_\beta}\right)-\gamma_{\alpha\beta}g_{\sigma\tau}\frac{\partial\gamma_{\mu\sigma}}{\partial x_\alpha}\frac{\partial\gamma_{\nu\tau}}{\partial x_\beta}\\&+\frac{1}{2}\left(\gamma_{\alpha\mu}\gamma_{\beta\nu}\frac{\partial g_{\sigma\tau}}{\partial x_\alpha}\frac{\partial\gamma_{\sigma\tau}}{\partial x_\beta}-\frac{1}{2}\gamma_{\mu\nu}\gamma_{\alpha\beta}\frac{\partial g_{\sigma\tau}}{\partial x_\alpha}\frac{\partial\gamma_{\sigma\tau}}{\partial x_\beta}\right),\end{aligned}\tag{779}$$

i.e., the expression in square brackets in expression (776) divided by $\sqrt{G}$, as the left-hand side of the field equations. This is indeed the left-hand side of the *Entwurf* field equations. To bring the right-hand side of equation (776) in form that can be compared with the right-hand side of equation (778), one has to relabel the summation index α by n, substitute $(\partial/\partial x_m) = g_{m\nu}\gamma_{\alpha\nu}(\partial/\partial x_\alpha)$ in the first term, and substitute $(\partial/\partial x_m) = g_{m\nu}\gamma_{n\nu}(\partial/\partial x_n)$ in the second term. Comparing the resulting expression,

$$\frac{\partial}{\partial x_n}\left(\sqrt{G}g_{m\nu}\left[\gamma_{n\beta}\gamma_{\alpha\nu}\frac{\partial\gamma_{\sigma\tau}}{\partial x_\beta}\frac{\partial g_{\sigma\tau}}{\partial x_\alpha}-\frac{1}{2}\gamma_{n\nu}\gamma_{\alpha\beta}\frac{\partial g_{\sigma\tau}}{\partial x_\alpha}\frac{\partial\gamma_{\sigma\tau}}{\partial x_\beta}\right]\right), \tag{780}$$

to the right-hand side of equation (778), one sees that one must choose the expression

$$\gamma_{\alpha\nu}\gamma_{\beta n}\frac{\partial g_{\sigma\tau}}{\partial x_\alpha}\frac{\partial\gamma_{\sigma\tau}}{\partial x_\beta}-\frac{1}{2}\gamma_{\nu n}\gamma_{\alpha\beta}\frac{\partial g_{\sigma\tau}}{\partial x_\alpha}\frac{\partial\gamma_{\sigma\tau}}{\partial x_\beta} \tag{781}$$

as the gravitational stress-energy pseudo-tensor $-2\kappa t_{\nu n}$. This is the definition of this quantity given in equation (13) of (Einstein and Grossmann 1913).[372] Notice that the expression in parentheses on the second line of expression (779) is equal to $-2\kappa t_{\mu\nu}$. If we now introduce the notation $D_{\mu\nu}(\gamma)$ for the expression on the first line,[373] the left-hand side of the field equations can be written as $D_{\mu\nu}(\gamma)-\kappa t_{\mu\nu}$, and the field equations themselves as

$$D_{\mu\nu}(\gamma) = \kappa(T_{\mu\nu}+t_{\mu\nu}), \tag{782}$$

which is the form in which the *Entwurf* field equations are given in equation (18) of (Einstein and Grossmann 1913).[374] Commenting on this equation (ibid., p. 17), Einstein pointed out that any acceptable field equations must be such that the stress-energy of matter and the stress-energy of the gravitational field enter the equations in the exact same way.[375]

Before pp. 26L–26R, this requirement had not explicitly played a role in Einstein's search for suitable field equations. But the way in which Einstein grouped the terms on the left-hand side of the identity (776)—with the terms in the first set of parentheses giving $D_{\mu\nu}(\gamma)$ and those in the second giving $-2\kappa t_{\mu\nu}$ — suggests that Einstein did consider this requirement when he wrote down the calculations on pp. 26L–26R.

One can understand why Einstein would have been pleased with these new candidate field equations. They satisfied the correspondence principle and they satisfied the conservation principle, not just in linear approximation but exactly. Since the new field equations were not extracted from some quantity of broad covariance with the help of a suitable coordinate restriction, it remained unclear whether they satisfy the relativity principle and the equivalence principle. The one encouraging result on this score (erroneous as it turns out) was that the divergence of the gravitational stress-energy pseudo-tensor vanishes for the rotation metric, a necessary condition for this

372 In the paper, the contravariant form is denoted by $\vartheta_{\mu\nu}$, and the covariant form by $t_{\mu\nu}$.

373 This is in keeping with (Einstein and Grossmann 1913), where this same quantity is introduced in equation (15) as $\Delta_{\mu\nu}(\gamma)$.

374 In (Einstein and Grossmann 1913), they are written as $\Delta_{\mu\nu}(\gamma) = \kappa(\Theta_{\mu\nu}+\vartheta_{\mu\nu})$.

375 A similar result holds in the final version of Einstein's earlier theory for static fields (see Einstein 1912b, 457) and in general relativity in its final form (see Einstein 1916, 807–808).

metric to be a vacuum solution of the *Entwurf* field equations.[376] More importantly, Einstein had not been able to find any acceptable field equations along the lines of his mathematical strategy that satisfied the relativity principle and the equivalence principle. And he had at best been able to show in linear approximation that these candidates satisfied the conservation principle. It is thus not surprising that Einstein gave up his attempt to construct field equations out of the Riemann tensor and decided to publish, in his joint paper with Marcel Grossmann, the field equations found along the lines of his physical strategy.

ACKNOWLEDGMENTS

The first draft of this commentary was prepared in the early 1990s by a working group of the *Arbeitsstelle Albert Einstein* (see preface). The working group—joined early on in its work on the notebook by John Norton and John Stachel and later by Michel Janssen—did not have to start from scratch. It had John Norton's "How did Einstein find his field equations?" (Norton 1984) to go on, as well as drafts of the annotation of the notebook prepared for Vol. 4 of *The Collected Papers of Albert Einstein*, both the footnotes added to the transcription (CPAE 4, Doc. 10) and the editorial note, "Einstein's Research Notes on a Generalized Theory of Relativity" (CPAE 4, 192–199). The joint work was continued in a number of workshops hosted by the *Max-Planck-Institut für Bildungsforschung* and later by the *Max-Planck-Institut für Wissenschaftsgeschichte*. On the basis of all this material, Michel Janssen and Jürgen Renn wrote the final version of the commentary (with the exception of sec. 4.5.8, which was written by Tilman Sauer). They produced the first half in close collaboration, partly at the University of Minnesota, partly (courtesy of Bernhard Schutz) at the *Max-Planck-Institut für Gravitationsphysik* (*Albert Einstein Institut*) in Golm. Michel Janssen wrote the second half (secs. 5 and 6) during extended stays at the *Max-Planck-Institut für Wissenschaftsgeschichte* in the late 1990s in close interaction with the scientists at the institute.

REFERENCES

Bergmann, Peter G., and Arthur Komar. 1972. "The Coordinate Group Symmetries of General Relativity." *International Journal of Theoretical Physics* 5: 15–28.

Bianchi, Luigi. 1910. *Vorlesungen über Differentialgeometrie*. 2nd expanded and rev. ed. Max Lukat (trans.). Leipzig: Teubner.

Budde, E. 1914. *Tensoren und Dyaden im dreidimensionalen Raum. Ein Lehrbuch*. Braunschweig: Vieweg.

Christoffel, Elwin Bruno. 1869. "Ueber die Transformation der homogenen Differentialausdrücke zweiten Grades." *Journal für die reine und angewandte Mathematik* 70: 46–70.

CPAE 1: John Stachel, David C. Cassidy, Robert Schulmann, and Jürgen Renn (eds.), *The Collected Papers of Albert Einstein*. Vol. 1. *The Early Years, 1879–1902*. Princeton: Princeton University Press, 1987.

CPAE 2: John Stachel, David C. Cassidy, Robert Schulmann, and Jürgen Renn (eds.), *The Collected Papers of Albert Einstein*. Vol. 2. *The Swiss Years: Writings, 1900–1909*. Princeton: Princeton University Press, 1989.

376 See p. 24R, expression (683) and the discussion in sec. 5.6.1 following expression (693).

CPAE 3: Martin J. Klein, A. J. Kox, Jürgen Renn, and Robert Schulmann (eds.), *The Collected Papers of Albert Einstein*. Vol. 3. *The Swiss Years: Writings, 1909–1911*. Princeton: Princeton University Press, 1993.

CPAE 4: Martin J. Klein, A. J. Kox, Jürgen Renn, and Robert Schulmann (eds.), *The Collected Papers of Albert Einstein*. Vol. 4. *The Swiss Years: Writings, 1912–1914*. Princeton: Princeton University Press, 1995.

CPAE 5: Martin J. Klein, A. J. Kox, and Robert Schulmann (eds.), *The Collected Papers of Albert Einstein*. Vol. 5. *The Swiss Years: Correspondence, 1902–1914*. Princeton: Princeton University Press, 1993.

CPAE 6: A. J. Kox, Martin J. Klein, and Robert Schulmann (eds.), *The Collected Papers of Albert Einstein*. Vol. 6. *The Berlin Years: Writings, 1914–1917*. Princeton: Princeton University Press, 1996.

CPAE 8: Robert Schulmann, A. J. Kox, Michel Janssen, and József Illy (eds.), *The Collected Papers of Albert Einstein*. Vol. 8. *The Berlin Years: Correspondence, 1914–1918*. Princeton: Princeton University Press, 1998.

Earman, John, and Michel Janssen. 1993. “Einstein’s Explanation of the Motion of Mercury’s Perihelion.” In (Earman et al. 1993, 129–172).

Earman, John, Michel Janssen, and John D. Norton (eds.). 1993. *The Attraction of Gravitation: New Studies in the History of General Relativity* (*Einstein Studies,* Vol. 5). Boston: Birkhäuser.

Einstein, Albert. 1907. “Über das Relativitätsprinzip und die aus demselben gezogenen Folgerungen.” *Jahrbuch der Radioaktivität und Elektronik* 4: 411–462 (CPAE 2, Doc. 47).

———. 1912a. “Lichtgeschwindigkeit und Statik des Gravitationsfeldes.” *Annalen der Physik* 38: 355–369 (CPAE 4, Doc. 3).

———. 1912b. “Zur Theorie des statischen Gravitationsfeldes.” *Annalen der Physik* 38 (1912): 443–458 (CPAE 4, Doc. 4).

———. 1913. “Zum gegenwärtigen Stande des Gravitationsproblems.” *Physikalische Zeitschrift* 14: 1249–1262 (CPAE 4, Doc. 17). (English translation in vol. 3 of this series.)

———. 1914a.“Bemerkungen” *Zeitschrift für Mathematik und Physik* 62: 260–261 (CPAE 4, Doc. 26).

———. 1914b. “Die formale Grundlage der allgemeinen Relativitätstheorie.” *Königlich Preußische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte*: 1030–1085 (CPAE 6, Doc. 9).

———. 1915a. “Zur allgemeinen Relativitätstheorie.” *Königlich Preußische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte*: 778–786. (CPAE 6, Doc. 21).

———. 1915b. “Zur allgemeinen Relativitätstheorie. (Nachtrag).” *Königlich Preußische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte*: 799–801. (CPAE 6, Doc. 22).

———. 1915c. “Erklärung der Perihelbewegung des Merkur aus der allgemeinen Relativitätstheorie.” *Königlich Preußische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte*: 831–839. (CPAE 6, Doc. 23).

———. 1915d. “Die Feldgleichungen der Gravitation.” *Königlich Preußische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte*: 844–847. (CPAE 6, Doc. 25).

———. 1916. “Die Grundlage der allgemeinen Relativitätstheorie.” *Annalen der Physik* 49: 769–822. (CPAE 6, Doc. 30).

———. 1918. “Prinzipielles zur allgemeinen Relativitätstheorie.” *Annalen der Physik* 55: 241–244. (CPAE 7, Doc. 4).

Einstein, Albert, and Adriaan D. Fokker. 1914. “Die Nordströmsche Gravitationstheorie vom Standpunkt des absoluten Differentialkalküls.” *Annalen der Physik* 44: 321–328. (CPAE 4, Doc. 28).

Einstein, Albert, and Marcel Grossmann. 1913. *Entwurf einer verallgemeinerten Relativitätstheorie und einer Theorie der Gravitation*. Leipzig: Teubner (CPAE 4, Doc. 13).

———. 1914. “Kovarianzeigenschaften der Feldgleichungen der auf die verallgemeinerte Relativitätstheorie gegründeten Gravitationstheorie.” *Zeitschrift für Mathematik und Physik* 63: 215–225 (CPAE 6, Doc. 2).

Eisenstaedt, Jean, and A. J. Kox (eds.). 1992. *Studies in the History of General Relativity* (*Einstein Studies,* Vol. 3). Boston: Birkhäuser.

Goenner, Hubert, Jürgen Renn, Jim Ritter, and Tilman Sauer, *The Expanding Worlds of General Relativity* (*Einstein Studies*, Vol. 7). Boston: Birkhäuser.

Grassmann, Hermann. 1862. *Die Ausdehnungslehre*. Berlin: Enslin.

Gray, Jeremy (ed.). 1999. *The Symbolic Universe: Geometry and Physics, 1890–1930*. Oxford: Oxford University Press.

Grossmann, Marcel. 1904. “Die fundamentalen Konstruktionen der nicht-Euklidischen Geometrie.” [Supplement to:] *Programm der Thurgauischen Kantonschule für das Schuljahr 1903/04*. Frauenfeld: Huber & Co.

Howard, Don, and John Stachel (eds.). 1989. *Einstein and the History of General Relativity* (*Einstein Studies*, Vol. 1). Boston: Birkhäuser.

Janssen, Michel. 1999. “Rotation as the Nemesis of Einstein’s *Entwurf* Theory.” In (Goenner et al., 127–157).

———. 2005. "Of Pots and Holes: Einstein's Bumpy Road to General Relativity." *Annalen der Physik* 14: *Supplement* 58–85. Reprinted in J. Renn (ed.) *Einstein's Annalen Papers. The Complete Collection 1901–1922*. Weinheim: Wiley-VCH, 2005.

Kollros, Louis 1956. "Albert Einstein en Suisse: Souvenirs." *Helvetica Physica Acta. Supplementum* 4: 271–281.

Kottler, Friedrich. 1912. "Über die Raumzeitlinien der Minkowski'schen Welt." *Kaiserliche Akademie der Wissenschaften* (Vienna). *Mathematisch-naturwissenschaftliche Klasse. Abteilung IIa. Sitzungsberichte* 121: 1659–1759.

Kretschmann, Erich. 1917. "Über den physikalischen Sinn der Relativitätspostulate. A. Einsteins neue und seine ursprüngliche Relativitätstheorie." *Annalen der Physik* 53: 575–614.

Laue, Max. 1911a. "Zur Dynamik der Relativitätstheorie." *Annalen der Physik* 35: 524–542.

———. 1911b. *Das Relativitätsprinzip*. Braunschweig: Vieweg.

Minkowski, Hermann. 1908. "Die Grundgleichungen für die elektromagnetischen Vorgänge in bewegten Körpern." *Königliche Gesellschaft der Wissenschaften zu Göttingen. Mathematisch–physikalische Klasse. Nachrichten*: 53–111. (English translation of the appendix "Mechanics and the Relativity Postulate" in vol. 3 of this series.)

Norton, John D. 1984. "How Einstein Found his Field Equations, 1912–1915." *Historical Studies in the Physical Sciences* 14: 253–316. Page references to reprint in (Howard and Stachel 1989, 101–159).

———. 1985. "What Was Einstein's Principle of Equivalence?" Studies in History and Philosophy of Science 16: 203–246. Reprinted in (Howard and Stachel 1989, 5–47).

———. 1992. "The Physical Content of General Covariance." In (Eisenstaedt and Kox 1992, 181–315).

———. 1993. "General Covariance and the Foundations of General Relativity: Eight Decades of Dispute." *Reports on Progress in Physics* 56: 791–858.

———. 1999. "Geometries in Collision: Einstein, Klein, and Riemann." In (Gray 1999, 128–144).

———. 2000. "'Nature is the Realisation of the Simplest Conceivable Mathematical Ideas': Einstein and the Canon of Mathematical Simplicity." *Studies in History and Philosophy of Modern Physics* 31: 135–170.

Reich, Karin. 1994. *Die Entwicklung des Tensorkalküls. Vom absoluten Differentialkalkül zur Relativitätstheorie*. Basel: Birkhäuser.

Riemann, Bernhard. 1867. "Über die Hypothesen, welche der Geometrie zu Grunde liegen." *Königliche Gesellschaft der Wissenschaften zu Göttingen. Abhandlungen* 13: 133–152.

Renn, Jürgen, and Tilman Sauer. 1999. "Heuristics and Mathematical Representation in Einstein's Search for a Gravitational Field Equation." In (Goenner et al., 127–157).

Sommerfeld, Arnold. 1910a. "Zur Relativitätstheorie I. Vierdimensionale Vektoralgebra." *Annalen der Physik* 32: 749–776.

———. 1910b. "Zur Relativitätstheorie II. Vierdimensionale Vektoranalysis." *Annalen der Physik* 33: 649–689.

Weyl, Hermann. 1918. *Raum. Zeit. Materie. Vorlesungen über allgemeine Relativitätstheorie*. Berlin: Springer.

Wright, Joseph Edmund. 1908. *Invariants of Quadratic Differential Forms*. Cambridge: Cambridge University Press.

JOHN D. NORTON

WHAT WAS EINSTEIN'S "FATEFUL PREJUDICE"?

In the later pages of the notebook, as Einstein let general covariance slip away, he devised and abandoned a new proposal for his gravitational field equations. This same proposal, revived nearly three years later, opened passage to his final theory. In abandoning it in the notebook, Einstein had all but lost his last chance of deliverance. This chapter reports and develops our group's accounts of this decision. Einstein's later accounts of this decision blame it upon what he called the "fateful prejudice" of misinterpreting the Christoffel symbols. We suggest that Einstein's aberrant use and understanding of coordinate systems and coordinate conditions was as important as another fateful prejudice.

INTRODUCTION

Under a decade of analysis, discussion and reflection, Einstein's Zurich Notebook has yielded. Strategies that were once enigmatic and pages that were once obscure have become familiar. For the great part, we understand the problems Einstein approached, how he sought to solve them, when these efforts succeeded, when they failed and even the hesitations behind the smallest markings. In other parts we may follow a calculation line by line but our view of his hopes and plans remain distant. Or he may abandon a calculation with just a few symbols surviving on the page. They can be deciphered only through luck or clairvoyance.

The boundary that fences in the clear from the obscure has grown so that less and less escapes it. The intriguing puzzles of the notebook remain at this boundary. They cannot be solved with the assurance that the weight of evidence admits no alternative. But they are not so distant that we must despair of any solution. We understand just enough of these puzzles to sense that a complete solution lies within our grasp. We may even articulate one or more candidates that are both plausible and attractive. Yet the evidence we cull from the notebook and elsewhere remains sufficient to encourage us, but insufficient to enable a final decision.

My purpose in this chapter is to review two of these problems. I will draw heavily on ideas that have circulated freely in our group and have grown, mutated and contracted as they passed between us.[1] This chapter will report on these communal ideas,

1 I gratefully acknowledge thoughtful discussion and responses on this chapter and its proposals from the members of this group (who are also co-authors in this volume) and also from Jeroen van Dongen.

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 2
Einstein's Zurich Notebook: Commentary and Essays.

while it gives my own particular viewpoint on them and adds a conjecture. Many of the ideas in this chapter are not reflected in our joint commentary because my viewpoints and conjecture represent a minority opinion. At the boundary, where categorical evidence is elusive, our intuitions and sensibilities decide. They differ as we pass through the group. We do not all know the same Einstein.

Two Puzzles

The problems meet on page 22R of the notebook. There we find Einstein generating the very same gravitational field equations of near general covariance that will reappear briefly in his publication of 4 November 1915, when he ruefully returned to general covariance. This supplies our first puzzle:

- Why were these field equations rejected in the notebook, when they were deemed admissible in November, 1915?

These equations did not employ the Ricci tensor as gravitation tensor, as would the source free field equations of Einstein's final theory. Famously, Einstein and Grossmann had mentioned but discarded this possibility in their joint *Entwurf* paper. The equations on page 22R employ a different gravitation tensor, which we have come to call informally the "November tensor." It was carefully and apparently successfully contrived to avoid exactly the problems they imagined for the Ricci tensor.

The calculations on page 22R differ in no essential way from those Einstein would publish in 1915. The calculations on the surrounding pages do differ. The absolute differential calculus makes it easy to write down expressions that are generally covariant; they hold in all coordinate systems. In the modern literature we routinely restrict these expressions to specialized coordinate systems by imposing freely chosen coordinate conditions. As Einstein's calculations in the notebook progressed, he became quite adept at the purely mathematical aspects of applying these conditions. Careful analysis of the pages show that his use of these conditions came to differ considerably from the modern usage and possibly with fatal consequences. Our second puzzle is to understand these differences:

- Did Einstein *choose* to use coordinate conditions in an idiosyncratic way later in the notebook? Or was he unaware of the modern usage?

In solving these puzzles, more is at stake than merely deciphering a few pages of a notebook that may not have long occupied Einstein. These pages mark Einstein's all but last chance to rescue himself from the misconceptions that led him to his *Entwurf* theory and to more than two years of distress as his greatest discovery eluded him. A solution to these puzzles will tell us if Einstein's final slide into the abyss rested on simple blunders, lack of imagination or creative misunderstandings that have yet to be appreciated in the historical literature.

Four Parts

This chapter is divided into four parts. In the first, I will review the circumstances that induced Einstein to the proposal of the "November tensor" T^{x}_{il} as gravitation tensor. Its rejection in the notebook will be explained partially by drawing on a proposal of Jürgen Renn's. At the time of the notebook, as Einstein later recalled, he failed to see that the Christoffel symbols were the natural expression for the components of the gravitational field, his "fateful prejudice." As a result, he was unable to see how to recover a stress-energy tensor for the gravitational field and the associated conservation laws from the "November tensor." The calculation just proved too complicated. This problem was resolved in November 1915 when Einstein had developed more powerful mathematical methods.

The second part outlines the puzzle surrounding Einstein's use of coordinates. I will distinguish the standard way in which coordinate conditions are used from the way that Einstein came to use them later in the notebook. It is so different that our group labels coordinate conditions used this way as "coordinate restrictions." This non-standard use of coordinate restrictions can aid us in explaining the notebook rejection of the "November tensor," if in addition we assume that Einstein was unaware that the same mathematical manipulations could be used in the modern manner as coordinate conditions. The evidence available to us admits no final decision over Einstein's awareness of this usage. I will suggest however that there are sufficient indications to make his supposed lack of awareness implausible and that page 22R of the notebook might well mark a transition from the use of coordinate conditions to coordinate restrictions.

The third part develops a conjecture on what might lie behind Einstein's idiosyncratic use of coordinate conditions in the notebook. In his later hole argument, Einstein erred in tacitly according an independent reality to coordinate systems. It is now speculated he may have committed this same error within the notebook while using coordinate conditions to extract the Newtonian limit from the "November tensor." The outcome would be that his theory overall would seem to gain no added covariance from the use of coordinate conditions rather than coordinate restrictions, to which Einstein reverted for their greater simplicity. Once again, the available evidence admits no final decision. I will suggest however that the conjecture is plausible since it requires us to suppose no additional errors by Einstein; he merely needs to follow through consistently on the misapprehensions we know he harbored in the context of the hole argument.

The fourth part offers a summary conclusion.

1. THE PUZZLE OF THE GRAVITATION TENSORS

Why did Einstein abandon the gravitational field equations in the notebook on page 22R that he later deemed suitable for publication on 4 November 1915? This is our first puzzle. In the first section of this part I will review the essential background to this puzzle. In the pages preceding page 22R, Einstein considered and rejected the natural candidate for a gravitation tensor, the Ricci tensor. It fell to misconceptions about static fields and the form of gravitational field equations in the case of weak fields. In the second section of this part I will describe how the proposal of page 22R was contrived ingeniously to circumvent the illusory flaws he had imagined for the Ricci tensor. In the third section I will review Einstein's later recollections concerning the notebook rejection and the central role that, as I shall call it, "{} prejudice" played in them. Drawing on a proposal by Jürgen Renn, I will advance what I believe is a plausible account of its significance. The difficulty was the recovery of an expression for the stress-energy tensor of the gravitational field and its associated conservation law. Because Einstein did not recognize that the Christoffel symbols are the natural structure for representing the components of the gravitational field, he thought this recovery required the algebraic expansion of the Christoffel symbols. That yielded such a surfeit of terms that Einstein despaired of completing the calculation. This difficulty, along with others to be reviewed in later parts of this chapter, led Einstein to abandon the proposed gravitation tensor. In 1914, in the course of his work on the *Entwurf* theory, Einstein developed more powerful variational methods. These enabled him to complete the calculation and to see the significance of the Christoffel symbols.

1.1 Background: The Rejection of the Ricci Tensor

The Entwurf Papers

In the *Entwurf* paper, Einstein and Grossmann famously report their failure to find generally covariant gravitational field equations. Their search had focused on finding a gravitation tensor, $G_{\mu\nu}$, constructed from the metric tensor and its derivatives, to be used in the gravitational field equations

$$G_{\mu\nu} = \kappa T_{\mu\nu}, \tag{1}$$

where $T_{\mu\nu}$ is the stress-energy tensor and κ is a constant. The absolute differential calculus of Ricci and Levi-Civita supplied the natural structure from which generally covariant gravitation tensors can readily be constructed. It is the Riemann tensor, written in Einstein's paper of 4 November 1915 (Einstein 1915a) as[2]

2 My policy with notation will be to follow the conventions used at the time of the work discussed. In November 1915, Einstein indicated contravariant and covariant components of a tensor by raised and lowered indices. Summation over repeated indices was *not* implied. The notation for the Riemann tensor and Christoffel symbols do not respect this raising and lowering convention.

$$(ik,lm) = \frac{\partial}{\partial x_m}\begin{Bmatrix} il \\ k \end{Bmatrix} - \frac{\partial}{\partial x_1}\begin{Bmatrix} im \\ k \end{Bmatrix} + \sum_\rho \begin{Bmatrix} il \\ \rho \end{Bmatrix}\begin{Bmatrix} \rho m \\ k \end{Bmatrix} - \begin{Bmatrix} im \\ k \end{Bmatrix}\begin{Bmatrix} \rho l \\ k \end{Bmatrix}, \tag{2}$$

where the Christoffel symbols of the second kind are

$$\begin{Bmatrix} il \\ k \end{Bmatrix} = \frac{1}{2}\sum_r g^{kr}\begin{bmatrix} il \\ r \end{bmatrix} = \frac{1}{2}\sum_r g^{kr}\left(\frac{\partial g_{ir}}{\partial x_k} + \frac{\partial g_{rl}}{\partial x_k} - \frac{\partial g_{il}}{\partial x_r}\right).$$

(The term $[il/r]$ is the Christoffel symbol of the first kind and is defined implicitly in this expression.) The Ricci tensor is the first nontrivial contraction, unique up to sign, of the Riemann tensor, written by Einstein as

$$G_{im} = \sum_{kl}(ik,lm). \tag{3}$$

Einstein later chose this tensor as the gravitation tensor in the source free case.

Einstein and Grossmann had revealed that they had considered this candidate for the gravitation tensor in preparing the *Entwurf* paper. They explained, in Grossmann's words, "...it turns out that this tensor does *not* reduce to the [Newtonian] expression $\Delta\phi$ in the special case of an infinitely weak, static gravitational field."[3] Einstein and Grossmann's explanation proved all too brief. It did not even mention the use of the coordinate conditions that are expected by the modern reader and that must be stipulated to restrict the coordinate systems of a generally covariant theory to those coordinate systems in which Newton's equations can hold. This omission even led to the supposition in the early history of this episode that Einstein was unaware of his freedom to apply these coordinate conditions.

With its earliest analyses,[4] we learned from the Zurich Notebook that Einstein understood all too well how to reduce generally covariant gravitational field equations to a Newtonian form by restricting the coordinate systems under consideration. In particular, he knew how to select what we now call "harmonic coordinates" to reduce the Ricci tensor to an expression analogous to the Newtonian $\Delta\phi$. With deeper analysis as developed in our commentary, the notebook provides a detailed account of how Einstein tested the Ricci tensor against his other expectations and how he was led to reject it.

Two Misconceptions: The Static Field…

What precluded acceptance of the Ricci tensor as the gravitation tensor were two interrelated expectations that proved to be incompatible with Einstein's final theory. On the basis of several apparently sound arguments, Einstein expected that static

---

3 "Allein es zeigt sich, daß sich dieser Tensor im Spezialfall des unendlich schwachen statischen Schwerefeldes nicht auf den Ausdruck $\Delta\phi$ reduziert." (Einstein and Grossmann 1913, 256–257)

4 See (Norton 1984) and also (Stachel 1980).

gravitational fields would be represented by a spatially flat metric, whose coefficients in a suitable coordinate system would be

$$g^{STAT}_{\mu\nu} = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & c^2(x,y,z) \end{bmatrix}, \tag{4}$$

where the g^{STAT}_{44} component $c^2(x,y,z)$ is some function of the three spatial coordinates $(x,y,z) = (x_1,x_2,x_3)$. The spatial flatness is represented by the constant value -1 for the other non-zero components, g^{STAT}_{11}, g^{STAT}_{22}, and g^{STAT}_{33}. This spatial flatness is not realized in general in the final theory.

We can understand exactly why Einstein would fail to anticipate this lack of spatial flatness. His explorations were based on the principle of equivalence, which asserted that a transformation to uniform acceleration in a Minkowski spacetime yielded a homogenous gravitational field (see Norton 1985). The Minkowski metric in standard coordinates is given by

$$g^{SR}_{\mu\nu} = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & c^2 \end{bmatrix} \tag{5}$$

for c now a constant interpreted as the speed a light. If one transforms to a coordinate system in uniform acceleration, the metric reverts to a form Einstein associated with a homogeneous gravitational field, $g^{HG}_{\mu\nu}$, which has the form of $g^{STAT}_{\mu\nu}$, but in which c is a linear function of the spatial coordinates, x_1, x_2, x_3. If the acceleration is in the direction of the $x_1 = x$ coordinate, for example, then $c = c_0 + ax$, for c_0 and a arbitrary constants whose values are set by the particulars of the transformation, so that

$$g^{HG}_{\mu\nu} = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & (c_0 + ax)^2 \end{bmatrix} \tag{6}$$

Einstein's early strategy in his work on gravitation had been to recover the properties of arbitrary gravitational field by judiciously generalizing those of $g^{HG}_{\mu\nu}$. His mistake, in 1912 and 1913, was to fail to anticipate that the spatial flatness of $g^{HG}_{\mu\nu}$, was not a property of all static fields, but a very special peculiarity of $g^{HG}_{\mu\nu}$.

…and the Field Equations for Weak Fields

Einstein's second expectation concerned how the gravitational field equations (1) would reduce to the Newtonian limit. In the weak field case, one supposes that one can find coordinate systems in which the metric adopts the form

$$g_{\mu\nu} = g^{SR}_{\mu\nu} + \varepsilon_{\mu\nu} \tag{7}$$

The quantities $\varepsilon_{\mu\nu}$ are of first order of smallness. For this weak field, Einstein supposed that the gravitation tensor of (1) would reduce to[5]

$$\Gamma_{\mu\nu} = \sum_{\alpha\beta} \frac{\partial}{\partial x_\alpha}\left(\gamma_{\alpha\beta}\frac{\partial \gamma_{\mu\nu}}{\partial x_\beta}\right) + \begin{pmatrix}\text{terms of}\\ \text{second order}\\ \text{and higher}\end{pmatrix}. \tag{8}$$

If the gravitation tensor reduced to this form in the weak field, then all that would remain to first order is the first term of (8), so that the gravitational field equations would reduce to the near-Newtonian expression

$$\sum_{\alpha\beta} \gamma^{SR}_{\alpha\beta} \frac{\partial^2}{\partial x_\alpha \partial x_\beta} g_{\mu\nu} = \left(\frac{1}{c^2}\left(\frac{\partial}{\partial x_4}\right)^2 - \left(\frac{\partial}{\partial x_1}\right)^2 - \left(\frac{\partial}{\partial x_2}\right)^2 - \left(\frac{\partial}{\partial x_3}\right)^2\right) g_{\mu\nu} = \kappa T_{\mu\nu} \tag{9}$$

or more simply expressed

$$\Box g_{\mu\nu} = \kappa T_{\mu\nu}.$$

It turns out that these most natural of intermediates in the transition to Newton's law of gravitation are not realized by the final theory. In it, the weak field equations corresponding to (9) include an extra trace term. See (Einstein 1922, 87).

$$\Box g_{\mu\nu} = \kappa(T_{\mu\nu} - (1/2) g_{\mu\nu} T). \tag{9'}$$

These two expectations concerning the static field and weak field are closely connected. In particular, as Einstein showed in Einstein (Einstein 1913, 1259), one recovers a spatially flat static metric $g^{STAT}_{\mu\nu}$ if one solves the weak field equation (9) for the case a of a time independent field produced by a static, pressureless dust

5 I revert to the notation of (Einstein and Grossmann 1913). Summation is *not* implied by repeated indices. All indices are written as subscript with the covariant and contravariant forms of a tensor represented by Latin and Greek letters respectively. Thus the modern $g_{\mu\nu}$ is written as $g_{\mu\nu}$, but the modern $g^{\mu\nu}$ is written as $\gamma_{\mu\nu}$. Coordinate indices are written as subscript as well.

cloud.[6] This recovery of a spatially flat solution is blocked by the added trace term T in (9') in the final theory.

That Prove Fatal

On page 19L of the notebook, Einstein showed that he knew how to reduce the Ricci tensor to the weak field form required by (9). Using a standard device in the literature, he simply restricted his coordinate systems to those in which the harmonic condition

$$\sum_{\kappa l} \gamma_{\kappa l} \begin{bmatrix} \kappa l \\ i \end{bmatrix} = 0 \tag{10}$$

is satisfied. He immediately found that he could eliminate all the second derivative terms that were not required by the operator (8) for the Newtonian limit. Disaster ensued over the pages 19R–21R for this promising combination of Ricci tensor as gravitation tensor and harmonic coordinate systems. Einstein sought to bring this combination into accord with his expectations (4) for static fields and for the weak field equations (9). He failed and inevitably so. The coordinate systems used to bring the static field into the form of $g^{STAT}_{\mu\nu}$ in (4) are not harmonic. That coordinate system does, however, satisfy a formally similar coordinate condition

$$\sum_{\kappa} \frac{\partial \gamma_{\kappa\alpha}}{\partial x_\kappa} = 0. \tag{11}$$

(We call this "Hertz condition" in this volume since it is mentioned by Einstein in a letter to Paul Hertz of 22 August 1915 (CPAE 8, Doc. 111).) What makes this condition attractive is that it entails the weak field form of the energy momentum conservation law[7]

$$\sum_{\nu} \frac{\partial \Theta_{\kappa\alpha}}{\partial x_\nu} = 0. \tag{12}$$

Einstein even realized that he could retain this form of the energy conservation law and the harmonic condition if he added the trace term in T in (9'), but the modified field equations were no longer compatible with his expectations for the weak static

6 The prediction of spatial flatness is almost immediate. The stress energy tensor $T_{\mu\nu}$ for this static dust cloud will satisfy $T_{\mu\nu} = 0$ excepting T_{44}. Thus we have immediately for all values of μ, ν excepting 4,4, that $\Delta g_{\mu\nu} = 0$ for all spacetime. With finite values at spatial infinity as a boundary condition, these last equations solve to yield $g_{\mu\nu} = \text{constant}$ for all μ, ν excepting 4,4, as required by $g^{STAT}_{\mu\nu}$ of (4).

7 $\Theta_{\mu\nu}$ is the contravariant form of the stress-energy tensor $T_{\mu\nu}$. The condition (11) combined with the field equation (9) yields the weak field form of the energy conservation law through

$$\kappa \sum_{\nu} \frac{\partial \Theta_{\mu\nu}}{\partial x_\nu} = \sum_{\nu} \frac{\partial}{\partial x_\nu}\left(\sum_{\alpha\beta} \gamma^{SR}_{\alpha\beta} \frac{\partial^2}{\partial x_\alpha \partial x_\beta} \gamma_{\mu\nu}\right) = \sum_{\nu\alpha\beta} \gamma^{SR}_{\alpha\beta} \frac{\partial^2}{\partial x_\alpha \partial x_\beta} \left(\frac{\partial \gamma_{\mu\nu}}{\partial x_\nu}\right) = 0.$$

field $g^{STAT}{}_{\mu\nu}$, so they could not stand. Harmonic coordinate systems no longer appear in the notebook.

1.2 The "November Tensor"

The outcome of Einstein's investigations of the Ricci tensor was disappointing. But his creative energies were far from spent. He then turned immediately to another proposal for a gravitation tensor, the one he would publish on 4 November 1915, upon his return to general covariance. It is laid out on page 22R of the notebook. Einstein shows how it is possible to split off a part of the Ricci tensor that is not a generally covariant tensor, but at least transforms tensorially under unimodular transformations.

Unimodular Transformations

The class of unimodular transformations has a simple defining property. A coordinate transformation $x_\alpha \to x'_\beta$ is fully specified by the associated matrix of differential coefficients $\partial x'_\beta / \partial x_\alpha$. A transformation is unimodular if the determinant of this matrix is unity:

$$\mathrm{Det}\left(\frac{\partial x'_\beta}{\partial x_\alpha}\right) = 1. \tag{13}$$

Unimodular transformations preclude transformations that uniformly expand the coordinate system, such as $x'_\alpha = 2x_\alpha$. They are volume preserving in spacetime.[8]

The coefficients of the metric tensor transform according to

$$g'_{\mu\nu} = \left(\frac{\partial x_\alpha}{\partial x'_\mu}\right)\left(\frac{\partial x_\beta}{\partial x'_\nu}\right) g_{\alpha\beta}.$$

Taking the determinants of these quantities we find that the (positive valued)[9] determinant G transforms according to

$$\sqrt{G'} = \mathrm{Det}\left(\frac{\partial x_\alpha}{\partial x'_\mu}\right)\sqrt{G}.$$

It now follows immediately that $\sqrt{G'} = \sqrt{G}$ for unimodular transformations, that is, when (13) holds. This equality tells us that $\sqrt{G}$ transforms as a scalar under unimodular transformation, as do functions of it such as $\log\sqrt{G}$. We can easily form unimodular vectors from this quantity. The coordinate derivative $\partial\phi / \partial x_i$ of a generally covariant scalar ϕ is a generally covariant vector. Similarly, the coordinate derivative

---

8 They are volume preserving in the coordinate space. A volume element $dx_1 dx_2 dx_3 dx_4$ for a region bounded by the four coordinate differentials dx_α in coordinate space is preserved since it transforms according to the rule $dx'_1 dx'_2 dx'_3 dx'_4 = \mathrm{Det}\,(\partial x'_\beta / \partial x'_\alpha) dx_1 dx_2 dx_3 dx_4$. The invariant volume element of a metrical spacetime, $\sqrt{G} dx_1 dx_2 dx_3 dx_4$ is also preserved since $\sqrt{G}$ is an invariant under unimodular transformation.

9 In other places, it is written as $-g$.

of a unimodular scalar is a unimodular vector. Therefore $\dfrac{\partial \log \sqrt{G}}{\partial x_i}$ is a unimodular vector. This result is the key to Einstein's plan.

Proposal: A Unimodular Tensor....

On page 22R of the notebook, Einstein took the Ricci tensor T_{il} and expressed it as the sum of two parts. He wrote

$$T_{il} = \underbrace{\left(\frac{\partial T_i}{\partial x_l} - \sum \begin{Bmatrix} il \\ \lambda \end{Bmatrix} T_\lambda\right)}_{\text{tensor 2nd rank}} - \underbrace{\sum_{\kappa l} \left(\frac{\partial \begin{Bmatrix} il \\ \kappa \end{Bmatrix}}{\partial x_\kappa} - \begin{Bmatrix} i\kappa \\ \lambda \end{Bmatrix} \begin{Bmatrix} l\lambda \\ \kappa \end{Bmatrix} \right)}_{\text{presumed gravitation tensor } T^x_{il}.} \tag{14}$$

His purpose is quite clear. And if there were any doubt, the proposal is explained in detail in (Einstein 1915a). The first term of T_{il} is a just the covariant derivative of the unimodular vector

$$T_i = \frac{\partial \log \sqrt{G}}{\partial x_i}$$

and therefore a tensor under unimodular transformations. Since the Ricci tensor T_{il} transforms as a tensor under all transformations, Einstein could infer that the second term of (14) must also transform as a tensor under unimodular transformations.[10] This second term, denoted as T^x_{il}, is chosen by Einstein as a candidate gravitation tensor. Because of its reappearance in November 1915, we have labeled it the "November tensor" in this volume. Its selection is compatible with Einstein's ambitions for extending the principle of relativity to acceleration. While not supplying general covariance, covariance under unimodular transformations is sufficiently expansive to capture transformations between inertial and accelerated coordinate systems. As Einstein shows in (Einstein 1915a, 786), these acceleration transformations include ones that set the spatial coordinate axes into rotation as well as ones that accelerate its spatial origin without rotation.[11]

…that Gives the Newtonian Limit and Energy Conservation

The remainder of the page explains why Einstein was attracted to this new candidate. He had been unable to reduce the entire Ricci tensor to the form (8) without employing a coordinate condition, the harmonic condition, that brought fatal problems. Ein-

10 The result is automatic. The quantity T^x_{il} can be expressed as a difference of two quantities, each of which are tensors at least under unimodular transformations.

stein now showed that he could reduce the tensor T^x_{il} to the form (8) if he considered coordinate systems which satisfied the coordinate condition (11) introduced above.

As Einstein proceeded to show, with the assumption of this condition, the candidate gravitation tensor T^x_{il} reduced to

$$-2T^x_{il} = \sum_{\alpha\beta}\gamma_{\alpha\beta}\frac{\partial^2 g_{il}}{\partial x_\alpha \partial x_\beta} + \left(\text{terms quadratic in } \frac{\partial g_{\mu\nu}}{\partial x_\alpha}\right). \tag{15}$$

In the weak field of (7), the terms quadratic in

$$\frac{\partial g_{\mu\nu}}{\partial x_\alpha}$$

will all be of order ε^2 and thus of the second order of smallness; the first term of (15) agrees with the first term of (8) in quantities of first order. The candidate tensor T^x_{il} has been reduced to the requisite form (8). In addition, the reduction has been effected by just the condition (11) needed to enforce energy conservation in the weak field. As Einstein had already found, that coordinate condition, in conjunction with the weak field equations (9) entailed energy conservation in the weak field form (12).

11 This last compatibility is not straightforward. The choice of T^x_{il} as gravitation tensor is not compatible with Einstein's favorite examples of a field produced by uniform, rectilinear acceleration in Minkowski spacetime, the static, homogeneous field, $g^{HG}_{\mu\nu}$, given as (6). One finds by explicit calculation that $g^{HG}_{\mu\nu}$ is not a solution of the source free field equations $T^x_{il} = 0$. This failure is already suggested by that fact that T^x_{il} is only a tensor under unimodular transformations and that the transformation from $g^{SR}_{\mu\nu}$ to $g^{HG}_{\mu\nu}$ is not unimodular. (Unimodular transformations preserve the determinant of the metric. But Det $(g^{SR}_{\mu\nu}) = c^2 = \text{constant}$, whereas $\text{Det}(g^{HG}_{\mu\nu}) = (c_0 + ax)^2 \neq \text{constant}$.) Now $g^{SR}_{\mu\nu}$ is obviously a solution of the source free field equations $T^x_{il} = 0$. So we cannot infer from the covariance properties of T^x_{il} that $g^{HG}_{\mu\nu}$ is also a solution.

If Einstein was aware of this problem, he did not find it immediately fatal to T^x_{il} as gravitation tensor. The problem should have been apparent as soon as Einstein contemplated a gravitation tensor covariant only under unimodular transformations. Yet he proceeded on page 22R with the elaborate recovery of the Newtonian limit. Again there is no trace of a concern over the homogeneous field, $g^{HG}_{\mu\nu}$, in the pages surrounding and following. (The concern is directed towards the coordinate restriction (11) and the rotation field $g^{ROT}_{\mu\nu}$ defined below.) The failure amounts to a failure of his principle of equivalence. But Einstein had already reconciled himself to such a failure in his theories of 1912 and it arose again in his *Entwurf* theory. See (Norton 1985, §4.3).

In the text I have explained his apparent indifference by assuming that he adopted the position expressed later in (Einstein 1915a, 786). Employing the same gravitation tensor T^x_{il}, the theory of that paper was also covariant under unimodular transformations. In order to affirm that the theory satisfied the relativity of motion, he observed (in part) that the coordinate transformation $x' = x - \tau_1$, $y' = y - \tau_2$, $z' = z - \tau_3$, $t' = t$, with τ_1, τ_2 and τ_3 arbitrary functions of t, is unimodular. We might note that this transformation corresponds to a large class of unidirectional accelerations. While the class does not include the transformation from $g^{SR}_{\mu\nu}$ to $g^{HG}_{\mu\nu}$, Einstein may well have simply assumed that it did include related transformation of comparable physical significance.

…Or Does It?

Einstein could hardly hope for a more satisfactory outcome. He was burdened by strict and unforgiving requirements on static fields and the weak field limit. Yet he found gravitational field equations of very broad covariance compatible with both. So satisfactory is this resolution that Einstein published it in November 1915 upon his return to general covariance. T^x_{il} is the gravitation tensor he proposed in his communication of 4 November to the Prussian Academy (Einstein 1915a). On 4 November, he had little choice. The natural gravitation tensors, the Ricci tensor and then the Einstein tensor, were still unavailable to him. He was still bewitched by his early, mistaken expectations concerning static fields and the weak field limit. These expectations were dispelled after that communication, over the course of that November. A rapid series of communications first brought him his successful explanation of the anomalous motion of Mercury and then his final, generally covariant field equations.

In the 4 November communication, Einstein paused to explain the transient charms of the "November tensor" T^x_{il}. He closed the communication of 4 November by showing (§4) that the coordinate condition (11), in the case of weak fields, reduces his field equations to the form (9).

Yet Einstein's achievement on page 22R of the Zurich Notebook proves to be as puzzling as it is impressive. For the proposal disappears as rapidly as it appeared; it receives no further serious consideration in the notebook.[12] The difficulties that led to its dismissal cannot be those that defeated the combination of the Ricci tensor and the choice of harmonic coordinate systems. These were the misleading expectations about static fields and the weak field limit. The gravitation tensor T^x_{il} was compatible with both. Why did Einstein so rapidly discard this promising candidate for his gravitation tensor? What changed to make it acceptable in November 1915?

1.3 The {} Prejudice

We have fragments of evidence that allow us to answer these questions. Some come from the pages of the notebook surrounding page 22R. The most important come in Einstein's later recollections.

A Letter to Sommerfeld of 28 November 1915

Einstein's most complete account comes in all too brief remarks in this letter. Having recounted the final field equations of his theory, Einstein continued:

> Of course it is easy to write down these generally covariant equations. But it is hard to see that they are the generalization of Poisson's [Newtonian] equations and not easy to see that they allow satisfaction of the conservation laws.

12 We shall see below in section 3.7 that T^x_{il} is reevaluated on the following page 23L, but now with the coordinate restriction (11) replaced by another. The candidate gravitation tensor reappears briefly on page 25L in an incomplete attempt to compute the stress energy tensor of the gravitational field associated with this gravitation tensor.

> Now one can simplify the whole theory significantly by choosing the reference system so that $\sqrt{-g} = 1$. Then the equations take on the form
>
> $$-\sum_{1}\frac{\partial\left\{\begin{matrix} im \\ 1 \end{matrix}\right\}}{\partial x_1}+\sum_{\alpha\beta}\left\{\begin{matrix} i\alpha \\ \beta \end{matrix}\right\}\left\{\begin{matrix} m\beta \\ \alpha \end{matrix}\right\} = -\kappa\left(T_{im}-\frac{1}{2}g_{im}T\right).$$
>
> I had already considered these equations 3 years ago with Grossmann up to the second term on the right hand side, but had then arrived at the result that they did not yield Newton's approximation, which was erroneous. What supplied the key to this solution was the realization that it is not
>
> $$\sum g^{l\alpha}\frac{\partial g_{\alpha i}}{\partial x_m}$$
>
> but the associated Christoffel symbols $\left\{\begin{matrix} im \\ 1 \end{matrix}\right\}$ that are to be looked upon as the natural expression for the "components" of the gravitational field. If one sees this, then the above equation becomes simplest conceivable, since one is not tempted to transform it by multiplying out [*Ausrechnen*] the symbols for the sake of general interpretation.[13]

Which equations had he considered three years before? "…[T]hese equations…up to the second term on the right hand side…," that is, excluding the trace term in T, coincide with the choice of the "November tensor" T^{x}_{il} as gravitation tensor on page 22R. Einstein tells Sommerfeld that he had considered these equations with Grossmann and that detail is affirmed by the appearance of Grossmann's name on the top of page 22R.[14]

The Fateful Prejudice

The elements of the account Einstein laid out to Sommerfeld reappear in other places in Einstein's writing. In his publication, (Einstein 1915a, 778), he also recounted his misidentification of the "'components' of the gravitational field." He recalled how he

---

13 "Es ist natürlich leicht, diese allgemein kovarianten Gleichungen hinzusetzen, schwer aber, einzusehen, dass sie Verallgemeinerungen von Poissons Gleichungen sind, und nicht leicht, einzusehen, dass sie den Erhaltungssätzen Genüge leisten. Man kann nun die ganze Theorie eminent vereinfachen, indem man das Bezugssystem so wählt, dass $\sqrt{-g} = 1$ wird. Dann nehmen die Gleichungen die Form an, […]. Diese Gleichungen hatte ich schon vor 3 Jahren mit Grossmann erwogen/ bis auf das zweite Glied der rechten Seite, war aber damals zu dem Ergebnis gelangt, dass sie nicht Newtons Näherung liefere, was irrtümlich war. Den Schlüssel zu dieser Lösung lieferte mir die Erkenntnis, dass nicht […] sondern die damit verwandten Christoffel'schen Symbole […] als natürlichen Ausdruck für die "Komponente" des Gravitationsfeldes anzusehen ist. Hat man dies gesehen, so ist die obige Gleichung denkbar einfach, weil man nicht in Versuchung kommt, sie behufs allgemeiner Interpretation umzuformen durch Ausrechnen der Symbole." Einstein to Arnold Sommerfeld, 28 November 1915 (CPAE 8, Doc. 153).

14 Presumably Einstein alludes to his earlier recovery of these equations in the introduction to his paper of 4 November 1915. Einstein recalls his work three years earlier with Grossmann and then claims: "In fact we had already then come quite close to the solution of the problem given in the following." ("In der Tat waren wir damals der im nachfolgenden gegebenen Lösung des Problems bereits ganz nahe gekommen.") (Einstein 1915a, 778)

had reformulated the energy conservation law in his earlier work, (Einstein 1914c). In the absence of non-gravitational forces, the law is just the vanishing of the covariant divergence of the stress-energy tensor $T_{\mu\nu}$. It could be re-expressed as[15]

$$\sum_{\nu} \frac{\partial \mathfrak{T}_{\nu}^{\sigma}}{\partial x_{\nu}} - \begin{Bmatrix} \nu\sigma \\ \tau \end{Bmatrix} \mathfrak{T}_{\tau}^{\nu} = \sum_{\nu} \frac{\partial \mathfrak{T}_{\nu}^{\sigma}}{\partial x_{\nu}} - \frac{1}{2} g^{\tau\mu} \frac{\partial g_{\mu\nu}}{\partial x_{\sigma}} \mathfrak{T}_{\tau}^{\nu} = 0,$$

where the tensor density $\mathfrak{T}_{\nu}^{\sigma} = \sqrt{-g} T_{\nu}^{\sigma}$. Einstein now reflected upon his earlier error:

> This conservation law had earlier induced me to view the quantities
>
> $$\frac{1}{2} \sum_{\mu} g^{\tau\mu} \frac{\partial g_{\mu\nu}}{\partial x_{\sigma}}$$
>
> as the natural expression for the components of the gravitational field, even though it is obvious, in view of the formulae of the absolute differential calculus, to introduce the Christoffel symbols
>
> $$\begin{Bmatrix} \nu\sigma \\ \tau \end{Bmatrix}$$
>
> instead of those quantities. This was a fateful prejudice.[16]

Einstein continues to argue for the naturalness of this choice. The Christoffel symbols are symmetric in the indices ν and σ and they reappear in the geodesic equation. However he does not explain precisely how this "prejudice" led him astray. For convenience, I will call this the "$\{\}$ prejudice."

A letter written to Lorentz the following January repeats and slightly clarifies the role of the $\{\}$ prejudice.

> I had already considered the present equations [of the final theory, not of 4 November] in their essentials 3 years ago with Grossmann, who had made me aware of Riemann's tensor. But since I had not recognized the formal meaning of the $\{\}$, I could achieve no overview and could not prove the conservation laws.[17]

The Problems Collected

If we assemble the clues, we find Einstein giving two reasons for his rejection of the "November tensor" T_{il}^{x} when he worked with Grossmann:[18]

---

15 Einstein refers back to the results in (Einstein 1914c). There the energy conservation law was written in terms of the covariant divergence of $T_{\mu\nu}$. In his paper of 4 November 1915, Einstein had discarded a term in $\sqrt{-g}$ to simplify the result at the expense of reducing its covariance to unimodular transformations only.

16 "Diese Erhaltungsgleichung hat mich früher dazu bereleitet, die Größen [eq.] als den natürlichen Ausdruck für die Komponenten des Gravitationsfeldes anzusehen, obwohl es im Hinblick auf die Formeln des absoluten Differenzialkalküls näher liegt, die Christoffelschen Symbole [eq.] statt jener Größen einzuführen. Dies war ein verhängnisvolles Vorurteil." (Einstein 1915a, 782)

17 "Die jetzigen Gleichungen hatte ich im Wesentlichen schon vor 3 Jahren zusammen mit Grossmann, der mich Riemanns Tensor aufmerksam machte, in Betracht gezogen. Da ich aber die formale Bedeutung der $\{\}$ nicht erkannt hatte, konnte ich keine Übersichtlichkeit erzielen und die Erhaltungssätze nicht beweisen." Einstein to Hendrik A. Lorentz, 1 January 1916 (CPAE 8, Doc. 177).

- He was unable to recover the Newtonian limit.
- The {} prejudice precluded recognition of the inherent simplicity of the equations and the recovery of the energy conservation law.

Both elements of Einstein's account are puzzling. A straightforward reading of page 22R shows Einstein recovering the Newtonian limit in exactly the same way as in his later publication of 4 November 1915. A more careful analysis will be needed, but that will be postponed to later parts of this chapter. Einstein's remarks about the {} prejudice are also puzzling at first. Einstein had a perfectly acceptable expression for the energy conservation law. It is just the vanishing of the covariant divergence of $T_{\mu\nu}$ and was introduced by Einstein on page 5R of the notebook. I believe that these last remarks admit a fairly simple explication.

Recovering Energy Conservation

To understand why this prejudice was fateful, we need to recall a major difference between Einstein's work in the notebook and in November 1915. Here I draw heavily on the insights of Jürgen Renn and Michel Janssen.[19] By 1915 Einstein had developed techniques of considerably greater sophistication for recovering energy conservation than he had used in 1913. Also, when Einstein talks of proving the conservation laws, we must understand him to mean a little more than merely recovering the standard result that the covariant divergence of $T_{\mu\nu}$ vanishes. We must understand an important part of the recovery to be the identification of a stress-energy tensor for the gravitational field, $t_{\mu\nu}$, that will figure in an alternate form of the energy conservation law (as given in Einstein and Grossmann 1913, 17)

$$\sum_{\mu\nu} \frac{\partial}{\partial x_\nu}\{\sqrt{-g}\ \gamma_{\sigma\mu}(T_{\mu\nu} + t_{\mu\nu})\} = 0.$$

At the time of the *Entwurf* theory, Einstein employed a simple device for generating this stress-energy tensor. It had been used on pages 19R, 20L and 21L of the notebook in the weak field, while Einstein weighed the fate of the Ricci tensor as gravitation tensor. Einstein took the expression for the gravitational force density in the weak field (7),

$$\frac{1}{2}\sum_{\mu\nu} \frac{\partial g_{\mu\nu}}{\partial x_\sigma}\Theta_{\mu\nu},$$

where $\Theta_{\mu\nu}$ is the contravariant form of the stress-energy tensor $T_{\mu\nu}$. He then substituted for $\Theta_{\mu\nu}$ using the gravitational field equation for the weak field (9). A simple

18 Since recollections are not infallible, there is always a possibility that the first difficulty with the Newtonian limit was misremembered and really pertained only to his problem with the Ricci tensor. We need not have such concerns for the {} prejudice. Since it was published on 4 November 1915, the notion was clearly formulated before Einstein had realized the problems with the Newtonian limit associated with his assumptions about the static field and the weak field equations.

19 See "Untying the Knot …" (What Did Einstein Know).

manipulation that preserved only terms of lowest order in the derivatives of $g_{\mu\nu}$ allowed this force density to be rewritten as the divergence of a tensor $t_{\alpha\sigma}$:[20]

$$\frac{1}{2}\sum_{\mu\nu\alpha\beta}\frac{\partial g_{\mu\nu}}{\partial x_\sigma}\left(\gamma_{\alpha\beta}\frac{\partial^2}{\partial x_\alpha \partial x_\beta}\gamma_{\mu\nu}\right)$$
$$= \frac{1}{2}\sum_{\mu\nu\alpha\beta\rho\tau}\frac{\partial}{\partial x_\beta}\left(\gamma_{\alpha\beta}\frac{\partial g_{\mu\nu}}{\partial x_\sigma}\frac{\partial \gamma_{\mu\nu}}{\partial x_\alpha} - \frac{1}{2}\delta_{\beta\sigma}\gamma_{\rho\tau}\frac{\partial g_{\mu\nu}}{\partial x_\rho}\frac{\partial \gamma_{\mu\nu}}{\partial x_\tau}\right) = \frac{1}{2}\gamma_{\alpha\beta}\frac{\partial t_{\alpha\sigma}}{\partial x_\beta}$$

$$\frac{1}{2}\sum_{\mu\nu\alpha\beta}\frac{\partial g_{\mu\nu}}{\partial x_\sigma}\left(\gamma_{\alpha\beta}\frac{\partial^2}{\partial x_\alpha \partial x_\beta}\gamma_{\mu\nu}\right)$$
$$= \frac{1}{2}\sum_{\mu\nu\alpha\beta\rho\tau}\frac{\partial}{\partial x_\beta}\left(\gamma_{\alpha\beta}\frac{\partial g_{\mu\nu}}{\partial x_\sigma}\frac{\partial \gamma_{\mu\nu}}{\partial x_\alpha} - \frac{1}{2}\delta_{\beta\sigma}\gamma_{\rho\tau}\frac{\partial g_{\mu\nu}}{\partial x_\rho}\frac{\partial \gamma_{\mu\nu}}{\partial x_\tau}\right) = \frac{1}{2}\frac{\partial}{\partial x_\beta}(\gamma_{\alpha\beta}t_{\alpha\sigma}).$$

Einstein identified that tensor with the stress-energy tensor of the gravitational field.

This equation holds only for quantities of second order of smallness (ε^2) in the metric tensor of (7) of the weak field. The major part of Einstein's strategy for deriving his *Entwurf* field equations was to determine what quantities must be added to the gravitation tensor of the weak field equations (9) to make the relation between force density and the divergence of $t_{\alpha\sigma}$ exact, that is, true for all orders. This strategy reappears after page 22R, on pages 24R and 25R, and then in the full derivation of the *Entwurf* gravitational field equations by this strategy on pages 26L–26R.

Why the $\{\}$ *Prejudice Was Fateful*

Now we can understand why the $\{\}$ prejudice was fateful as Einstein inspected the candidate gravitation tensor T_{il}^{x} on page 22R. On his account, he was unable to see how to recover the Newtonian limit, a problem we shall return to. He also needed to assure himself that the gravitation tensor was compatible with energy conservation and that included admission of a well-defined stress-energy tensor $t_{\alpha\sigma}$ for the gravitational field. Following his standard practice, that would mean that he must be able to reformulate the expression for gravitational force density as a divergence. We can immediately see the problem Einstein would face. The tensor T_{il}^{x} is expressed fully in terms of Christoffel symbols, with each representing a sum of three terms in

$$\frac{\partial g_{\mu\nu}}{\partial x_\tau}.$$

The product of two Christoffel symbols would yield nine of these derivative terms.[21] Faced with so many terms, we could well imagine Einstein's sense that he had no

20 The symbol $\delta_{\beta\sigma}$, where $\delta_{\beta\sigma} = 1$ when $\beta = \sigma$ and zero otherwise, was not then used by Einstein, but is introduced here for simplicity.

21 Einstein's *Entwurf* gravitation tensor has one second derivative term and three first derivative terms. Unless there are duplications, the November tensor would have three second derivative terms and nine first derivative terms.

overview (as he wrote to Lorentz above) or that this was certainly not the simplest conceivable equation (as he wrote to Sommerfeld above). We could well imagine that this difficulty, along with failure of the Newtonian limit, would be sufficient grounds for him to abandon the candidate tensor.

What changed by November 1915? In the course of 1914, Einstein developed powerful variational methods for recovering energy conservation and expressions for the stress-energy tensor of the gravitational field. (Einstein and Grossmann 1914; Einstein 1914c). He applied those to his field equations of 4 November 1915, and found that the expressions took on just about the simplest form one could expect—as long as all quantities were expressed in terms of the Christoffel symbols. His field Lagrangian was just

$$\sum_{\sigma\tau\alpha\beta} g^{\sigma\tau} \begin{Bmatrix} \sigma\beta \\ \alpha \end{Bmatrix} \begin{Bmatrix} \tau\alpha \\ \beta \end{Bmatrix}.$$

It is one of the simplest fully contracted expressions quadratic in the Christoffel symbols. (The Lagrangian must be quadratic if it is to return field equations linear in the second derivatives of the metric tensor.) His expression for the canonical stress-energy tensor of the gravitational field was scarcely more complicated.

Einstein's analysis in the notebook began with a force density

$$\frac{1}{2}\sum_{\mu\nu} \frac{\partial g_{\mu\nu}}{\partial x_\sigma} \Theta_{\mu\nu}$$

expressed in terms of the derivatives of the metric tensor. It overwhelmed him and he abandoned it. Einstein's analysis in November 1915 retained the Christoffel symbols and, using his more powerful methods, yielded just about the simplest expressions he could expect. In hindsight, Einstein diagnosed the error to lie in his starting point. Had he not misidentified the components of the gravitational field, would he have resisted the temptation to expand the Christoffel symbols? Would he have come to see that he had the right equations before him?

Einstein used the term "prejudice"—a belief not properly grounded in evidence. It is a fitting label for the error we reconstruct. He was not assured that energy conservation would fail for this tensor in the notebook. He had no firm proof, no result around which to maneuver. He merely balked at a very complicated calculation that could have, in principle, been completed. He had no good reason to abandon the tensor other than the hunch that the true way could not be that complicated. And he later found that it was not at all complicated when viewed from another perspective.

2. COORDINATE CONDITIONS AND COORDINATE RESTRICTIONS

On page 22R of the notebook, Einstein shows how to use the coordinate condition (11) to reduce the gravitation tensor T_{il}^{x} to the requisite Newtonian form (8). Why does he report to Sommerfeld that he and Grossmann had originally thought the resulting gravitational field equations incompatible with the Newtonian limit? In this part of the chapter and in the part to follow, I will describe two explanations, both requiring that Einstein did not use coordinate conditions in the modern way. The explanation to be developed in this part is the simpler of the two. It asserts that Einstein understood field equations to be compatible with the Newtonian limit if they had the form (8) not just in some specialized coordinate systems, but in all coordinate systems. A cursory inspection would reveal that T_{il}^{x} does not have this form (8), rendering it incompatible with the Newtonian limit.

For this account to be tenable, we must now explain why page 22R displays the apparently successful reduction of the tensor T_{il}^{x} to the Newtonian form (8) using coordinate condition (11). This part will supply that explanation by suggesting that Einstein did not use the coordinate condition (11) in the standard way, as it was later in Einstein's paper of 4 November 1915. It was not a condition just to be invoked in the case of the Newtonian limit. It was a postulate to be used universally. In part one of this section, I will review the these two ways of using conditions such as (11). In this volume, we reserve the term "coordinate condition" for the standard usage and "coordinate restriction" for the other usage suggested here. This notion of coordinate restriction was introduced by Jürgen Renn and Tilman Sauer.[22] We will see in the second section of this part that there is clear evidence that Einstein used coordinate restrictions in the notebook on page 22 and afterwards.

In the third section of this part, I will describe how we can use the notion of coordinate restrictions to explain why Einstein abandoned the gravitation tensor T_{il}^{x}. To do so, we need a further assumption. Einstein did not just use (11) as a coordinate restriction on page 22. We must assume that he was unaware of the other possibility of using (11) as a coordinate condition. Then his rejection of T_{il}^{x} as gravitation tensor is automatic; it does not have the Newtonian form (8). This account is the majority viewpoint within our group.

The account depends upon the assumption that Einstein was unaware of the possibility of using (11) as a coordinate condition. In the fourth section of this part, I will explain why I do not believe the assumption. There is no single piece of evidence that allows us to decide either way on the assumption. It lies on the boundary. However I believe that there are so many indications that speak against it, that their combined weight makes the assumption untenable. The most plausible account, I believe, is that page 22 of the notebook marks a turning point. Prior to it, Einstein used coordinate conditions; after he reverted to the use of coordinate restrictions.

2.1 Two Uses of One Equation

Four Equations Select a Coordinate System…

The equations of a generally covariant spacetime theory hold in arbitrary coordinate systems. In applying the theory, we may pick the coordinate system freely. The four coordinates are just four real valued functions x_{α} defined on the spacetime manifold.

22 See "Pathways out of Classical Physics …" (in vol. 1 of this series).

Therefore a coordinate system can be chosen with four conditions $x_\alpha(p) = f_\alpha(p)$, where the $f_\alpha(p)$ are four arbitrary real valued functions of suitable differentiability defined over events p. Thus four arbitrary conditions are associated with the choice of a coordinate system.

This simple fact about coordinate systems is often rendered as the much looser idea that there are four degrees of freedom in a generally covariant theory associated with the freedom of choice of the coordinate system.[23] These four degrees of freedom are more usually exploited indirectly by specifying four differential conditions on quantities defined in spacetime. Examples are the harmonic condition (10) and the condition (11) used on page 22R. They do not fully exhaust the freedom. Since they are differential conditions on the metric, they do not force a unique choice of coordinate system; differential equations admit many solutions according to the choice of boundary conditions. So each of (10) and (11) admit many coordinate systems. For example, if one admits a coordinate system x_α, it also admits any coordinate system linearly related to it. This follows immediately from the covariance of (10) and (11) under linear coordinate transformations.

In the case of the harmonic condition (10), the relation between the different forms of the condition can be made more explicit. We can define the natural, generally covariant analog of the d'Alembertian operator $\square$ used in (9) as follows. If φ is a scalar, we take its covariant derivative twice and contract with $\gamma_{\mu\nu}$ over the two resulting indices. In Einstein's notation of 1913, this gives

$$\square\varphi = \sum_{\alpha\beta\rho\sigma} \gamma_{\alpha\beta}\frac{\partial^2\varphi}{\partial x_\alpha \partial x_\beta} - \gamma_{\alpha\beta}\begin{bmatrix}\alpha\beta\\ \rho\end{bmatrix}\gamma_{\rho\sigma}\frac{\partial\varphi}{\partial x_\sigma}.$$

If we now form $\square x_\tau$ for each of the four coordinates, we quickly see that the harmonic condition (10) is equivalent to[24]

$$\square x_\tau = \sum_{\alpha\beta\rho\sigma} \gamma_{\alpha\beta}\frac{\partial^2 x_\tau}{\partial x_a \partial x_\beta} - \gamma_{\alpha\beta}\begin{bmatrix}\alpha\beta\\ \rho\end{bmatrix}\gamma_{\rho\sigma}\frac{\partial x_\tau}{\partial x_\sigma} = 0. \tag{10'}$$

23 This slogan—four degrees of freedom associated with the choice of coordinates—must be approached with some caution. It does not mean we can adjoin any four equations we like to our theory under the guise of choosing the coordinate system. Adding the single equation $R = 0$, where R is the curvature scalar, does a great deal more than restrict the coordinate system. One might imagine that restricting the equations to first order derivatives in $g_{\mu\nu}$ would protect us from these problems, since, at any event in spacetime, we can always transform such derivatives to zero. But it does not. Imagine that we have 100 such conditions, $C_1 = 0$, $C_2 = 0$, …, $C_{100} = 0$ that more than exhaust the freedom to choose coordinates. They can be disguised as a single equation $(C_1)^2 + \ldots + (C_{100})^2 = 0$.

24 To see the equivalent, notice that the first term of (10') vanishes for any coordinate system. The second term vanishes if (10) holds. So (10) entails (10'). Conversely, if (10') holds, its second term must vanish, which immediately entails (10).

One sees from this equation that the harmonic condition cannot fix the coordinate system uniquely.[25] If the condition is satisfied by x_τ, it will also be satisfied by any linear transform of it. Other transforms will also be admissible. The condition cannot fix the coordinate system up to linear transformation unless one invokes further restrictive conditions (see Fock 1959, §93).

We know that Einstein was aware of this form (10') of the harmonic condition on coordinate systems, then routinely available in the literature on infinitesimal geometry as the "isothermal" coordinates. At the bottom of page 19L on which he introduced the condition in form (10), he wrote "…Holds for coordinates which satisfy the eq[uation] $(\Delta\varphi = 0)$."

We see how equations (10) and (11) allow us choose a restricted set of coordinates. There are two ways relevant to our present interests that these equations may be used: as coordinate conditions and as coordinate restrictions.

…As Coordinate Conditions

Einstein later used a standard procedure for recovering the Newtonian limit from his generally covariant general theory of relativity. See for example, (Einstein 1922, 86–87). That theory must revert to Newton's theory of gravitation in the special circumstance of weak static fields, that is, under the assumption that the metric has the form (7) and is static. In addition, Newton's theory is not generally covariant, but is covariant under Galilean transformation only.[26] Therefore the covariance of Einstein's theory must be restricted if Newton's theory is to be recovered.

That covariance is already restricted in part by the presumption that the metric have the form (7). That form is not preserved under arbitrary transformations. The restriction to the weak field metric (7) is not, however, sufficient to reduce the covariance of the theory to the Galilean covariance of Newton's theory. That form is preserved by any transformation which introduces small changes of order of the $\varepsilon_{\mu\nu}$ to the coefficients of the metric. This last freedom is eliminated by imposing a coordinate condition, such as the harmonic condition (10). We have already seen the direct effect of this condition. It eliminates all second derivative terms from the Ricci tensor beyond those in the Newtonian like form (8). In so far as Einstein expected his *Entwurf* theory to have broad covariance, he must have believed the restriction of the metric to the weak field form (7) was sufficient restriction on its covariance for the recovery of the weak field limit.

A coordinate condition is used only in the special circumstances of the Newtonian limit; it is not imposed universally on the theory.

---

25 Notice that the operator $\sum_{\alpha\beta\rho\sigma} \gamma_{\alpha\beta}\begin{bmatrix}\alpha\beta\\ \rho\end{bmatrix}\gamma_{\rho\sigma}\frac{\partial}{\partial x_\sigma}$ is invariant under linear transformation.

26 At least, this is the way it seemed to Einstein in the 1910s. Cartan and Friedrichs later showed that Newtonian theory could be given a generally covariant formulation, so that the problem of recovering the Newtonian limit from Einstein's theory takes on a different cast. See (Havas 1964) and also "The Story of Newstein …" (in vol. 4 of this series).

...As Coordinate Restrictions

The same equation (10) and (11) can be used in a different way. Einstein's goal in the notebook is a theory with sufficient covariance to satisfy a generalized principle of relativity. General covariance supplies more covariance than he needs; it includes covariance under transformations not associated with changes of states of motion, such as the transformation from Cartesian spatial coordinates to radial coordinates. So Einstein can afford to use the generally covariant expressions of the Ricci Levi Civita calculus merely as intermediates. If those expressions are not themselves suitable for his theory, then he can simplify them to generate others of somewhat less covariance that are. The generation of the November tensor T_{il}^{x} on page 22R is an example. The Ricci tensor itself appeared unsuitable as a gravitation tensor. There proved to be a way of splitting the tensor into two parts each of which is a tensor under unimodular transformations. Since Einstein was willing to accept unimodular covariance instead of general covariance, he could select one of these parts as his gravitation tensor.

The equations (10) and (11) could be used in the same way. If the Ricci tensor or the tensor T_{il}^{x} proved unsuitable as a gravitation tensor, why not sacrifice a little more covariance to produce expressions that are suitable? Conjoining (10) or (11) to their associated tensors produces simpler expressions. The tensor T_{il}^{x}, for example is reduced to (15). If Einstein selected this reduced form as his gravitation tensor, then he assured recovery of the Newtonian limit. The gravitation tensor has the required form (8).

The cost of using equations (10) and (11) in this way is a further sacrifice of covariance. The final equations will have no more covariance than the coordinate restrictions (10) and (11). Whether these have sufficient covariance to support an extension of the principle of relativity cannot easily be read by inspecting equations (10) and (11). It is a matter of computation.

2.2 The Evidence for Einstein's Use of Coordinate Restrictions

There is strong evidence that Einstein used equation (11) and another similar restriction as a coordinate restriction, that is, as a universal restriction not limited to the special case of the Newtonian limit.

The Non-Linear Transformation of Equation (11)

The first piece of evidence is on page 22L. There Einstein undertakes a simple calculation. He writes down two equations. The second is $|p_{\mu\nu}| = 1$. Since, in Einstein's notation,

$$|p_{\mu\nu}| = \frac{\partial x'_{\mu}}{\partial x_{\nu}},$$

this is just the condition that the transformation $x_{\nu} \rightarrow x'_{\mu}$ be unimodular. The first is equation (11) in the primed coordinate system. Einstein then expands this equation in terms of unprimed quantities and the coefficients of the transformation $p_{\mu\nu}$ and their inverses.

The calculation is incomplete and its outcome obscure. Its purpose is not obscure and that is all that matters here. Einstein is checking the covariance of equation (11) within the domain of unimodular transformations. If Einstein intended (11) to be a coordinate condition, it is hard to see why he would concern himself with its transformation properties. The role of equation (11) as a coordinate condition is merely to assist in reducing the covariance of the theory to enable recovery of the Newtonian limit. Galilean covariance only is required in that Newtonian limit. Einstein can be assured of this much covariance. Galilean transformations are a subset of the linear coordinate transformations. Einstein can determine rapidly that equation (11), used as a coordinate condition, will give him that much covariance. The calculation is trivial. It merely requires noticing that the coefficients

$$|p_{\mu\nu}| = \frac{\partial x'_{\mu}}{\partial x_{\nu}}$$

and their inverses

$$\pi_{\mu\nu} = \frac{\partial x_{\nu}}{\partial x'_{\mu}}$$

are constants under linear transformation. Therefore the quantity in equation (11) transforms as a vector under linear transformation since

$$\sum_{\kappa}\frac{\partial \gamma'_{\kappa\alpha}}{\partial x'_{\kappa}} = \sum_{\kappa\rho\sigma\tau}\pi_{\kappa\rho}\frac{\partial(p_{\kappa\sigma}p_{\alpha\tau}\gamma_{\sigma\tau})}{\partial x_{\rho}} = \sum_{\kappa\rho\sigma\tau}p_{\kappa\sigma}p_{\alpha\tau}\pi_{\kappa\rho}\frac{\partial\gamma_{\sigma\tau}}{\partial x_{\rho}} = \sum_{\sigma\tau}p_{\alpha\tau}\frac{\partial\gamma_{\sigma\tau}}{\partial x_{\sigma}},$$

where we use $\sum_{\kappa}P_{\kappa\sigma}\pi_{\kappa\rho} = \delta_{\sigma\rho}$. Hence, if this quantity vanishes in one coordinate system as (11) requires, it will vanish in any coordinate system to which one transforms with a linear transformation.

Einstein cannot have had this simple linear case in mind on page 22L. For the calculation there clearly allows non-constancy of the coefficients $p_{\mu\nu}$; he does not eliminate the derivative terms

$$\frac{\partial P_{\mu\alpha}}{\partial x_i},$$

which vanish automatically for linear transformations. This concern is unintelligible if equation (11) is being used as a coordinate condition. The concern is explained quite simply if that equation is being used as a coordinate restriction. The quantity in (15), the tensor T_{il}^{x} after reduction by coordinate restriction (11), is his gravitation tensor. He is computing its covariance the easy way. By its construction, this candidate gravitation tensor will transform as a tensor under unimodular transformations that leave equation (11) unchanged. If the candidate gravitation tensor is to allow a generalization of the principle of relativity, its covariance must include non-linear transformations.

The Theta Requirement

The result of the calculation on page 22L cannot have been encouraging for the combination of the tensor T_{il}^{x} and condition (11) receive no further serious attention in the notebook. Instead, on page 23L, Einstein introduced another way of recovering the Newtonian like expression (8) from T_{il}^{x} that did not require use of equation (11). That it not be required was apparently of some importance since, in a document of pure calculation with essentially no explanatory text at all, Einstein went to the trouble to explain in writing

$$\text{“}\sum \frac{\partial \gamma_{\kappa\alpha}}{\partial x_{\kappa}} = 0 \text{ is not necessary.”}$$

In its place Einstein introduced a coordinate restriction of another type. He constructed the quantity

$$\vartheta_{i\kappa\lambda} = \frac{1}{2}\left(\frac{\partial g_{i\kappa}}{\partial x_{\lambda}} + \frac{\partial g_{\kappa\lambda}}{\partial x_{i}} + \frac{\partial g_{\lambda i}}{\partial x_{\kappa}}\right) \tag{16}$$

and required that transformations between coordinates be so restricted that this quantity $\theta_{i\kappa\lambda}$ transform as a tensor. (We shall call this the "theta requirement," the "theta condition" or the "theta restriction" according to its interpretation.) He then proceeded to show by adding and discarding terms in $\theta_{i\kappa\lambda}$ from T_{il}^{x} how one could construct a quantity

$$\sum \frac{\partial}{\partial x_{\kappa}}\left(\gamma_{\kappa\alpha}\frac{\partial g_{il}}{\partial x_{\alpha}}\right) + \sum \gamma_{\rho\alpha}\gamma_{\kappa\beta}\frac{\partial g_{i\kappa}}{\partial x_{\alpha}}\frac{\partial g_{l\rho}}{\partial x_{\beta}} \tag{17}$$

that is a tensor under unimodular transformations for which $\theta_{i\kappa\lambda}$ transforms tensorially. Einstein's efforts have produced another expression in the form of (8), apparently yet another candidate for the gravitation tensor, at least in the Newtonian limit.

Its Relation to Rotational Covariance…

Through another part of the notebook we also learn what apparently interested Einstein in the requirement that $\theta_{i\kappa\lambda}$ transforms tensorially. The simplest requirement of this type would be to ask that the quantity $\partial g_{i\kappa}/\partial x_{\lambda}$ transform as a tensor. But that, perhaps, was an excessively restrictive. It is easy to see that this quantity transforms as a tensor only under linear transformations of the coordinate systems. If one sought a natural weakening of this requirement, the simplest weakening is to consider just the symmetric part of $\partial g_{i\kappa}/\partial x_{\lambda}$, which is (up to multiplicative factor) the quantity $\theta_{i\kappa\lambda}$.[27] One might hope that the weakened requirement would now admit other interesting transformations, such as those to coordinate systems in uniform rotation. More explicitly, these are the transformations that take the coordinates $x_{\alpha} = (x, y, z, t)$ to a new coordinate system $x'_\alpha = (x', y', z', t')$ in uniform rotation at angular velocity ω about the z axis

$$x' = x\cos\omega t + y\sin\omega t \quad y' = -x\sin\omega t + y\cos\omega t \quad z' = z \quad t' = t. \tag{18}$$

That this is Einstein's hope is revealed, apparently, by calculations on pages 42L–42R of the notebook. Einstein sets up and solves the following problem: what are all the metrics of unit determinant that satisfy the conditions[28]

$$\frac{\partial g_{ik}}{\partial x_4} = 0 \text{ and } \vartheta_{i\kappa\lambda} = \frac{1}{2}\left(\frac{\partial g_{i\kappa}}{\partial x_\lambda} + \frac{\partial g_{\kappa\lambda}}{\partial x_i} + \frac{\partial g_{\lambda i}}{\partial x_\kappa}\right) = 0? \tag{19}$$

The problem posed by Einstein is a reformulation of this more interesting problem: assume we start from the metric $g^{SR}_{\mu\nu}$. To which time (= x_4 coordinate) independent metrics can we transform by means of unimodular coordinate transformations for which $\theta_{i\kappa\lambda}$ transforms as a tensor? Since $g^{SR}_{\mu\nu}$ has constant coefficients, we have $\theta_{i\kappa\lambda} = 0$, so that the requirement that $\theta_{i\kappa\lambda}$ transforms as a tensor reduces to the requirement that $\theta_{i\kappa\lambda}$ remain the zero tensor. Thus the metrics to which we can transform must satisfy (19). Apparently Einstein hoped that these transformations would include the unimodular transformation (18), so that this class of metrics would include what we can call a "rotation field", the form of the metric $g^{SR}_{\mu\nu}$ that results when it is transformed by the rotation transformation (18)

$$g^{ROT}_{\mu\nu} = \begin{bmatrix} -1 & 0 & 0 & \omega y' \\ 0 & -1 & 0 & -\omega x' \\ 0 & 0 & -1 & 0 \\ \omega y' & -\omega x' & 0 & c^2 - \omega^2(x'^2 + y'^2) \end{bmatrix}. \tag{20}$$

…Is Not Close Enough

And Einstein's hopes were almost realized. The result of his calculation was that the two conditions (19) were satisfied by a metric whose coefficients in its *covariant* form equaled those of the *contravariant* form of $g^{ROT}_{\mu\nu}$ that is $\gamma^{ROT}_{\mu\nu}$. This was close to showing that the transformations under consideration would allow the transformation from $g^{SR}_{\mu\nu}$ to $g^{ROT}_{\mu\nu}$. But it is not good enough for a mathematical result such as this to be close. It either succeeds or fails—and this one failed. Einstein revealed his frustration by remarking in one of the few textual comments in the notebook of calculations, "Schema of γ for a rotating body identical with the adjacent g- schema!" The

27 Another advantage is that the symmetrized form $\theta_{\iota\kappa\lambda}$ of $\partial g_{i\kappa}/\partial x_\lambda$ is very similar in structure to the Christoffel symbols, so that the Christoffel symbols in T^x_{il} can readily be rewritten in terms of $\theta_{i\kappa\lambda}$, easing the course of the calculations.

28 Einstein also suppresses the x_3 coordinate.

exclamation remark is very unusual and flags Einstein's surprise and, probably, disappointment.[29]

The Theta Requirement is Not a Coordinate Condition

We can reconstruct the content of these calculations fairly confidently. But their purpose is quite mysterious if we assume that the theta requirement is simply a coordinate condition being used to reduce T_{il}^{x} to the Newtonian form (8) for the case of the Newtonian limit. There are two problems. First, if the theta requirement has this purpose, then there is no need to investigate its covariance under rotation transformations (18) or, for that matter, to contrive the condition to have this covariance. Linear covariance is sufficient for the Newtonian limit and it is obvious without calculation that the theta condition has that much covariance.[30] Nonetheless, lack of rotational covariance seems to have been fatal to the proposal of the theta condition.

The second problem is that the reduction of T_{il}^{x} to (17) is not the calculation that would be undertaken if the theta condition were a coordinate condition. In that case, one would merely seek the expression to which T_{il}^{x} reduced in coordinate systems compatible with the condition. Expression (17) is not that expression. In generating it, Einstein freely added terms in $\theta_{i\kappa\lambda}$ so contrived as not to disturb the covariance of the resulting expression under these transformations. One cannot revert to T_{il}^{x} merely by relaxing the constraint of the theta restriction. In short, (17) is guaranteed to transform tensorially under these restricted transformations, but it is not the quantity one would seek if one chose T_{il}^{x} as the gravitation tensor and sought its Newtonian limit through a coordinate condition.

Both these problems are resolved immediately if we assume that Einstein is using the theta requirement as a coordinate restriction. The expression (17) is his candidate gravitation tensor. It can have no more covariance than the theta condition, so an investigation of the latter's covariance is, indirectly, an investigation of the covariance of the candidate gravitation tensor. Moreover T_{il}^{x} is merely an intermediate used in the construction of the candidate gravitation tensor (17). There is no need to ensure that this latter expression be a form of T_{il}^{x} in a restricted class of coordinates. Einstein's goal is merely a quantity of Newtonian form (8) with as much covariance as the theta condition. Einstein can add terms in theta freely if they allow a simpler final result, for these terms do not compromise the covariance of the result.

29 That this result proved fatal to the proposal of the theta restriction is confirmed by the calculations that follow on page 43L. There Einstein attempts to define a contravariant form of $\theta_{i\kappa\lambda}$ and begins to check whether it might be able to reduce the tensor T_{il}^{x} if used in the same way as $\theta_{i\kappa\lambda}$ in the original theta restriction. Presumably Einstein chose a contravariant form of $\theta_{i\kappa\lambda}$ as a replacement of the failed $\theta_{i\kappa\lambda}$ in the hope that a calculation analogous to that on pages 42L–42R would yield the correct covariant form of $g_{\mu\nu}^{ROT}$.

30 The deep concern with the covariance of the theta condition is also evident on the page facing the one on which the theta restriction is introduced. That facing page, 23R, is given over to computation of the transformation behavior of $\theta_{i\kappa\lambda}$ under infinitesimal transformations.

2.3 The Problem of the Newtonian Limit

How can the notion of coordinate restriction help us understand why Einstein rejected T^x_{il} as a candidate gravitation tensor in the notebook? In particular, how can it help us to understand Einstein's remark to Sommerfeld that the tensor did not yield the Newtonian limit when page 22R of the notebook appears to contain the calculation that shows how to recover the Newtonian limit? That is, it shows how to use equation (11) to reduce T^x_{il} to a Newtonian form, just as Einstein would in his paper of 4 November 1915.

The answer is simple. The expression T^x_{il} does not have the Newtonian form (8) and that may already be sufficient to explain Einstein's remark. Indeed, in addition to problems of energy conservation, Einstein may also have succumbed at this point to the temptation to multiply out the Christoffel symbols in an effort to get closer to an expression of the Newtonian form (8). If equation (11) is being used as a coordinate restriction in this effort, then T^x_{il} has ceased to be Einstein's candidate gravitation tensor. The new gravitation tensor is its reduced form, expression (15). While the formal manipulation of the reduction to expression (15) is the same in the notebook and in the 4 November 1915, paper, their interpretations would be very different. In 1915, the calculation shows how to recover the Newtonian limit from T^x_{il}. In the notebook, the calculation merely used T^x_{il} as an intermediate to generate a new candidate gravitation tensor, expression (15).

What was the fate of this new candidate gravitation tensor? It does not survive beyond page 22R. The notion of coordinate restriction can help us surmise its fate. If expression (15) is to be used as a gravitation tensor, it is of the greatest importance to determine its covariance. As we have seen, that is determined indirectly by investigating the covariance of the coordinate restriction (11). Presumably this was Einstein's goal on the facing page 22L when he probed the covariance of equation (11). We do not know how far Einstein went in these investigations. But we do know the results he would have found had he persisted. It is not too hard to see that coordinate restriction (11) is not covariant under rotation transformation (18). The simplest way to see this is to substitute $g^{ROT}_{\mu\nu}$ directly into (11). Since (11) vanishes for $g^{SR}_{\mu\nu}$, if it is covariant under rotation transformation (18), then it must also vanish for $g^{ROT}_{\mu\nu}$. But it does not. We have

$$\sum_{\kappa}\frac{\partial\gamma^{ROT}_{\kappa\alpha}}{\partial x'_{\kappa}} = \left(-\frac{\omega^2}{c^2}x', -\frac{\omega^2}{c^2}y', 0, 0\right) \neq 0.$$

We know that the rotation transformation (18) and the rotation field $g^{ROT}_{\mu\nu}$ became a topic of continued concern to Einstein on the pages following page 22. The rotation field enters indirectly on page 23L through the connection of the theta condition to the rotation field on pages 42L–42R. The rotation field is explicitly the subject of pages 24L, 24R and 25L.

It is natural to suppose that Einstein somehow came to see that his coordinate restriction (11) lacked rotational covariance, although we cannot identify a particular calculation in the notebook that unequivocally returns the result. The supposition that he had found the result would explain the strategy of the introduction of the theta condition on page 23L. Having found that his coordinate restriction (11) fails to satisfy rotational covariance, Einstein would respond by introducing a new coordinate restriction explicitly contrived to have rotational covariance. The theta condition is formulated directly as a covariance condition—Einstein will consider coordinate systems for which $\theta_{i\kappa\lambda}$ transforms as a tensor. As we saw above, the quantity $\theta_{i\kappa\lambda}$ was plausibly chosen exactly because it might yield covariance under rotation transformation (18). And we saw that Einstein remarked with evident satisfaction on page 23L that equation (11) was not needed, affirming his goal of replacing it with the theta condition.

This account of the failure of T_{il}^{x} as a gravitation tensor in the notebook is both simple and appealing. It depends crucially on one assumption: *Einstein was unaware of how to use conditions like (11) as a coordinate condition at the time of the writing of the notebook*. Without this assumption, we cannot use the notion of coordinate restrictions to explain Einstein's remark that the candidate gravitation tensor T_{il}^{x} does not yield the Newtonian limit. For, if he then understood the use of coordinate conditions, the calculation of page 22R supplied everything needed for recovery of the Newtonian limit. We must assume that he was unaware of the use to which his formal manipulation could be put.

2.4 Was Einstein Unaware of Coordinate Conditions?

I know of no evidence that decisively answers this question. So my final assessment is that we just do not know. There are weak indications, however, that point in both directions and I will try to assess them here. My view is that the case for the negative is stronger; that is, I find it most credible that Einstein was aware of possibility of using coordinate conditions.

In the Notebook

Requirements that may be either coordinate conditions or coordinate restrictions play a major role in the notebook on pages 19–23. The harmonic condition/restriction persists on pages 19–21, the requirement (11) on page 22 and then the theta requirement on page 23. The theta requirement was used as coordinate restriction and Einstein's calculation admitted no alternative interpretation of its use as a coordinate condition because of the way he added terms in $\theta_{i\kappa\lambda}$ in the course of his calculation. The calculation that used requirement (11) on page 22R is compatible with the requirement being used as either coordinate condition or coordinate restriction *or both*; the interest in the non-linear transformation of (11) on page 22L suggests its use as a coordinate restriction. There seems to be no indication that lets us decide whether the harmonic condition is used as a coordinate condition or restriction.[31] In particular, if it were used as a coordinate restriction, we might expect Einstein at some point to check its covariance in the way that he checked the covariance of requirement (11) and the theta restriction. The pages 19–21 contain no such check. Was that because he was using the requirement as a coordinate condition so that it needed no such check?

Or was is that he was too preoccupied with the ultimately fatal difficulty of recovering the Newtonian limit and energy conservation to proceed to a test of covariance?

While Einstein certainly used coordinate restrictions in the notebook, nothing in the above precludes his awareness of coordinate conditions and that he may have *also* thought of using the harmonic requirement and equation (11) as coordinate restrictions.

"Presumed Gravitation Tensor"

Of the fragments of relevant evidence in the notebook, the most important is Einstein's labeling on page 22R. There, as we saw above in expression (14), Einstein splits the Ricci tensor into two parts. The first is easily seen to be a tensor under unimodular transformation. Therefore the second is also such a tensor. Einstein labels this second quantity "Vermutlicher Gravitationstensor T_{il}^{x}" —"presumed gravitation tensor. T_{il}^{x}".

If Einstein is unaware of the use of coordinate conditions, then the identification of T_{il}^{x} as a gravitation tensor is very hard to understand. It does not have the Newtonian form (8). The derivative of the Christoffel symbol will immediately contribute three second derivative terms in the metric tensor, two more than (8) allows. This failure is not difficult to see. A Christoffel symbol is the sum of three first derivative terms. Its derivative will contain three second derivative terms in the metric tensor. Perhaps a novice in these calculations might overlook it. But Einstein is not a novice in these calculations at this stage in the notebook. In the pages preceding in the notebook he has become increasingly adept at more and more elaborate calculations involving the expansion of Christoffel symbols. On the following page 23L he devises the theta requirement. It depends on the recognition that the quantity $\theta_{i\kappa\lambda}$ and a Christoffel symbol have very similar structures so that the latter could be reexpressed profitably in terms of the former.

31 We might clutch at straws. If the harmonic requirement is used as a coordinate condition merely for the Newtonian limit, one needs to recover only the second derivative terms in the metric tensor and not the full reduced expression with first derivative terms, as Einstein does on page 19L. Or is this just Einstein being thorough and carrying a simple computation through to completion, wondering, perhaps, if the full result has an especially simple form? If the harmonic requirement were used as a coordinate restriction, then the full result would be needed, but that would still not preclude the possibility that Einstein weighed the use of the harmonic requirement as both coordinate condition and coordinate restriction. At the top of page 19R Einstein decomposes the harmonic requirement in the weak field into two equations comprising five conditions in all. That is one more than is allowed for a coordinate condition but admissible for coordinate restrictions. But since one of the new equation sets is just energy momentum conservation in the weak field, the decomposition is not necessarily an illegitimate strengthening of a coordinate condition as supplementing it with a physical requirement he demands on other grounds. Alternately but in the same spirit, that same condition, which is just equation (11) in the weak field, is a differential condition that must be satisfied by any static metric of form (4), as Einstein has already found earlier on page 39R of the notebook. On this same page 19R, he calls the harmonic requirement a "Nebenbedingung"—a "supplementary condition." If the requirement is a coordinate restriction, that is an odd label for what is as much a physical postulate as the original gravitational field equations they modify. But then perhaps that is the right way to view their action—as a universal supplement to those equations.

Perhaps this was just an oversight by Einstein. Perhaps it was haste that led him to label a manifestly inadmissible term as his presumed gravitation tensor. This supposition of haste becomes harder and harder to reconcile with what we know. At least the top half of page 22R is fairly neatly written and compact in argument, suggesting that it is not a live calculation but the record of deliberations elsewhere. Perhaps they record the outcome of a meeting with Grossmann—this is suggested by Grossmann's name on the top of the page and Einstein's later report to Sommerfeld of 28 November 1915, that he and Grossmann together had considered the gravitation tensor of this page. Einstein's failure to notice the two additional second derivative terms would have to survive whatever deliberation or meeting that produced the result and its transcription.

Yet more curious is the success of the equation (11) in reducing the tensor T_{il}^{x} to the Newtonian form (8). If Einstein chose T_{il}^{x} as a candidate gravitation tensor in haste, what sublime good fortune came with the equation (11). It just happens to be a restriction compatible both with the form he demanded for the static field and with energy conservation in the weak field, the problems that proved fatal to the harmonic requirement. And it just happened to the one that rescued his poor choice of T_{il}^{x} as gravitation tensor and allowed him to use it as an intermediate on the way to a better choice. On the supposition that Einstein was unaware of the use of coordinate conditions, we cannot presume that he already knew that equation (11) would effect this reduction. For if he already knew that, he would not label T_{il}^{x} his presumed gravitation tensor. It would just be an intermediate as the Ricci tensor itself is.

The supposition of haste and unanticipated good fortune seems necessary to make the page compatible with a lack of awareness of the use of coordinate conditions. I find this supposition incredible. I find it much more credible that Einstein simply wrote what he meant. He chose T_{il}^{x} as his gravitation tensor, fully aware of the surplus second derivative terms and fully aware, by the time of the writing of page 22R, that they could be eliminated by the condition (11). That condition (11) has this power need longer be a fortuitous coincidence. After the failure of the harmonic requirement, we may suppose that Einstein sought a tensor that could be reduced to the Newtonian form by equation (11), for that was the requirement that was manifestly compatible with energy conservation in the weak field. Surely what attracted Einstein to the gravitation tensor T_{il}^{x} was exactly the fact that condition (11) allowed its reduction to the Newtonian form (8) and its selection as a presumed gravitation tensor resulted from working backwards from this result.

If we accept this last version of the story, then we accept that Einstein intended to use requirement (11) on page 22R as a coordinate condition and only later considered using it as a coordinate restriction.

Einstein's Later Discussion and Treatment of Coordinate Conditions

If the content of the notebook allows no final decision, we might look for evidence elsewhere. If Einstein were unaware of the use of coordinate conditions and this played some role in his failure, we might expect some trace of it in his later recollections and writings. He would have failed to see what later became his standard method for recovery of the Newtonian limit. Many of the errors of the notebook and *Entwurf* theory are mentioned later. He remarks both in correspondence and in his publications on his surprise that static fields turn out not to be spatially flat, (see Norton 1984, §8). He eventually also puts some effort into explaining to his correspondents how he erred in the "hole argument" and an enduring trace of this correction was his publication of the "point-coincidence argument," see section 3.2 below and (Norton 1987). I know of no place in which Einstein directly allows that he was unaware of the use of coordinate conditions when he devised the *Entwurf* theory.

What were the errors he corrected when he returned to the tensor T^x_{il}? A problem with the Newtonian limit accrues a brief mention in his letter to Sommerfeld. The real force of Einstein's correction in that letter lies in his confession of the $\{\}$ prejudice. That he regarded this error as decisive is affirmed by the fact that it also is discussed at some length in the text of the paper of 4 November 1915, as we say above. In stark contrast, the use of coordinate conditions gets no mention in this correspondence. In the 4 November paper, the correct use of coordinate conditions is introduced with an indifference that suggests he thinks their use entirely obvious.[32] His *complete* discussion is merely:

> [Through this previous equation] the coordinate system is still not determined, in that four equations are needed for their determination. Therefore we may arbitrarily stipulate for the first approximation
>
> $$\sum_{\alpha\beta} \frac{\partial g^{\alpha\beta}}{\partial x_\beta} = 0.^{33} \qquad [(11)]$$

If Einstein had suffered a failure to see that equation (11) could be used this way for almost three years, would he not offer some elaboration if only to assure readers that the manipulation is admissible? Or should we assume that Einstein was feeling too vulnerable at this crucial time in his theory's development to admit all his former errors?[34]

---

32 This nonchalant attitude persisted into his review article (Einstein 1916, E§21), where the recovery of the Newtonian limit is formally incomplete exactly because Einstein neglects to invoke a coordinate condition. Einstein considers just the first term of the tensor T^x_{il} as part of his recovery of the Newtonian limit. He observes correctly that from it one recovers Poisson's equation of Newtonian theory, $\Delta g_{44} = \kappa\rho$ (where Δ is the Laplacian, κ a constant and ρ the mass density) by considering just the 44 component in the case of a time (x_4) independent field. That observation is insufficient for the recovery of the Newtonian limit. One must also establish that the remaining components of the field equations do not impose further conditions that restrict Poisson's equation in a way incompatible with the Newtonian limit. This further step requires a coordinate condition and that Einstein simply neglects to introduce or even mention. Einstein's later textbook exposition (Einstein 1922, 87) does give a serviceable account of how coordinate conditions are used to reduce the gravitational field equations to a Newtonian form, but without any special care that would suggest he thought the matter delicate.

33 "Hierdurch ist das Koordinatensystem noch nicht festgelegt, indem zur Bestimmung desselben 4 Gleichungen nötig sind. Wir dürfen deshalb für die erste Näherung willkürlich festsetzen [(11)]." (Einstein 1915a, 786)

The Entwurf Theory

What is striking about the *Entwurf* theory is that it does not require coordinate conditions for the recovery of the Newtonian limit. Its gravitation tensor already has the Newtonian form (8). So merely presuming a weak field of form (7) indirectly introduces enough of a restriction on the coordinate system to allow recovery of the Newtonian limit.

This striking feature of the *Entwurf* theory and Einstein's silence on coordinate conditions would be explained quite simply by the supposition that Einstein was then unaware of the use of coordinate conditions. But both could also be explained in another way. If he decided in favor of the *Entwurf* field equations for other reasons, then he might well never mention the use of coordinate conditions simply because he had no occasion to. Indeed, even in his later theory which did require them, he seemed quite reluctant to say anything more than the absolute minimum about them.

One thing that we cannot infer from the *Entwurf* theory and his writings associated with it is that Einstein was somehow naive about coordinate systems and how one might introduce a specialized coordinate system. We shall see below in section 3.6 that Einstein explained both in print and correspondence that he understood that equations of restricted covariance must correspond to generally covariant equations if they are to be anything more than just a restriction on the choice of coordinate system. He also made quite clear that he understood the subtleties of introducing specialized coordinate systems. That is, they might be introduced in two ways. In one way, one merely chooses to consider a restricted class of the coordinate systems already available; this decision does not alter the geometry of the spacetime. In the second way, one demands that this geometry must be such that it admits a coordinate system of a particular type; this demand indirectly applies a further and often profound restriction to the geometry of the spacetime.

If Einstein was unaware of the possibility of using coordinate conditions, it was not part of a broader blindness or lack of sophistication concerning coordinate systems.

What is More Plausible?

Since none of these items of evidence is decisive, we should also ask after the plausibility of different answers. Here our personal Einsteins speak as much as evidence. One might be comfortable with an Einstein unaware of the possibility of coordinate conditions. They never appear unequivocally in the notebook—although the labeling of T_{il}^{x} on page 22R as the "presumed gravitation tensor" is, in my view, very hard to explain if the initial intent was not to use a coordinate condition. So perhaps, on a principle of parsimony, we attribute the least knowledge we need to Einstein.

I find the supposed lack of awareness quite implausible. Coordinate systems and covariance requirements are Einstein's great strength and the locus of his deepest

---

34 Einstein did not explain in this paper where his "hole argument" against general covariance had erred. Below (see section 3.7) I will suggest that this reticence may have reflected a difficulty in seeing clearly what the problem was and this difficult will be a part of the account developed there.

reflection. As we shall recall below in section 3.4, the essential goal in devising his general theory of relativity was the elimination of the preferred inertial coordinate systems of Newtonian theory and special relativity, which is in turn sustained by their limited covariance. It is fundamental to his entire research project that his final theory not harbor them. So how then is Einstein to recover the Newtonian limit from his theory? He must introduce specialized assumptions that obtain only in the case of the Newtonian limit and restores the characteristically Newtonian elements. One assumption is that the metric field have the specialized weak form of (7). He must also reduce the covariance of theory and thereby reintroduce exactly the preferred coordinate systems he had labored so hard to eliminate. Einstein's knew how to restrict covariance. It is done partly in the coordinate dependence of the metric given as (7). It is done explicitly through a set of four equations such as the harmonic requirement or equation (11). But is it really possible that Einstein would fail to notice that he need only impose these covariance restricting requirements in the circumstances of the Newtonian limit? He would see that a specialized form of the metric is admissible in these special circumstances. But he must somehow overlook that a restriction of covariance in these special circumstances is also admissible.

Mistakes and oversights are all too common in science. We enter them into the historical record readily when we have evidence for them. This is one for which we have no unequivocal evidence and we have indications that speak against it. It must happen in Einstein's area of greatest expertise and concern. And it must not be a momentary lapse. It must persist[35] into the development of the *Entwurf* at least up to the development of his general arguments against general covariance later in 1913.

A Transition from Coordinate Conditions to Coordinate Restrictions?

Our evidence on Einstein's awareness or lack of awareness of the use of coordinate conditions remains incomplete. Yet all these considerations make it most credible that Einstein was aware of their use and could have considered using requirements such as the harmonic and equation (11) as both coordinate conditions and coordinate restrictions. Let us go a little further. If we had to choose a single narrative that would fit best with the progression of calculations in the notebook, it would be this.

When the harmonic requirement is introduced on page 19L, it is used as a coordinate condition, with Einstein perhaps reserving the possibility of using it as a coordinate restriction if that should prove viable and simpler. On page 22R, requirement (11) is introduced as a coordinate condition with T^x_{il}, chosen as the gravitation tensor. However he is unable to see how to use T^x_{il} as his gravitation tensor. So he decides he must look for simpler expressions. He reverts to use of requirement (11) as a coordinate restriction so that he can use the simpler gravitation tensor (15), the

35 Thoughts of the use of condition (11) did not leave Einstein after the *Entwurf* theory was completed. As late as August 1915, he recalled in a letter to Paul Hertz how he had struggled with this condition, Einstein to P. Hertz, 22 August 1915, (CPAE 8, Doc. 111). Presumably this continued presence facilitated revival of T^x_{il} in November 1915.

reduced form of T_{il}^{x}. That also proves inadmissible, presumably because of its restricted covariance. So, on the following page, Einstein introduces the theta restriction, which can only be a coordinate restriction. It is especially contrived to have the covariance that requirement (11) lacked.

What makes it credible that page 22R is the turning point is Einstein's labeling of T_{il}^{x} as the "presumed gravitation tensor" when he must have known already that equation (11) would reduce it to the Newtonian form. That suggests that equation (11) was first introduced as a coordinate condition. The investigation of its covariance properties on page 22L marks the decision to treat the requirement as a coordinate restriction.[36] In the earlier pages 19–21, the harmonic requirement could have been either coordinate condition or restriction. Nothing in the calculations would have committed Einstein to either. The lack of interest in the covariance properties of the harmonic requirement suggests that Einstein had less interest in its use as a coordinate restriction.

If these suppositions are correct, then they bear directly on the "mathematical" and "physical strategy" we describe Einstein as using elsewhere in these volumes. The use of coordinate conditions would be associated with the mathematical strategy in its purest form. If recovery of the Newtonian limit will be through the harmonic condition, for example, then Einstein is able to use the full Ricci tensor as his gravitation tensor and not a simpler reduced form. With his failure to see that the Ricci tensor or that T_{il}^{x} are viable gravitation tensors, Einstein begins to withdraw from the mathematical strategy towards the physical strategy. The use of coordinate restrictions represents an intermediate stage in that withdrawal. He is still trying to use the gravitation tensors naturally suggested by the mathematical formalism, but now in reduced form. The failure of these last efforts leads him to revert to the physical strategy.

3. A CONJECTURE: THE HOLE ARGUMENT AND THE INDEPENDENT REALITY OF COORDINATE SYSTEMS

The Puzzle Continues

If we accept that Einstein knew about the possibility of using coordinate conditions, page 22R once again presents us with a troubling puzzle. In his later recollection to

36 Is there a trace of two stages of calculation on page 22R? The calculations there are divided by a horizontal rule. The calculations above the rule deal only with the term in T_{il}^{x} that contains second derivatives of the metric tensor and its reduction by equation (11) to the Newtonian form (8). Those below deal with expansion of the terms quadratic in the Christoffel symbols in T_{il}^{x}. The calculations above the rule are the ones needed if equation (11) is to be used as a coordinate condition; in the Newtonian limit all that matters are the terms in the second derivatives of the metric. The ones below are needed in addition if (11) is used as a coordinate restriction; they give the full expression for the reduced form of T_{il}^{x}. The calculations above the rule are neater and, as I suggested earlier, may just report discussions and calculations conducted elsewhere. Those below the rule are massively corrected and have the look of live calculations. Were they undertaken later after Einstein had decided to revert from the use of (11) as a coordinate condition to a coordinate restriction?

Sommerfeld, Einstein reports that he had been unable to recover the Newtonian limit from the gravitation tensor T^x_{il}. But page 22R contains just the calculation that seems to do this. As we saw in section 2, the remark could be explained using the notion of coordinate restrictions. But that explanation fails if we accept that Einstein was aware of the use of coordinate conditions. So how can we reconcile his later recollection with the content of the notebook?

There is a further aspect of page 22R that bears cautious reflection. Page 22R should have been a great triumph for Einstein. He had labored since page 14L through calculations of great complexity in an effort to recover a gravitation tensor from the Riemann tensor. The problem seemed to yield on page 19L with the introduction of the harmonic condition and the easy reduction of the Ricci tensor to a quantity of Newtonian form (8). But the success faded over the following pages in the face of a final hitch that grew to be fatal. He could not see how to reconcile the harmonic condition with the form he expected for the static field, the weak field equations and energy conservation in the weak field. On page 22R he finally had the answer to that last hitch. By choosing T^x_{il} as his gravitation tensor, he could replace the harmonic condition with condition (11) and this new condition resolved all the earlier problems, since it was both compatible with the form expected for static fields and with energy conservation in the weak field. The solution was so unobjectionable that he published it upon his return to general covariance in November 1915. But in the notebook, that successful solution is just abandoned and apparently quite hastily. His later recollections explain this decision in terms of the $\{\}$ prejudice. Just when he had everything else working, he gave up because, on the best reconstruction, he could not see how to extract an energy-momentum tensor for the gravitational field from the tensor. He gave up more than just the gravitation tensor T^x_{il}. He seems to have given up the use of coordinate conditions entirely and with it the easy access to the gravitation tensors of broad covariance naturally suggested by the mathematical formalism. If the $\{\}$ prejudice was all there was to it, Einstein had lost his customary tenacity and become fickle or faint hearted or both.

Might there have been a further difficulty that compromised the recovery of the Newtonian limit and that he did not report?

Another Error?

Might we find another error or misconception that Einstein may have committed that would give answers to both the above questions? Of course it is always possible to invent hidden errors varying from the trivial slip to the profound confusion, tailor made to fit this or that aberration. The real difficulty is to establish that the error was really committed.

If there is such an error, we would expect it to be somehow associated with the use and understanding of coordinate systems. We do know of a serious misconception concerning coordinate systems that drove Einstein away from general covariance during the years of the *Entwurf* theory. This was the misconception that supported the hole argument. Months after the completion of the *Entwurf* theory, Einstein introduced this argument as a way of showing that the achievement of general covariance in his gravitation theory would be physically uninteresting. After he had returned to general covariance Einstein explained the error of the hole argument. He had unwittingly accorded an independent reality to spacetime coordinate systems and this had compromised his understanding of what is represented physically in a transformation of the fields of his theory.[37] In our histories to date, this error affected Einstein only through the hole argument and thus well after Einstein's turn away from general covariance in 1912 and 1913. However Einstein's theory was, on his own report, dependent intimately and fundamentally on the transformation of fields and spacetime coordinates. Is it possible that Einstein's misconception on the independent reality of coordinate systems had earlier damaging effects?

The Conjecture

The conjecture to be advanced here is that Einstein's misconception about the independent reality of coordinate systems did not just exert its harmful influence with Einstein's discovery of the hole argument, well after the *Entwurf* theory was in place. Rather I shall urge that it decisively misdirected Einstein's investigations at a much earlier stage, the time of the calculation of the notebook. I believe that it can explain why Einstein abandoned the use of coordinate conditions so precipitously, why he would have judged the calculation concerning the Newtonian limit of page 22R to be a failure and why he acquiesced so readily to the gravely restricted covariance of the *Entwurf* theory. Einstein failed to see this error until 1915. Until then it precluded his use of coordinate conditions. It led him to expect that any coordinate condition must have sufficient covariance to support an extension of the principle of relativity to acceleration.

More specifically, I will suggest that when Einstein applied a coordinate condition such as (11), he unwittingly accorded an independent existence to the coordinate systems picked out by the condition. Then, merely by repeating the same way of thinking about transformations as used in the hole argument, he would end up entertaining extraordinary expectations for each of these special coordinate systems. If some metric field $g_{\mu\nu}$ is a solution of his field equations in one of these special coordinate systems, then he would expect all its transforms (in a sense I will make clear below) $g'_{\mu\nu}$ also to be realizable as solutions in this coordinate system. A failure of the coordinate system to admit these transforms would appear as an objectionable, absolute property of the coordinate system. Such properties are just the type that Einstein had denounced in the inertial systems of classical physics and special relativity and which he prom-

37 More precisely stated: A particular set of coordinate values in a coordinate system will designate a definite physical event in spacetime. In Einstein's later view and our modern view, the physical event designated depends on the metric field; an alteration of the metric field changes which physical event is designated by these coordinate values. Einstein initially believed, however, that these same coordinate values could continue to pick out the same physical event even though the metrical field in that coordinate system was changed. That is, the coordinate system's power to pick out events is independent of the metrical field.

ised his new theory would eliminate. Now the transforms $g'_{\mu\nu}$ will only be admissible in the special coordinate system if they are compatible with the coordinate condition that defines the special coordinate system. Thus the covariance of theory as a whole had effectively been reduced to the covariance of the coordinate condition used in extracting the Newtonian limit. That condition had to be of sufficient covariance to support Einstein's hopes for a generalization of the principle of relativity to acceleration. In spite of proposals of great ingenuity in his preparation for the *Entwurf* theory, Einstein could find no combination of gravitational field equations of broad covariance and coordinate condition that satisfied these extraordinary demands.

The effect of the misconception conjectured is that coordinate conditions would lose their appeal. If a coordinate condition was used to extract the Newtonian limit, the covariance of the theory as a whole would now be reduced to that of the coordinate condition. As a result, Einstein would acquire no additional covariance for his theory in using a requirement as a coordinate condition rather than a coordinate restriction. The advantage of the latter, however, is that it delivers a gravitation tensor of considerably simpler form. Therefore I suggest that Einstein's recognition of this outcome, quite plausibly on page 22R itself, would explain why he so precipitously abandoned coordinate conditions in the notebook. The extraction of the Newtonian limit from tensor T^x_{il} via equation (11), whether it is construed as a coordinate condition or restriction, would fail for the same reason. Equation (11) would fail to have sufficient covariance.

Its Tacit Character

In the hole argument, the independent reality of the coordinate systems has a tacit, hidden character. Indeed Einstein found it hard to express explicitly what he meant. Even something as simple as the exact steps of his construction really only became clear with publication of the fourth version of the argument. It was not until after his return to covariance and possibly some prompting from his correspondents that he seemed able to give a clear account of where the argument erred. We must surely presume that, at the time of the hole argument, Einstein was simply not aware that his manipulations presumed an independent reality for his coordinate systems. It is an essential part of the present conjecture that he was not aware of the corresponding presumption earlier at the time of the calculations of the notebook. The hole argument was first offered in a hasty, ill-digested form that still led to a powerful conclusion, the inadmissibility of general covariance. The same would be true in the notebook. A similarly hasty check of the covariance of the coordinate condition would suffice to convince Einstein that disaster had struck. Its haste would allow him to overlook that his conclusion depended upon an assumption about the independent reality of coordinate systems that he would surely never endorse if it were articulated clearly.

In the Sections to Follow…

I will layout the background, context and elaboration of the conjecture. In section 3.1, I will describe the hole argument and, in section 3.2, how Einstein later diagnosed his error as the improper attribution of an independent reality to coordinate systems. In section 3.3 I will lay out the content of the conjecture in greater detail. Einstein's treatment of coordinate systems founders since it ends up ascribing absolute properties to certain coordinate systems. In section 3.4, I will review Einstein's insistence on the inadmissibility of such absolute properties, for that inadmissibility is what defeats his use of coordinate conditions. In section 3.5, I will review Einstein's early remarks on the restricted covariance of his *Entwurf* theory and his recognition that the restricted equations must correspond to generally covariant equations. I will use Einstein's mistaken attitude to the independent reality of the coordinate systems to explain his evident indifference towards finding those equations. During the reign of the *Entwurf* theory, Einstein gave several accounts of the introduction of specialized coordinate systems. In section 3.6, I will review these remarks to show that they are compatible with the present conjecture concerning Einstein's attitude to coordinate systems. Finally in section 3.7, I will review our evidence concerning the conjecture. I will conclude that we have neither decisive evidence in favor of it or against it, but weaker indications that both benefit and harm it.

3.1 The Hole Argument

Its Fullest Statement

Einstein and Grossmann's *Entwurf* paper was published mid 1913 as a separatum by Teubner (Einstein and Grossmann 1913).[38] There they reported their failure to find acceptable, generally covariant gravitational field equations. By late 1913, Einstein had found what soon became his favored vehicle for excusing this lack of general covariance, the "hole argument," which purported to show that all generally covariant gravitational field equations would be physically uninteresting.[39] Of its four presen-

38 In a letter of 28 May 1913 to Paul Ehrenfest, Einstein promises that paper will appear "in at least a few weeks" (CPAE 5, Doc. 441).

39 The earliest written and unambiguously dated mention of the hole argument is in a letter of 2 November 1913, from Einstein to Ludwig Hopf, (CPAE 5, Doc. 480). Einstein is not likely to have had the hole argument in hand much earlier than this. The hole argument supplanted another means of exonerating his theory's lack of general covariance, a consideration based on the law of conservation of energy-momentum. We know from a letter of 16 August 1913, from Einstein to Lorentz that Einstein only hit upon this earlier consideration on 15 August (CPAE 5, Doc. 470). For further discussion see (Norton 1984, §5). (The hole argument is also mentioned in the printed version of a lecture delivered on 9 September to the 96th annual meeting of the Schweizerische Naturforschende Gesellschaft in Frauenfeld (Einstein 1914b, 289). But we cannot be assured that Einstein had the hole argument at the time of the lecture since the printed version of the lecture was published many months later on 16 March 1914, see (CPAE 4, 484). Also the hole argument is not mentioned in another, briefer, printed version of the talk (Einstein 1913). That briefer version does call for a restriction on the basis of the conservation laws. It is curious that the mention of the hole argument in the printed version of (Einstein 1914b) appears in the context of the discussion of the conservation laws. In this longer and presumably later version, did Einstein strike out the consideration based on the conservation laws and write in a mention of the hole argument?)

tations, the clearest is the final version of November 1914:

> *§12. Proof of the Necessity of a Restriction on the Choice of Coordinates.*
>
> We consider a finite region of the continuum Σ, in which no material process takes place. Physical occurrences in Σ are then fully determined, if the quantities $g_{\mu\nu}$ are given as functions of the x_ν in relation to a coordinate system K used for description. The totality of these functions will be symbolically denoted by $G(x)$.
>
> Let a new coordinate system K' be introduced, which agrees with K outside Σ, but deviates from it inside Σ in such a way that the $g'_{\mu\nu}$ related to the K' are continuous everywhere like the $g_{\mu\nu}$ (together with their derivatives). We denote the totality of the $g'_{\mu\nu}$ symbolically with $G'(x')$. $G'(x')$ and $G(x)$ describe the same gravitational field. In the functions $g'_{\mu\nu}$ we replace the coordinates x'_ν with the coordinates x_ν, i.e. we form $G'(x)$. Then, likewise, $G'(x)$ describes a gravitational field with respect to K, which however does not agree with the real (or originally given) gravitational field.
>
> We now assume that the differential equations of the gravitational field are generally covariant. Then they are satisfied by $G'(x')$ (relative to K'), if they are satisfied by $G(x)$ relative to K. Then they are also satisfied by $G'(x)$ relative to K. Then relative to K there exist the solutions $G(x)$ and $G'(x)$, which are different from one another, although both solutions agree in the boundary region, *i.e. occurrences in the gravitational field cannot be uniquely determined by generally covariant differential equations for the gravitational field*. [Einstein's emphasis][40]

A Notational Convenience

The content, interpretation and significance of the hole argument has been examined extensively elsewhere.[41] Thus I will concentrate on those aspects of importance to the present conjecture. The argument depends on exploiting the defining property of a covariance group to generate new solutions of the gravitational field equations from old solutions. Assume that a transformation from coordinate system x^μ to x'^μ is

---

40 "§12. Beweis von der Notwendigkeit einer Einschränkung der Koordinatenwahl.
Wir betrachten einen endlichen Teil Σ des Kontinuums, in welchem ein materieller Vorgang nicht stattfindet. Das physikalische Geschehen in Σ ist dann vollständig bestimmt, wenn in bezug auf ein Funktion der x, gegeben werden. Die Gesamtheit dieser Funktionen werde symbolisch durch $G(x)$ bezeichnet.
Es werde ein neues Koordinatensystem K' eingeführt, welches außerhalb Σ mit K übereinstimme, innerhalb Σ aber von K abweiche, derart, daß die auf K' bezogenen $g'_{\mu\nu}$ wie die $g_{\mu\nu}$ (nebst ihren Ableitungen) überall stetig sind. Die Gesamtheit der $g'_{\mu\nu}$ bezeichnen wir symbolisch durch $G'(x')$. $G'(x')$ und $G(x)$ beschreiben das nämliche Gravitationsfeld. Ersetzen wir in den Funktionen $g'_{\mu\nu}$ die Koordinaten x'_ν durch die Koordinaten x_ν, d. h. bilden wir $G'(x)$, so beschreibt $G'(x)$ ebenfalls ein Gravitationsfeld bezüglich K, welches aber nicht übereinstimmt mit dem tatsächlichen (bzw. ursprünglich gegebenen) Gravitationsfelde.
Setzen wir nun voraus, daß die Differentialgleichungen des Graviationsfeldes allgemein kovariant sind, so sind sie für $G'(x')$ erfüllt (bezüglich K'), wenn sie bezüglich K für $G(x)$ erfüllt sind. Sie sind dann die voneinander verschiedenen Lösungen $G(x)$ und $G'(x)$, trotzdem an den Gebietsgrenzen beide Lösungen übereinstimmten, d. h. *durch allgemein kovariante Differentialgleichungen für das Gravitationsfeld dann das Geschehen in demselben nicht eindeutig festgelegt werden*." (Einstein 1914c, 1067)

41 See, for example, (Stachel 1980, §§3–4; Norton 1984, §5; 1987).

within the covariance group of the gravitational field equation and that a metric field $g_{\mu\nu}(x_\alpha)$ in the coordinate system x_μ satisfies the field equations. It follows that the metric $g'_{\mu\nu}(x'_\alpha)$ in the coordinate system x'_μ, defined by the tensor transformation law

$$g'_{\mu\nu} = \frac{\partial x_\alpha}{\partial x'_\mu}\frac{\partial x_\beta}{\partial x'_\nu}g_{\alpha\beta} \tag{21}$$

is also a solution of the field equations. These two solutions of the field equations are merely representations in different coordinate systems of the same gravitational field; it is represented by $g_{\mu\nu}(x_\alpha)$ in the coordinate system x_α and by $g'_{\mu\nu}(x'_\alpha)$ in the coordinate system x'_α. In an attempt to reduce distracting notational complications, Einstein represented the two metrics as "$G(x)$" and "$G'(x')$". His point was to draw attention to the functional dependence of the $g_{\mu\nu}$ on the coordinates x_α with the latter considered as variables; and the functional dependence of the $g'_{\mu\nu}$ on x'_α. The device is helpful, since it suppresses the various indices that play no role in Einstein's argument. I will use it below but with lower case g instead of upper case G:

$$\begin{aligned} g_{\mu\nu}(x_\alpha) &\text{ is represented by } g(x) \\ g_{\mu\nu}(x'_\alpha) &\text{ is represented by } g(x') \\ g'_{\mu\nu}(x'_\alpha) &\text{ is represented by } g'(x'). \end{aligned} \tag{22}$$

How the Argument Works

This functional dependence allows Einstein to generate a further solution of the gravitational field equations that is, apparently, physically distinct from the original field described by $g(x)$ and $g'(x')$. It is constructed by considering the solution $g'(x')$ as a set of ten functions of the four independent variables comprising the coordinate system x'. One then replaces the independent variables x' by x, so that Einstein recovers a new field in the original coordinate system x, which is $g'(x)$. Now $g'(x')$ is a solution of the gravitational field equations not because of any special properties of the coordinate system x' but merely because of the functional dependence of the $g'(x')$ on the independent variables x'. That functional form is all that generally covariant gravitational field equations consider in determining whether $g'(x')$ is admissible. By construction, $g'(x)$ shares exactly the same functional dependence on its independent variables as $g'(x')$. Thus if $g'(x')$ is a solution of the field equations so is $g'(x)$.[42]

Einstein can now complete his argument. He has two solutions of his gravitational field equations $g(x)$ and $g'(x)$, both in the *same* coordinate system x. These two solutions were constructed from the transformation x to x'. This transformation had a special property. By supposition the transformation is the identity everywhere but inside a matter free region of spacetime Σ (the "hole"), where it comes smoothly to differ from the identity. This special property confers a corresponding property on the two solutions $g(x)$ and $g'(x)$: they agree outside the hole, but they come smoothly to disagree within, for the g and g' are different functions within that hole. And since they are defined in the same coordinate system, this difference entails, Einstein urged, that they represent physically distinct gravitational fields. The result is a violation of determinism. The metric field and matter distribution outside the matter free hole fails to determine how the metric field will extend into the hole; it may extend as $g(x)$ or $g'(x)$. Einstein deemed this circumstance sufficiently troublesome to warrant rejection of all generally covariant gravitational field equations, for all generally covariant field equations will admit solution pairs $g(x)$ and $g'(x)$.

The Hole Construction

For our purposes what is important is that Einstein saw in the general covariance of the gravitational field equations an immediate license to construct the field $g'(x)$ from $g(x)$. This construction will be the focus of our attention, so I will restate it:

> *If*
>
> (a) a transformation x to x' is within the covariance group of the gravitational field equations and
>
> (b) a metric field $g(x)$ in the coordinate system x satisfies the field equations,
>
> *then*
>
> the metric field $g'(x)$ is also a solution of the gravitational field equations in the *original* coordinate system x, where the functions g' are defined by the tensor transformation law (21).

Einstein's Difficulty in Expressing the Argument

Einstein found it very hard to make clear that his hole argument depended essentially on the use of the hole construction. Rather, the three earlier versions of the hole argu-

---

42 An example illustrates the reasoning. The metric

$$g'(x') = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & (x'^1)^2 \end{bmatrix}$$

happens to be a solution of the generally covariant gravitational field equations $R_{\mu\nu} = 0$, where $R_{\mu\nu}$ is the Ricci tensor, in a coordinate system x'. What makes this g' a solution is the way that each coefficient of g' depends functionally on the coordinates x'. All coefficients are 0 or -1 excepting g'_{44} which is the square of the coordinate x'^1. We find $R_{\mu\nu}$ vanishes if we compute it for a g' with this functional dependence. It now follows immediately that the metric

$$g'(x) = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & (x^1)^2 \end{bmatrix}$$

in the coordinate system x is also a solution since it shares this same functional dependence.

ment seemed to depend on merely noticing that the two coordinate representations $g(x)$ and $g'(x')$ of the same gravitational field employed different functions. In that case the hole argument becomes the elementary blunder of failing to notice that the one gravitational field has different representations in different coordinate systems. I take this as evidence that, in his own work, Einstein did not clearly distinguish the two types of transformations $g(x)$ to $g'(x')$ and $g(x)$ to $g'(x)$. His invocation of the transformation law (21) could refer to either, without the need for explanation or apology. As we shall see below, Einstein's early presentations of the hole argument merely invoked (21) and Einstein must have presumed that readers would follow him and understand the transformation under consideration to be $g(x)$ to $g'(x)$.

The hole argument appears in Einstein's 1914 addendum to (Einstein and Grossmann 1913) where the crucial passage reads "...one can obtain $\gamma'_{\mu\nu} \neq \gamma_{\mu\nu}$ [for the metric field $\gamma_{\mu\nu}$] for at least a part of [the hole] L ...it follows...that more than one system of the $\gamma_{\mu\nu}$ is associated with the same [matter distribution]."[43] In a later version (Einstein and Grossmann 1914, 218), the corresponding passage reads "at least for a part of [the hole] L $\gamma_{\mu\nu}' \neq \gamma_{\mu\nu}$ ... so that more than one system of $\gamma_{\mu\nu}$ is associated with the same system of [stress-energy tensor] $\Theta_{\mu\nu}$ ..." Again, in the version of the hole argument of Einstein in (Einstein 1914a), the corresponding passage reads "... even though we do have $\mathfrak{T}_{\sigma\nu}' = \mathfrak{T}_{\sigma\nu}$ everywhere [for stress-energy tensor density $\mathfrak{T}_{\sigma\nu}$], the equations $g_{\mu\nu}' = g_{\mu\nu}$ are certainly not all satisfied in the interior of [the hole] Φ. This proves the assertion."[44] Fortunately Einstein gave a cryptic but sufficient clue in this last instance that he intended the failure of the equality $g_{\mu\nu}' = g_{\mu\nu}$ to be understood in the manner of the hole construction above, for he appended a footnote to the sentence containing the inequality $g_{\mu\nu}' = g_{\mu\nu}$ that read: "The equations are to be understood in such a way that the same numerical values are always assigned to the independent variables x'_ν on the left sides as to the variables x_ν on the right sides."

These presentations were sufficiently ambiguous to confuse the early historical literature on the hole argument. It interpreted Einstein as believing that the two coordinate representations of the same field, $g(x)$ and $g'(x')$, somehow represented physically distinct fields. One of the achievements of Stachel in his path-breaking paper (Stachel 1980) was to demonstrate that Einstein was not guilty of this novice blunder.[45]

43 "Daß wenigstens für einen Teil von L $\gamma'_{\mu\nu} \neq \gamma_{\mu\nu}$ ist. … zu dem nämlichen System der $\Theta_{\mu\nu}$ mehr als ein System der $\gamma_{\mu\nu}$ gehört. (Einstein and Grossmann 1913; translation from CPAE 4E, 289)

44 "… so ist zwar überall $\mathfrak{T}_{\sigma\nu}' = \mathfrak{T}_{\sigma\nu}$, dagegen werden im Inneren von Φ die Gleichungen $g_{\mu\nu}' = g_{\mu\nu}$ sicherlich nicht alle erfüllt sein. Hieraus folgt die Behauptung." (Einstein 1914a, 178; translated in CPAE 4E, 285)

45 As Stachel showed, the transformation from $g(x)$ to $g'(x')$ corresponded to what we now call a passive transformation in which the coordinate system changes but not the field. The transformation of the hole construction from $g(x)$ to $g'(x)$ corresponds to an active transformation in which the coordinate system remains unchanged but the field alters. See (Norton 1987; 1989, §2). However, as I argue in (Norton 1989, §5), it is possible to remain faithful to Einstein's purpose and wording without explicitly introducing the notions of active and passive transformations.

3.2 The Error of the Hole Argument: The Independent Reality of Coordinate Systems

Why the Hole Argument Fails

Of course the hole argument fails to establish that all generally covariant gravitational field equations are physically uninteresting. The standard resolution allows that the two fields $g(x)$ and $g'(x)$ are mathematically distinct but counters that they represent the same physical field. Thus the hole argument does not show that the field and matter distribution outside the hole leave the field within underdetermined. It just shows that the mathematical description of the field within the hole is undetermined. After his return to general covariance, Einstein argued for the physical equivalence of the fields $g(x)$ and $g'(x)$ with his "point-coincidence argument;" the two fields are equivalent since they must agree on the disposition of all coincidences, such as the intersections of the world lines of particles.[46] Alternatively, following the approach favored in Göttingen by the Hilbert school, we could argue for the equivalence of the two fields by noting that they agree on all invariant properties.[47]

Letters to Ehrenfest and Besso Explain Einstein's Error

The point coincidence argument explains how we should understand the system described in the hole argument. But it does not diagnose the error of thought that lured Einstein to interpret the system differently prior to November 1915. That diagnosis came in Einstein's letters in late 1915 and early 1916 when he explained to his correspondents how he had erred in the hole argument. In preparing his correspondent Paul Ehrenfest for the point coincidence argument, Einstein explained in a letter of 26 December 1915:

> In §12 of my work of last year, everything is correct (in the first three paragraphs) up to the italics at the end of the third paragraph. One can deduce no contradiction at all with the uniqueness of occurrences from the fact that both systems $G(x)$ and $G'(x)$, related to the same reference system, satisfy the conditions of the grav. field. The apparent force of this consideration is lost immediately if one considers that
>
> (1) the reference system signifies nothing real
>
> (2) that the (simultaneous) realization of the two different g- systems (better said, two different gravitational fields) in the same region of the continuum is impossible according to the nature of the theory.
>
> In the place of §12 steps the following consideration. The physical reality of world occurrences (in opposition to that dependent on the choice of reference system) consists *in spacetime coincidences*... [Einstein's emphasis][48]

He wrote an essentially identical explanation to his friend Michele Besso a little over a week later on 3 January 1916:

---

46 See (Norton 1987; Howard and Norton 1993, §7) for a proposal on the origin of the point-coincidence argument.

47 See (Howard and Norton 1993) for the proposal of a premature communication of this viewpoint to an unreceptive Einstein by Paul Hertz in the late summer of 1915.

> Everything was correct in the hole argument up to the last conclusion. There is no physical content in two different solutions $G(x)$ and $G'(x)$ existing with respect to the *same* coordinate system K. To imagine two solutions simultaneously in the same manifold has no meaning and the system K has no physical reality. In place of the hole argument we have the following. *Reality* is nothing but the totality of spacetime point coincidences... [Einstein's emphasis][49]

Ehrenfest proved difficult to convince of the correctness of Einstein's new way of thinking over the hole argument and Einstein needed to enter into a more detailed exchange that centered on the example of light from a star passing through an aperture onto a photographic plate.[50] In his letter of 5 January 1916, Einstein noted the instinctive attractiveness of the notion of the reality of the coordinate system:

> I cannot blame you that you still have not seen the admissibility of generally covariant equations, since I myself needed so long to achieve full clarity on this point. Your problem has its root in that you instinctively treat the reference system as something "real."[51]

Surely we are to read in this that Einstein too was misled by this instinct.

On Being Real

We learn from these letters that Einstein was under the influence of a deep-seated prejudice at the time of formulation of the hole argument: he improperly accorded a physical reality to coordinate systems. It can often be difficult to decipher precisely what is meant by an attribution of reality or non-reality—one need only recall the extended debates over realism in philosophy of science to be reminded of these diffi-

---

48 "In §12 meiner Arbeit vom letzten Jahre ist alles richtig (in den ersten 3 Absätzen) bis auf das am Ende des dritten Absatzes gesperrt Gedruckte. Daraus, dass die beiden Systeme $G(x)$ und $G'(x)$, auf das gleiche Bezugssystem bezogen, den Bedingungen des Grav. Feldes genügen, folgt noch gar kein Widerspruch gegen die Eindeutigkeit des Geschehens. Das scheinbar Zwingende dieser Überlegung geht sofort verloren, wenn man bedenkt, dass
1) das Bezugssystem nichts Reales bedeutet
2) dass die (gleichzeitige) Realisierung zweier verschiedener g- Systeme (besser gesagt zweier verschiedener Grav. Felder) in demselben Bereiche des Kontinuums der Natur der Theorie nach unmöglich ist.
An die Stelle des §12 hat folgende Überlegung zu treten. Das physikalisch Reale an dem Weltgeschehen (im Gegensatz zu dem von der Wahl des Bezugssystem Abhängigen) besteht *in raumzeitlichen Koinzidenzen*." Einstein to Paul Ehrenfest, 26 December 1915 (CPAE 8, Doc. 173). Adjusted translation from (Norton 1987, 169).

49 "An der Lochbetrachtung war alles richtig bis auf den letzten Schluss. Es hat keinen physikalischen Inhalt, wenn inbezug auf dasselbe Koordinatensystem K zwei verschiedene Lösungen $G(x)$ und $G'(x)$ existieren. Gleichzeitig zwei Lösungen in dieselbe Mannigfaltigkeit hineinzudenken, hat keinen Sinn und das System K hat ja keine physikalische Realität. Anstelle der Lochbetrachtung tritt folgende Überlegung. *Real* ist physikalisch nichts als die Gesamtheit der raumzeitlichen Punktkoinzidenzen." Einstein to Michele Besso, 3 January 1916 (CPAE 8, Doc. 178).

50 For details, see (Norton 1987, §4).

51 "Das Du die Zulässigkeit allgemein kovarianter Gleichungen noch nicht eingesehen hast, kann ich Dir nicht verübeln, weil ich selbst so lange brauchte, um über diesen Punkt volle Klarheit zu erlangen. Deine Schwierigkeit hat ihre Wurzel darin, dass Du instinktiv das Bezugssystem als etwas "Reales" behandelst." Einstein to Paul Ehrenfest, 5 January 1916 (CPAE 8, Doc. 180).

culties. But in this case the attribution of reality has quite precise consequences. When Einstein accords physical reality to a coordinate system x, this entails that the coordinate system can support two distinct fields, $g(x)$ and $g'(x)$. In particular, Einstein is committed to the x in each system of metrical coefficients representing the *same* coordinate system. This sameness entails that the two mathematical structures, $g(x)$ and $g'(x)$, represent different physical fields. Some particular set of coordinate values, such as $x^{\alpha} = (0,0,0,0)$, will pick out the same point of spacetime in each field. But, since the g and g' are different functions of the same coordinates in a neighborhood of the point, they will each attribute different properties to that point, revealing that they represent different physical fields.

In Einstein's later view it no longer makes sense to say that x represents the same coordinate system in each structure $g(x)$ and $g'(x)$. Thus we can no longer conclude that some particular set of coordinate values picks out the same point in each field and the inference to their physical distinctness is blocked.

Einstein's misconception about the independent reality of coordinate systems was clearly firmly in place towards the end of 1913, the time of his creation of the hole argument.[52] Nothing we have seen indicates that this misconception arose at that time. Rather his description of its "instinctive" character suggests that Einstein had tacitly harbored this misconception beforehand. Might this misconception have misdirected Einstein's work on his *Entwurf* theory at an earlier stage? In the following I will conjecture that it did in a quite precise way.

3.3 The Conjecture: How the Independent Reality of Coordinate Systems Defeats the Use of Coordinate Conditions

I have urged that Einstein knew about the possibility of coordinate conditions, that he used them in the notebook and then abandoned them in favor of the use of coordinate restrictions. I have even suggested that this transition may have taken place on page 22R of the notebook, in which the same requirement (11) might have been used first as a coordinate condition and then as a coordinate restriction. I now conjecture that Einstein abandoned the use of coordinate conditions because of the same error committed in the context of the hole argument. Einstein unwittingly attributed an independent reality to the coordinate systems introduced by coordinate conditions. The effect was that he mistakenly believed that the covariance of his entire theory was reduced to that of the coordinate condition. The reversion to coordinate restrictions is now natural. He mistakenly thought that using coordinate conditions to recover the

52 In describing Einstein's earlier misconception I will speak of his belief that the coordinate system has "*independent* reality," which is to be understood as asserting reality independent of the metrical field. This is because Einstein's later denial of the physical reality of the coordinate system can only apply to a reality independent of the metric. For it is entirely compatible with Einstein's later views that a coordinate system can reflect an element of reality, but only if it does so indirectly by virtue of its relation to the metric defined on the spacetime. For example, the possibility of a coordinate system in which the metrical coefficients are all constant, reflects a real property of the spacetime, its flatness.

Newtonian limit provided no greater covariance for this theory and the use of coordinate restrictions had the advantage of simplifying the equations of his theory.

The Example of T^x_{il}

To see how this notion of the independent reality of coordinate systems would defeat the use of coordinate conditions, we will look at the example of coordinate condition (11) applied to the candidate gravitation tensor T^x_{il}. The example illustrates clearly the general argument. It is also of interest in itself since I believe Einstein may well have fallen into the general mistake outlined while considering this very example.

Einstein's essential purpose in considering a structure as complicated at T^x_{il} is to achieve the broadest covariance possible for his gravitational field equations. By construction, T^x_{il} is covariant under unimodular transformations. We have seen that one particular unimodular transformation comes to special prominence in the pages immediately following the proposal of the gravitation tensor T^x_{il}. That is the transformation (18) to uniformly rotating coordinates that brings a rotation field g^{SR} (20) into being in a Minkowski spacetime.[53]

The simple reading of this covariance of the gravitational field equations in the case of a Minkowski spacetime is that it admits the transformation of $g^{SR}(x)$ to the rotation field $g^{ROT}(x')$ under the coordinate transformation (18). They are just the same Minkowski spacetime represented in two coordinate systems x and x'. However we have already seen that when Einstein speaks of such a simple transformation he may actually be referring to a more complicated transformation. In the context of the hole argument, as we saw above, when Einstein wrote about the transformation of a metric g under the transformation of the coordinates x to x', he did not just refer to the transformation of $g(x)$ to $g'(x')$. He also tacitly referred to construction of a new solution of the field equations $g'(x)$ in the original coordinate system x. Indeed Einstein seemed to treat the construction of the new field $g'(x)$ as an automatic consequence of the covariance of the gravitational field equations—so much so that, in three of four presentations of the hole argument, Einstein appears just to refer to the transformation $g(x)$ to $g'(x')$ whereas he intended to refer to the construction of the new field $g'(x)$. Thus Einstein would read the covariance of his gravitational field equations under transformation (18) as the license to take the solution $g^{SR}(x)$ of these equations and construct a new solution $g^{ROT}(x)$, both in the same coordinate system x.

Applying the Hole Construction

Einstein would see this construction as an automatic part of the covariance of his field equations, although its construction requires some manipulation as codified in what I called the "hole construction" above. We may pause here for a moment to affirm that the construction of the new solution $g^{ROT}(x)$ follows directly from the hole con-

53 I shall continue to use the abbreviation (22), so that $g^{SR}(x)$ stands for $g^{SR}_{\mu\nu}(x_\alpha)$.

struction, although Einstein would surely not have resorted to such a labored development. The two antecedent conditions (a) and (b) are satisfied as:

a) If T^{x}_{il} is chosen as the gravitation tensor, then the gravitational field equations are covariant under the transformation (18) from inertial to uniformly rotating coordinates, for this is a unimodular transformation.

b) $g^{SR}(x)$ is a solution of the source free field equations $T^{x}_{il} = 0$.[54]

It now follows that $g^{ROT}(x)$ will also be a solution of the source free field equations in the original coordinate system x.

The Independent Reality of the Coordinate Systems x^{LIM} of the Newtonian Limit…

In his evaluation of T^{x}_{il}, Einstein would have a particular class of coordinate systems in mind as admitting g^{SR} as a solution. These are the coordinate systems in which the candidate gravitation tensor T^{x}_{il} reduces to (8) in preparation for recovery of the Newtonian limit. Let us label one of these coordinate systems x^{LIM}. Thus Einstein's field equations must admit both $g^{SR}(x^{LIM})$ and $g^{ROT}(x^{LIM})$ as solutions of the source free field equations in the same coordinate system x^{LIM}.

While these results follow from a straightforward application of Einstein's 1913 understanding of covariance and coordinate systems, they have brought us close to disaster for the candidate gravitation tensor T^{x}_{il}. To complete the journey to disaster we now must now ask what it would mean to say that these source free field equations must admit both $g^{SR}(x^{LIM})$ and $g^{ROT}(x^{LIM})$ as solutions. In Einstein's *later* view (and the modern view), this could mean nothing more than the following: there exists coordinate systems x in which $g^{SR}(x)$ solves the source free field equations; and there exists coordinate systems y in which $g^{ROT}(y)$ solves the source free field equations. But there can be no physical sense in the notion that the coordinate systems x and y are the same coordinate systems. Yet the Einstein of 1912 and 1913 would be committed to the notion that the coordinate systems x^{LIM} appearing in each solution are the same coordinate systems.

There is only one resource available to give meaning to this sameness. The coordinate systems x^{LIM} of the Newtonian limit are introduced and identified in calculation by satisfaction of the coordinate condition (11). If it is really the same coordinate systems x^{LIM} appearing in each of $g^{SR}(x^{LIM})$ and $g^{ROT}(x^{LIM})$, then coordinate condition (11) must be satisfied by both $g^{SR}(x^{LIM})$ and $g^{ROT}(x^{LIM})$. In hindsight, we know that this demand is excessive. But, I conjecture, the Einstein of 1912 and 1913 did not realize this. There is a natural robustness to the application of coordinate conditions such as (11) in the modern sense that is easily mistaken for the troublesome use of the condition that I attribute to Einstein. It was legitimate in 1912 and 1913 and remains legitimate today to use the same coordinate condition to pick out the coordinate systems for the Newtonian limit in a diverse array of distinct physical situations: in the source free case, in the case of static fields, in the case of fields with propagat-

54 $g^{SR}_{\mu\nu}$ has all constant coefficients; so all its derivatives vanish and T^{x}_{il} along with them.

ing gravitational waves, in the case of a field produced by a single mass or in the case of a field produced by distributed matter; and in many more cases. Now we might use a condition such as the harmonic coordinate condition rather than Einstein's (11) but that difference is inessential to the point. In using the same harmonic condition in each of these distinct physical cases, we routinely say that we choose harmonic coordinates. Are we always aware that the harmonic coordinates of a Minkowski spacetime are not the same in any physical sense as the harmonic coordinates of a Minkowski spacetime perturbed ever so slightly by the most minute of gravitational waves? Proceeding with the tacit assumption of the independent reality of coordinate systems, Einstein could easily overlook this subtlety. It would surely be quite natural for him to presume that his coordinate condition (11) would pick out the same coordinate systems x^{LIM} in all these cases and also in the case of g^{SR} and g^{ROT}.

Treated this way, the coordinate condition (11) becomes a physical postulate that picks out a real entity, the class of coordinate systems x^{LIM}, much as the gravitational field equations pick out the gravitational fields that can be realized physically. This character of the coordinate condition (11) does not compromise our freedom to stipulate the coordinate systems that we will use in describing our fields. We are still free to choose which coordinate systems we will use and that choice can be made by accepting or rejecting a coordinate condition such as (11). But that choice is among entities that enjoy some physical reality.

…Brings Disaster and Explains Why Einstein Would Check the Covariance of His Coordinate Condition

Thus I infer that the Einstein of 1912 and 1913 would expect that the condition (11) picks out the same coordinate systems x^{LIM} in the cases of the solutions $g^{SR}(x^{LIM})$ and $g^{ROT}(x^{LIM})$. This is the disastrous conclusion. While the coordinate condition (11) holds for $g^{SR}(x^{LIM})$, we saw above that it fails for $g^{ROT}(x^{LIM})$. Einstein has arrived at a contradiction that serves as a *reductio ad absurdum* of his choice of T_{il}^{x} as gravitation tensor and the expectation that his theory is covariant under all unimodular transformations. If the theory has that degree of covariance, $g^{ROT}(x^{LIM})$ must be a solution of its source free field equations in the coordinate system x^{LIM}. But it is not. The proposed gravitation tensor has failed.

This is a failure of coordinate condition (11) to have sufficient covariance. Under the normal understanding of coordinate conditions, Einstein would have no reason to check the covariance of (11). But if Einstein accords independent reality to the coordinate system x^{LIM}, then the natural outcome is to check its covariance. If the present conjecture is correct, this explains why Einstein checked the covariance of condition (11) on page 22L, the one facing the page on which condition (11) is used to reduce T_{il}^{x} to a Newtonian form.

This contradiction between the expected and actual covariance of Einstein's theory would appear to have a particular character to Einstein, a conflict between the covariance of his theory and the ability to recover the Newtonian limit. Upon choosing T_{il}^{x} as the gravitation tensor, his entire gravitation theory would be covariant at least under unimodular transformations. That is, the gravitational field equations are covariant under unimodular transformations and the remaining equations governing energy-momentum conservation, the motion of particles and the electromagnetic field are generally covariant. However if the theory admits coordinate systems in which the Newtonian limit can be realized, then the theory loses its broad covariance. In particular, it loses covariance under transformations to uniform rotation, so that Einstein could no longer conceive of uniform rotation as a rest state, in contradiction with his requirement of a generalized principle of relativity.

The Problem Generalized

The power attributed to the coordinate condition (11) does not depend on any specific properties of the gravitation tensor T_{il}^{x} or the coordinate condition (11). The arguments rehearsed here would proceed equally with any candidate gravitation tensor of suitably broad covariance and any coordinate condition able to reduce that gravitation tensor to the form (8). Again, the argument does not require that the transformation be a rotation transformation (18). Any transformation in the covariance group of the gravitational field equations could be used. Thus, if the conjecture is correct, Einstein must have held very restrictive expectations for the covariance of his emerging general theory of relativity, whatever its gravitation tensor might be.

To find these expectations, we generalize the argument above for any gravitation tensor, any transformation in the covariance group of the resulting field equations and any coordinate condition that reduces the gravitation tensor to the form (8). For the case of the gravitation tenor T_{il}^{x}, the coordinate condition (11) picks out the class of coordinate systems x^{LIM} in which the Newtonian limit obtains and the gravitation tensor has form (6). Correspondingly for some gravitation tensor $G_{\mu\nu}$ of broad covariance, a coordinate condition $C_{\alpha} = 0$ will pick out the coordinate systems in which the Newtonian limit obtains and the gravitation tensor reduces to form (8). Since rotation transformation (18) is in the covariance group of T_{il}^{x}, Einstein would expect through the hole construction that the two metrics g^{SR} and g^{ROT}, related by this transformation, are admissible as solutions in this coordinate system x^{LIM}. But this can only obtain if the coordinate condition (11) is covariant under rotation transformation (18). Correspondingly, if g and g' are solutions of the (source free) gravitational field equations based on the gravitation tensor $G_{\mu\nu}$, Einstein would expect, through the hole construction, that they are solutions of the reduced gravitational field equations in the limit coordinate system. But this can only obtain if the coordinate condition $C_{\alpha} = 0$ is covariant under the transformation that takes g to g'. That is, Einstein would expect the following results (C1), (C2) and (C3), to obtain:

> (C1) The covariance of the theory as a whole is limited to the covariance of the coordinate condition used to pick out the coordinate systems in which the Newtonian limit is realized.

For the covariance of that coordinate condition delimits the transformations admissible for solution of the field equations in those coordinate systems.

(C2) The covariance of the gravitational field equations, *after* they have been reduced by the coordinate condition to the form (8), defines the covariance of the theory as a whole.

For these reduced gravitational field equations just result from the conjunction of the unreduced gravitational field equations and the coordinate condition so that their covariance is limited by the covariance of the coordinate condition. (In both (C1) and (C2), if the unreduced gravitational fields equations have restricted covariance, then these conditions also limit the covariance of the theory as a whole.)

(C3) In a viable theory, the coordinate condition used and the resulting reduced gravitational field equations will still exhibit broad covariance, including covariance under the rotation transformations (18), so that they admit $g_{\mu\nu}^{ROT}$ as a solution.

If the covariance required in (C1) or (C2) does not include acceleration transformations, such as the rotation transformation (18), then the theory fails to meet the demands of a generalized principle of relativity. It harbors covariance restricting coordinate systems akin to the objectionable, absolute inertial systems of classical mechanics and special relativity (see below).

If the present conjecture is correct, Einstein would adopt (C1), (C2) and (C3). The immediate outcome would be that there is no gain is using a requirement like (11) as a coordinate condition rather than a coordinate restriction. In either use, the equation will impose the same restriction on his gravitation theory's covariance. But the advantage of using coordinate restrictions is that they allow for simpler gravitational field equations.

Moreover, let us suppose that Einstein came to see (C1), (C2) and (C3) as a part of his evaluation of the candidate gravitation tensor T_{il}^{x} on page 22R. Then his natural response would be to discontinue the use of coordinate conditions, as he does after page 22R. Indeed his construction of the theta condition on page 23L would be a natural next step. He abandons coordinate conditions in favor of coordinate restrictions, so he contrives a coordinate restriction specifically to have the rotational covariance lacked by (11).

3.4 The Problem of Absolute Coordinates

The cause of the difficulty is the coordinate systems x^{LIM}, essential for the recovery of the Newtonian limit. Throughout his scientific life Einstein had railed against the objectionable, absolute properties of inertial coordinate systems. The coordinate systems x^{LIM} had now adopted just those objectionable properties and Einstein could not tolerate their presence in his theory. Einstein had made quite clear that the fundamental goal of his general theory of relativity was to eliminate exactly these preferred systems of coordinates.

His Denunciations Persist from his Early Work…

Typical of his denunciations of such systems were his remarks written in the early days of the *Entwurf* theory:

> The theory presently called "the theory of relativity" [special relativity] is based on the assumption that there are somehow preexisting "privileged" reference systems K with respect to which the laws of nature take on an especially simple form, even though one raises in vain the question of what could bring about the privilegings of these reference systems K as compared with other (e.g., "rotating") reference systems K'. This constitutes, in my opinion, a serious deficiency of this theory.[55]

The privileging of the reference system K in special relativity resides in the fact that only in K do free bodies move inertially (the "specially simple form" of the laws of motion of free bodies), whereas in K' they move under the influence of a rotation field. K and K' cannot switch roles. K cannot admit a rotation field while bodies move inertially in K'. Of course Einstein was not referring in these remarks to the special coordinate systems x^{LIM} introduced in the Zurich Notebook. However, these special coordinate systems have exactly the properties that Einstein found objectionable in K: the coordinate systems x^{LIM} admit g^{SR} so that free bodies will move inertially in x^{LIM}. But x^{LIM} does not admit the rotation field g^{ROT}.

The presence of such absolute coordinate systems would cut Einstein to the quick. In the course of nearly half a century of writing on the general theory of relativity, Einstein found the need to reappraise much of what he wrote on the foundations of his theory. His vacillations on Mach's principle are probably the best known instance. But he never wavered in his insistence that the absolute of the inertial system must be eliminated. These sentiments supported the need for a generalization of the principle of relativity to acceleration when Einstein wrote his explanatory texts:

> All of the previous considerations have been based upon the assumption that all inertial systems are equivalent for the description of physical phenomena, but that they are preferred, for the formulation of the laws of nature, to spaces of reference in a different state of motion. We can think of no cause for this preference for definite states of motion to all others, according to our previous considerations, either in the perceptible bodies or in the concept of motion; on the contrary, it must be regarded as an independent property of the

---

55 "Die gegenwärtig als "Relativitätstheorie" bezeichnete Theorie ist auf die Annahme gegründet, daß es gewissermaßen präexistierende "bevorzugte" Bezugssysteme K gebe, auf die bezogen die Naturgesetze eine besonders einfache Form annehmen, trotzdem man vergeblich die Frage aufwirft, wodurch die Bevorzugungen jener Bezugsysteme K gegenüber anderen Bezugssystemen K' (z. B. "rotierenden") bedingt sein könnte. Es liegt hierin meiner Ansicht nach ein schwerer Mangel dieser Theorie." (Einstein 1914a, 176; translation in CPAE 4E, 282). Again writing at the time of the *Entwurf* theory, Einstein expressed similar sentiments when he spoke of "...reference systems with respect to which freely moving mass points carry out rectilinear uniform motion (inertial systems). What is unsatisfactory is that it remains unexplained *how* the inertial systems can be privileged with respect to other systems." ("… Bezugssysteme zu entschlüpfen, in bezug auf welche kräftefrei bewegte Massenpunkte eine geradlinig gleichförmige Bewegung ausführen (Inertialsysteme). Das Unbefriedigende liegt dabei darin, daß unerklärt bleibt, *wieso* die Inertialsystemen ausgezeichnet sein können.") Einstein's parentheses and emphasis, (Einstein 1913, 1260; translation in CPAE 4E, 219).

> spacetime continuum. The principle of inertia, in particular, seems to compel us to ascribe physically objective properties to the spacetime continuum. Just as it was consistent from the Newtonian standpoint to make both the statements, *tempus est absolutum, spatium est absolutum,* so from the standpoint of the special theory of relativity we must say, *continuum spatii et temporis est absolutum*. In this latter statement *absolutum* means not only "physically real," but also "independent in its physical properties, having a physical effect, but not itself influenced by physical conditions."[56]

…To His Final Years

These sentiments persist essentially unchanged in the final years of his life. In a letter of 28 December 1950, Einstein explained to D. W. Sciama his concern over the latter's theory of restricted covariance; the equations held in coordinate systems in the set M but not in the forbidden set M^x:

> We now ask: on what basis can natural laws hold with respect to M but not with respect to M^x? (Logically considered, both sets M and M^x are after all completely equivalent.) If one takes the theory really seriously, there is only <u>one</u> answer: the preference for M over M^x is an independent physical property of space, which must be added as a postulate to the field equations, so that the physical theory as a whole can have a clear meaning. Newton recognized this with complete clarity ("Spacium est absolutum"). In fact, each theory based on a subgroup introduces an "absolute space", only one that is "less absolute" than classical mechanics.
>
> It was first achieved in G. R., that a space with independent (absolute) properties is avoided. There first are the laws, as they are expressed through the field equations, *complete* and require no augmenting assumptions over physical space. "Space" subsists then only as the continuum property of the physical-real (field), not as a kind of container with independent existence, in which physical things are placed.[57]

These same sentiments would apply to x^{LIM}. In resisting admission of g^{ROT}, the coordinate system would be restoring independent, absolute properties to spacetime, properties that went beyond what was given through the field equations. Einstein

---

56 "Alle bisherigen Überlegungen beruhen auf der Voraussetzung, daß die Inertialsysteme für die physikalische Beschreibung gleichberechtigt, den Bezugsräumen von anderen Bewegungszuständen für die Formulierung der Naturgesetze aber überlegen seien. Für diese Bervorzugung bestimmter Bewegungszustände vor allen anderen kann gemäß unseren bisherigen Betrachtungen in den wahrnehmbaren Körpern bzw. in dem Begriff der Bewegung eine Ursache nicht gedacht werden; sie muß vielmehr auf eine selbständige, d. h. durch nichts anderes bedingte Eigenschaft des raumzeitlichen Kontinuums zurückgeführt werden. Insbesondere scheint das Trägheitsgesetz dazu zu zwingen, dem Raum-Zeit-Kontinuum physikalisch-objektive Eigenschaften zuzuschreiben. War es vom Standpunkt Newtons konsequent, die beiden Begriffe auszusprechen; "Tempus absolutum, spatium absolutum", so muß man auf dem Standpunkt der speziellen Relativitätstheorie von "continuum absolutum spatii et temporis est" sprechen. Dabei bedeutet "absolutum" nicht nur "physikalisch-real", sondern auch "in ihren physikalischen Eigenschaften selbständig, physikalisch bedingend, aber selbst nicht bedingt"." From the popular (Einstein 1917, Ch.XXI) and the textbook (Einstein 1922, 55).

57 The typescript of the letter is EA 20–469. The autograph manuscript, EA 20 470, contains an extra sentence given in parentheses here as the second sentence, ("Logically considered..."). (EA 20–469 denotes the item with control number 20–469 in the Einstein Archive.)

would shortly characterize just such behavior as a reversion to the flawed viewpoints of Antiquity. To George Jaffé on 19 January 1954, he wrote:

> You consider the transition to the special theory of relativity as the most essential of all the ideas of the theory of relativity, but not the transition to the general theory of relativity. I hold the reverse to be true. I see the essential in the conquest of the inertial system, a thing that acts on all processes but experiences no reaction from them. This concept is in principle no better than that of the central point of the world in Aristotelian physics.[58]

3.5 The Structure and Program of the Entwurf Theory

Explaining Einstein's Indifference to General Covariance

According to the accounts developed in this volume, at the time of the creation of his *Entwurf* theory, Einstein thought rather differently from his later views on coordinate systems. There appears to be a trace of this difference in his early discussion of the limited covariance of his *Entwurf* theory. That is, he was curiously indifferent about discovering the generally covariant gravitational field equations that he believed must correspond to his *Entwurf* equations. Once Einstein has developed general arguments against the admissibility of general covariance, we need not search for a reason for this indifference. But these arguments emerged only later in 1913, after the *Entwurf* was published. We need some explanation for this indifference in the intervening months.

The accounts discussed in this paper supply them. The *Entwurf* equations would be recovered from generally covariant equations by application of a coordinate condition. So, if Einstein accorded an independent reality to the coordinate systems so introduced, then his indifference would be explained by the misapprehension that his theory overall would gain no added covariance from the transition to these generally covariant equations. Or, more simply, if Einstein was just unaware of the use of coordinate conditions, then he would be unaware of how to retain the *Entwurf* gravitational field equations for the essential case of the Newtonian limit, so the generally covariant equations would appear unusable within his theory.

The Restricted Covariance of the Entwurf theory

Einstein's exploration of T_{il}^{x} and the theta restriction are some of his final efforts in the Zurich Notebook to recover gravitational field equations from covariance considerations. These efforts halt decisively on pages 26L–26R, where Einstein laid out in capsule the derivation of the gravitational field equations of the *Entwurf* theory. This derivation uses no covariance considerations at all. It is based essentially on the

58 "Sie betrachten den Uebergang zur speziellen Relativitätstheorie als den wesentlichen Gedanken der Relativität überhaupt, nicht aber den Uebergang zur allgemeinen Relativitätstheorie. Ich halte das Umgekehrte für richtig. Das Wesentliche sehe ich in der Ueberwindung des Inertialsystems, eines Dinges das auf alle Vorgänge wirkt, von diesen aber keine Rückwirkung erfährt. Dieser Begriff ist im Prinzip nicht besser als der des Weltmittelpunktes in der Aristotelischen Physik." (EA 13 405)

demand of the Newtonian limit and energy-momentum conservation. Einstein and Grossmann arrived at a gravitation tensor of form (8)

$$\sum_{\alpha\beta}\frac{1}{\sqrt{-g}}\cdot\frac{\partial}{\partial x_\alpha}\left(\gamma_{\alpha\beta}\sqrt{-g}\cdot\frac{\partial\gamma_{\mu\nu}}{\partial x_\beta}\right)-\sum_{\alpha\beta\tau\rho}\gamma_{\alpha\beta}g_{\tau\rho}\frac{\partial\gamma_{\mu\tau}}{\partial x_\alpha}\frac{\partial\gamma_{\nu\rho}}{\partial x_\beta}$$
$$+\frac{1}{2}\sum_{\alpha\beta\tau\rho}\gamma_{\alpha\mu}\gamma_{\beta\nu}\frac{\partial g_{\tau\rho}}{\partial x_\alpha}\frac{\partial\gamma_{\tau\rho}}{\partial x_\beta}-\frac{1}{4}\sum_{\alpha\beta\tau\rho}\gamma_{\mu\nu}\gamma_{\alpha\beta}\frac{\partial g_{\tau\rho}}{\partial x_\alpha}\frac{\partial\gamma_{\tau\rho}}{\partial x_\beta}.$$

Unfortunately Einstein and Grossmann were unable to specify the covariance group of the resulting gravitational field equations. They were able to assure the reader only of covariance under linear transformation. Of course Einstein was apologetic over their failure to discover the covariance of these equations. In closing his critique of any gravitation theory based on a scalar gravitation potential, Einstein candidly conceded how this omission had crippled Einstein's program:

> Of course, I must admit that, for me, the most effective argument for the rejection of such a theory rests on the conviction that relativity holds not only with respect to orthogonal linear substitutions but also with respect to a much wider group of substitutions. But already the mere fact that we were not able to find the (most general) group of substitutions associated with our gravitational equations makes it unjustifiable for us to press this argument.[59]

What is puzzling is that the deficiency could be set aside with such a simple disclaimer. The driving force of Einstein's program was the conviction that the relativity of motion must be extended to acceleration and that this would be realized by a theory covariant under non-linear coordinate transformations, for only the latter corresponded to transformations to accelerated states of motion.

To see just how puzzling this is, we need to recall two of Einstein's commitments at this time. First we are assured by Einstein's remarks in a letter to Lorentz of 14 August 1913 of his continued commitment to a broader covariance and his alarm at his continued failure to affirm the broader covariance of the theory:

> *But the gravitational equations themselves unfortunately do not have the property of general covariance*. Only their covariance under *linear* transformations is assured. However the whole trust in the theory rests on the conviction that acceleration of the reference system is equivalent to a gravitational field. Therefore if all the systems of equations of the theory, thus also equation (18) [gravitational field equations], do not admit still other transformations aside from the linear, then the theory contradicts its own starting point; it's left hanging in the air. [Einstein's emphasis][60]

It is a measure of Einstein's frustration and desperation that the following day—15 August 1913[61]—he fell into an embarrassing error. He thought that he could

59 "Ich muß freilich zugeben, daß für mich das wirksamste Argument dafür, daß eine derartige Theorie zu verwerfen sei, auf der Überzeugung beruht, daß die Relativität nicht nur orthogonalen linearen Substitutionen gegenüber besteht, sondern einer viel weiteren Substitutionsgruppe gegenüber. Aber wir sind schon deshalb nicht berechtigt, dieses Argument geltend zu machen, weil wir nicht imstande waren, die (allgemeinste) Substitutionsgruppe ausfindig zu machen, welche zu unseren Gravitationsgleichungen gehört." (Einstein and Grossmann 1913, I§7; translation in CPAE 4E, 170–171)

establish from the requirement of energy conservation that his gravitation theory could be at *most* covariant under linear transformations. He retracted this trivially flawed argument in a paper published the following May (Einstein and Grossmann 1914, 218), but not before the argument had appeared several times in print.[62]

Correspondence with Generally Covariant Equations

Second, Einstein expressed his belief that his *Entwurf* field equations must correspond to generally covariant equations. Having presented his *Entwurf* gravitational field equations, he continued:

> It is beyond doubt that there exists a number, even if only a small number, of generally covariant equations that correspond to the above equations, but their derivation is of no special interest either from a physical or from a logical point of view, as the arguments presented in point 8 clearly show.[63] However, the realization that generally covariant equations corresponding to [these gravitational field equations] must exist is important to us in principle. Because only in that case was it justified to demand the covariance of the rest of the equations of the theory with respect to arbitrary substitutions. On the other hand, the question arises whether those other equations might not undergo specialization owing to the specialization of the reference system. In general, this does not seem to be the case.[64]

Although Einstein does not make explicit what the relation of correspondence is between the *Entwurf* equations and their generally covariant counterparts, it would surely be that the former are recovered from the latter by some kind of coordinate condition or restriction.

While these remarks come from a paper of January 1914, we have no reason to doubt that they reflected Einstein's feelings just a few months earlier at the time of completion of the *Entwurf* paper. They provide a natural interpretation of remarks

---

60 "*Aber die Gravitationsgleichungen selbst haben die Eigenschaft der allgemeinen Kovarianz leider nicht*. Nur deren Kovarianz *linearen* Transformationen gegenüber ist gesichert. Nun beruht aber das ganze Vertrauen auf die Theorie auf der Überzeugung, dass Beschleunigung des Bezugssystems einem Schwerefeld äquivalent sei. Wenn also nicht alle Gleichungssysteme der Theorie, also auch Gleichungen (18) ausser den linearen noch andere Transformationen zulassen, so widerlegt die Theorie ihren eigenen Ausgangspunkt; sie steht dann in der Luft." (CPAE 5, Doc. 467)

61 The dating is derived from Einstein's report to Lorentz in a letter of 16 August 1913 (CPAE 5, Doc. 470).

62 For discussion see (Norton 1984, §6).

63 In his point 8, Einstein had stated the hole argument and the argument against general covariance based on the conservation of energy-momentum.

64 "Es ist zweifellos, daß diesen Gleichungen eine, wenn auch geringere Zahl von allgemein kovarianten Gleichungen entspricht, deren Aufstellung aber weder vom physikalischen noch vom logischen Standpunkte von besonderem Interesse ist, wie aus den unter 8 gegebenen Überlegungen deutlich hervorgeht. Prinzipiell wichtig aber ist uns die Erkenntnis, daß den Gleichungen (6) entsprechende allgemein kovariante existieren müssen. Denn nur in diesem Falle war es gerechtfertigt, die Kovarianz der übrigen Gleichungen der Theorie beliebigen Substitutionen gegenüber zu fordern. Es entsteht andererseits die Frage, ob jene anderen Gleichungen durch die Spezialisierung des Bezugssystems keine Spezialisierung erfahren. Dies scheint im allgemeinen nicht der Fall zu sein." (Einstein 1914a, 179; translation in CPAE 4E, 286)

made by Einstein when he reflected on their failure to find generally covariant gravitational field equations:

> To be sure, it cannot be negated a priori that the final, exact equations of gravitation could be of higher than second order. Therefore there still exists the possibility that the perfectly exact differential equations of gravitation could be covariant with respect to *arbitrary* substitutions. But given the present state of our knowledge of the physical properties of the gravitational field, the attempt to discuss such possibilities would be premature. For that reason we have to confine ourselves to the second order, and we must therefore forgo setting up gravitational equations that are covariant with respect to arbitrary transformations. [Einstein's emphasis][65]

Einstein cannot mean by this that the higher order equations are incompatible with the *Entwurf* equations. For then solutions of the *Entwurf* equations would not be solutions of the higher order equations, so that each would admit a different class of physical fields. In this case, the selection of the *Entwurf* equations is just the selection of the wrong equations. It is hard to imagine that Einstein would dismiss correcting such an outright error by calling the correction "premature." But the dismissal is more intelligible if these higher order equations are the generally covariant equations that reduce to the *Entwurf* equations with the application of a coordinate condition or restriction. For then all solutions of the *Entwurf* equations would be solutions of the higher order equations; transition to the higher order equations would merely admit more coordinate representations of the same physical fields into the theory.

The Incongruity of Einstein's Approach…

If Einstein held these two views at the time of publication of the *Entwurf* theory and he also held to an essentially modern view of coordinate systems and coordinate conditions, then his assessment of the theory's state and his further development of the theory is quite mysterious. For the sole effect of a coordinate condition, in this modern view, is to obscure the covariance of the theory. As long as the coordinate condition does not extend beyond the four equations routinely allowed, it does not preclude any physical field; it merely reduces the range of coordinate representations of each

65 "A priori kann allerdings nicht in Abrede gestellt werden, daß die endgültigen, genauen Gleichungen der Gravitation von höherer als zweiter Ordnung sein könnten. Es besteht daher immer noch die Möglichkeit, daß die vollkommen exakten Differentialgleichungen der Gravitation *beliebigen* Substitutionen gegenüber kovariant sein könnten. Der Versuch einer Diskussion derartiger Möglichkeiten wäre aber bei dem gegenwärtigen Stande unserer Kenntnis der physikalischen Eigenschaften des Gravitationsfeldes verfrüht. Deshalb ist für uns die Beschränkung auf die zweite Ordnung geboten und wir müssen daher darauf verzichten, Gravitationsgleichungen aufzustellen, die sich beliebigen Transformationen gegenüber als kovariant erweisen." (Einstein and Grossmann 1913, I.§5; translation in CPAE 4E, 160) Michel Janssen has suggested an alternative interpretation: Einstein may merely mean that his *Entwurf* field equations might be good empirical approximations in the domain of weaker fields for some set of generally covariant gravitational field equations of higher order. If this interpretation is correct, we still have ample evidence from his other remarks that Einstein also expected the *Entwurf* field equations to be recoverable from generally covariant equations by means of a coordinate condition. See, for example, the remarks quoted in section 3.6.

physical field.[66] In so far as the field equations, after reduction by the coordinate condition, are intended to yield the Newtonian limit, they need only exhibit covariance under linear transformation. It might just happen that the reduced field equations exhibited greater covariance so that they might play a direct role in the generalization of the principle of relativity. But there is no reason to expect this. The only sure way to expand the covariance of the theory is to find the unreduced, generally covariant form of the gravitational field equations. That is the obvious and natural way to develop the *Entwurf* theory.

This was not Einstein's approach. Rather than seeking out these generally covariant equations, he let all his hopes hang on a slender thread: the *Entwurf* equation might just have sufficient covariance to support a generalized principle of relativity. So Einstein devoted his efforts to two tasks, both of which came to fruition after he had hit upon the hole argument. First he sought to discover the extent of the covariance of his *Entwurf* equations, describing this as the most important problem to be solved in the context of this theory.

> ...the equation of the gravitational field that we have set up do not possess this property [of general covariance]. For the equations of gravitation we have only been able to prove that they are covariant with respect to arbitrary *linear* transformations; but we do not know whether there exists a general group of transformations with respect to which the equation are covariant. The question as to the existence of such a group for the system of equations (18) and (21) [gravitational field equations] is the most important question connected with the considerations presented here. [Einstein's emphasis][67]

These efforts culminated in the discovery with Grossmann (Einstein and Grossmann 1914) that the covariance of his theory extends to what they call "adapted coordinate systems;" that is, coordinate systems that satisfy

$$\sum_{\alpha\beta\mu\nu} \frac{\partial^2}{\partial x_\nu \partial x_\alpha}\left(\sqrt{-g}\gamma_{\alpha\beta} g_{\sigma\mu} \frac{\partial \gamma_{\mu\nu}}{\partial x_\beta}\right) = 0. \tag{23}$$

Second he threw himself into the task of establishing that whatever limited covariance the *Entwurf* theory may have is good enough, for further covariance would be

66 For example, if our "field equation" is just a flatness requirement, the vanishing of Riemann-Christoffel curvature tensor, then one of its solutions is a Minkowski spacetime, whose coordinate representations include $g^{SR}_{\mu\nu}$ and $g^{ROT}_{\mu\nu}$. The effect of a coordinate condition such as (11) is not to eliminate a physical possibility such as this solution. It precludes the representation $g^{ROT}_{\mu\nu}$ with which is it incompatible; but it admits $g^{SR}_{\mu\nu}$.

67 "… daß die von uns aufgestellten Gleichungen des Gravitationsfeldes diese Eigenschaft nicht besitzen. Wir haben für die Gravitationsgleichungen nur beweisen könne, daß sie beliebigen linearen Transformationen gegenüber kovariant sind; wir wissen aber nicht, ob es eine allgemeine Transformationsgruppe gibt, der gegenüber die Gleichungen kovariant sind. Die Frage nach der Existenz einer derartigen Gruppe für das Gleichungssystem (18) bzw. (21) ist die wichtigste, welche sich an die hier gegebenen Ausführungen anknüpft." (Einstein and Grossmann 1913, I.§6; translation in CPAE 4E, 167) The continuation of the letter to Lorentz of 14 August 1913, quoted above describes some of his efforts to uncover these covariance properties.

physically uninteresting. Here Einstein had more success than his material warranted. He first showed in a trivially flawed and soon retracted argument that one can expect no more than linear covariance. Then the hole argument showed that generally covariance would be physically uninteresting and his analyses of 1914 showed that the *Entwurf* theory has the maximum covariance compatible with the hole argument.[68]

… is Explained

While Einstein's approach is baffling if we assume that he had a modern understanding of coordinate systems and coordinate conditions, it becomes entirely reasonable in the light of the conjecture of this part. He believed his *Entwurf* field equations to result from some set of unknown generally covariant equations reduced by a coordinate condition, presumably what turned out to be the adapted coordinate condition (23). In accord with (C1) and (C2), Einstein would pay no penalty in using the reduced form of the field equations in his theory. The covariance of the theory as a whole is just the covariance of the reduced equations (or, equivalently, the covariance of the coordinate condition (23). So the reduced form of these equations is not obscuring the true covariance of the theory as a whole, contrary to the modern view. And, since the effect of a coordinate condition (23) is just to restrict the covariance of the generally covariant equations, the reduced equations are not eliminating any physical fields; the limitation is just that each physical field arises in the theory in fewer coordinate representations. Thus, with the completion of the *Entwurf* theory in mid 1913, Einstein could have entered into the search for the generally covariant equations that correspond to his *Entwurf* equations. But there would have been little to gain from finding them. Finding them would not alter the covariance of the theory as a whole and it would not admit into the theory any new physical fields.[69]

There was a more pressing problem that had to absorb his immediate attention. Einstein did not know the covariance of the *Entwurf* theory. According to (C3), Einstein hoped that this covariance would extend to include transformations representing acceleration, for otherwise Einstein's hopes of extending the principle of relativity to acceleration would not be met by his theory. More was at stake. Einstein believed that his *Entwurf* gravitational field equation were unique; that is, they were the only equations employing a gravitation tensor of form (8) compatible with energy-momentum conservation.[70] Thus if the *Entwurf* equations failed to have sufficient covariance, then Einstein's entire project would be called into doubt. He could not just reject the

68 For discussion, see (Norton 1984, §6).

69 Or more simply, if Einstein was unaware of the use of coordinate conditions, the use of the generally covariant field equations, unsupplemented by adapted coordinate condition (23), would be incompatible with recovery of the Newtonian limit, since those equations would be unlikely to have the Newtonian form (8).

70 The uniqueness of these equations is suggested by the description of the identities (12) of Einstein and Grossmann (Einstein and Grossmann 1913, §5) used in the derivation of these equations as "uniquely determined" and then directly affirmed by Einstein (Einstein 1914b, 289). See (Norton 1984, §4) for discussion; the equations prove not to be unique, although this is not easy to see.

Entwurf field equations and seek a better alternative. He now believed that he had no option other than the *Entwurf* equations. Thus Einstein had to find the covariance of the *Entwurf* equations and, if his efforts to extend the principle of relativity were to succeed, it had to include acceleration transformations.

Thus the conjecture explains exactly the direction of Einstein's research on completion of the *Entwurf* theory.[71] He would gain nothing of significance from finding the generally covariant equations corresponding to his *Entwurf* equations. The problem urgently needing his attention was the discovery of the extent of the covariance of his *Entwurf* equations. These efforts of discovery soon transformed into the arguments that sought to established the need, in physical terms, for a restriction on covariance: that is, the arguments from the conservation laws and the hole argument. As Einstein's remarks from early 1914 quoted above indicate, these arguments establish that the quest for the generally covariant equations is of "no special interest"—a conclusion that I urge had already been forced implicitly by his according independent physical reality to the coordinate systems arising in the process of extracting the Newtonian limit.

3.6 Einstein's Pronouncements on the Selection of Specialized Coordinate Systems

The conjecture advanced here requires that Einstein's 1912–1915 understanding of coordinate systems in quite irregular. It is essential that this conjecture be compatible with Einstein's pronouncements on coordinate systems from this period. As it turns out, Einstein made few such pronouncements—so few, that it was initially thought in the history of science literature that Einstein was unaware of how to use four conditions to constrain the choice of coordinate systems. My purpose in this section is to review Einstein's most important pronouncements on the selection of specialized coordinate systems from this period and to show that they are quite compatible with the conjecture advanced here, although they neither speak for nor against it.

Two Ways to Introduce Specialized Coordinate Systems

Best known of these pronouncements is a distinction made in (Einstein 1914a, 177–178). Since this last pronouncement turns out to be a somewhat awkward statement of the same distinction explained more clearly in a later letter to Lorentz, I shall consider the later remarks first. In a letter of 23 January 1915 to Lorentz, Einstein sought to explain that his "choice of coordinates makes no assumption physically about the world." He used a "geometric comparison" to illustrate the possibilities:

> I have a surface of unknown kind upon which I want to carry out geometrical investigations. If I require that a coordinate system (p, q) on the surface can be so chosen that
>
> $$ds^2 = dp^2 + dq^2$$

71 A supposed lack of awareness of the use of coordinate conditions would also explain this direction.

then I thereby assume that the surface can be developed onto a plane. However if I require only that the coordinates can be so chosen that

$$ds^2 = A(p,q)dp^2 + B(p,q)dq^2$$

i.e. that the coordinates are orthogonal, I thereby assume nothing about the nature of the surface; one can realize them on any surface.[72]

Einstein's remark is a commonplace of differential geometry and applies equally in the geometry of two-dimensional surfaces and in the geometry of spacetimes. In presuming the existence of a particular coordinate system, we might be tacitly restricting the geometry of the space, or we might not. So, as in Einstein's first example, if we assume that there is a coordinate system in which the metrical coefficient $g_{\mu\nu}$ are constant, then we are assuming that the space is also metrically flat.[73] For constancy of the $g_{\mu\nu}$ is necessary and sufficient for metrical flatness. Other coordinate systems, however, can be realized in any space, so that the presumption of their existence does not restrict the geometric properties of the space.

To proceed to Einstein's (1914) remarks, we express the constraint that picks out a coordinate system in which the metrical coefficients are all constant as

$$\frac{\partial\gamma_{\alpha\beta}}{\partial x_\kappa} = 0. \tag{24}$$

This condition is equivalent to metrical flatness, which is a condition that can be given in invariant or generally covariant form, that is, as the vanishing of the Riemann-Christoffel curvature tensor

$$R^{\alpha}_{\beta\mu\nu} = 0. \tag{24'}$$

However the now familiar

$$\sum_\kappa \frac{\partial\gamma_{\kappa\alpha}}{\partial x_\kappa} = 0 \tag{11}$$

consumes just the four degrees of freedom available in selection of a coordinate system in any four-dimensional spacetime and thus places no restriction on its geometry. Whatever (11) states cannot be re-expressed by a non-vacuous invariant or generally covariant relation.

---

72 "Es liegt mir eine Fläche unbekannter Art vor, auf der ich geometrische Untersuchungen machen will. Verlange ich, es solle auf der Fläche ein Koordinatensystem (p,q) so gewählt werden, dass $ds^2 = dp^2 + dq^2$, [s]o setzte ich damit voraus, dass die Fläche auf eine Ebene abwickelbar sei. Verlange ich aber nur, dass die Koordinaten so gewählt seien, dass $ds^2 = A(p,q)dp^2 + B(p,q)dq^2$ ist, d. h. dass die Koordinaten orthogonal seien, so setzte ich damit über die Natur der Fläche nichts voraus; man kann dies auf jeder Fläche erzielen." Einstein to H. A. Lorentz, 23 January 1915 (CPAE 8, Doc. 47).

73 That a surface "can be developed onto a plane" is synonymous with flatness.

Working Backwards

This is the distinction that Einstein describes in (Einstein 1914a). The difference is that Einstein starts with an expression of restricted covariance and then works backwards, asking if the expression came from a generally covariant expression by restriction of the coordinate system.

> If we are given equations connecting any quantities whatsoever[74] that are valid only for a special choice of the coordinate system, then one has to distinguish between two cases:
>
> 1. To these equations there correspond generally covariant equations, i.e. equations valid with respect to arbitrary reference systems;
>
> 2. There are no generally covariant equations that can be deduced from the equations given for the specially chosen reference frame.
>
> In case 2, the equations say nothing about the things described by the quantities in question; they only restrict the choice of reference system. If the equations say anything at all about the things represented by the quantities, then we are dealing with case 1, i. e., in that case, there always exist generally covariant equations between the quantities.[75]

The constraint (24) is an instance of a non-generally covariant equation of case 1. Its existence does restrict the quantities involved, for it entails the flatness of the metric. Thus there is a corresponding generally covariant relation (24'). The requirement (11), however, generates no restriction on these quantities and thus corresponds to no (non-vacuous) generally covariant requirement.

The distinction outlined here does not bear on the reading of coordinate restrictions I urge Einstein held in 1912–1915. The requirement (11), places no restriction on the geometric properties represented by the metric $g_{\mu\nu}$. That is an issue independent of how the requirement picks out particular coordinate systems. To parrot Einstein, the requirement "says nothing" about the metrical quantities, but it certainly "says some-

74 Einstein's footnote: "Of course, the transformation properties of the quantities themselves must be considered here as being given for arbitrary transformations." ("Die Transformationseigenschaften der Größen selbst müssen natürlich hierbei als für beliebige Transformationen gegeben betrachtet werden.")

75 "Wenn Gleichungen zwischen irgendwelchen Größen gegeben1) [see previous note] sind, die nur bei spezieller Wahl des Koordinatensystems gültig sind, so sind zwei Fälle zu unterscheiden:
1. Es entsprechen den Gleichungen allgemein kovariante, d. h. bezüglich beliebiger Bezugssysteme gültige Gleichungen;
2. es gibt keine allgemein kovarianten Gleichungen, die aus den für spezielle Wahl des Bezugssystems gegebenen Gleichungen gefolgert werden können.
Im Falle 2 sagen die Gleichungen über die durch die Größen dargestellten Dinge gar nichts aus; sie beschränken nur die Wahl des Bezugssystems. Sagen die Gleichungen über die durch die Größen dargestellten Dinge überhaupt etwas aus, so liegt stets der Fall 1 vor, d. h. es existieren dann stets allgemein kovariante Gleichungen zwischen den Größen." (Einstein 1914a, 177–178; translation in CPAE 4E, 284) Einstein's purpose is to assert that his non-generally covariant gravitational field equations of the *Entwurf* theory do make some assertion about the quantities involved. Thus they are an instance of case 1. and there must exist corresponding generally covariant equations.

thing" about the coordinate systems, for it admits some and precludes others. Deciding just what it says about them is the issue that defeated Einstein in 1912–1915.

Specialized Coordinate Systems and Nordström's Theory of Gravitation

There is an important instance of case 1 in (Einstein and Fokker 1914), submitted for publication in February 1914, a month after the submission of (Einstein 1914a). Their work pertains to Nordström's latest theory of gravitation, which Einstein judged the most viable of the gravitation theories then in competition with the Einstein and Grossmann *Entwurf* theory.[76]

Nordström's theory had been developed by Nordström and Einstein as a Lorentz covariant theory of gravitation. With Fokker, Einstein now showed that the theory could be recovered in the generally covariant framework of the *Entwurf* theory, complete with its generally covariant energy conservation law. In place of the Einstein-Grossmann gravitational field equations, Einstein and Fokker adopted the single field equation $R = kT$, where R is the Riemann curvature scalar, T the trace of the stress-energy tensor and k a constant. That single equation would be insufficient to fix the ten coefficients of the metric tensor, so additional constraints were needed. "It turns out," Einstein and Fokker observed in their introductory summary, "that one arrives at the Nordström theory instead of the Einstein-Grossmann theory, if one makes the sole assumption that it is possible to choose preferred coordinate systems in such a way that the principle of the constancy of the speed of light obtains."[77] They interpreted the presumption of such a coordinate system as equivalent to assuming the existence of coordinate systems in which the spacetime's line element has the form[78]

$$ds^2 = \phi^2(dx_1^2 + dx_2^2 + dx_3^2 - dx_4^2). \tag{25}$$

That a spacetime admits a line element of this form greatly restricts its geometry; it is equivalent to conformal flatness. As Einstein suggests, this restriction can be written in generally covariant form. It was later found to be equivalent to the vanishing of the Weyl conformal tensor.

Einstein does not mention the Nordström theory in remarks in a letter to Planck 7 July 1914, written about six months after publication of Einstein and Fokker's paper. However his remarks describe exactly the specialized coordinate system introduced in the Einstein-Fokker formulation of the Nordström theory.

> There is a fundamental difference between that specialization of the reference system that classical mechanics or [special] relativity theory introduces and that which I apply in

76 For an account of Nordström's theories, see (Norton 1992; 1993).

77 "Es erweist sich hierbei, daß man zur Nordströmschen Theorie statt zur Einstein-Großmannschen gelangt, wenn man die einzige Annahme macht, es sei eine Wahl bevorzugter Bezugssysteme in solcher Weise möglich, daß das Prinzip von der Konstanz der Lichtgeschwindigkeit gewahrt ist." (Einstein and Fokker 1914, 321)

78 Then a light signal, for which $ds^2 = 0$, propagates with unit coordinate velocity. For example, if it propagates along the x_1 axis, the light signal satisfies $dx_1/dx_4 = \pm 1$.

> the theory of gravitation. That is, one can always introduce the latter, no matter how the $g_{\mu\nu}$ may be chosen. However the specialization introduced by the principle of the constancy of the speed of light presumes differential relations between the $g_{\mu\nu}$, and indeed relations whose physical interpretation would be very difficult. The satisfaction of these relations cannot be enforced for every given manifold through suitable choice of the reference system. According to the latter understanding, there are two heterogeneous conditions for the $g_{\mu\nu}$
>
> 1) the analog of Poisson's equation
>
> 2) the conditions that enable the introduction of a system of constant c.[79]

These two "heterogeneous conditions" correspond exactly with the two laws of the Nordström theory. The first, the field equation $R = kT$, is the analog of Poisson's equation. The second is the presumption that we can introduce a coordinate system in which the line element takes the conformally flat form (25). Its introduction is enabled by further conditions, which were later found to be expressible as "differential relations between the $g_{\mu\nu}$," the vanishing of the Weyl conformal tensor.

3.7 Was Einstein Really Defeated by According an Independent Reality to Coordinate Systems?

That is, is the conjecture of this part true? In sum, the answer is similar to the one given in section 2.4 to the question of whether Einstein was aware of coordinate conditions. There is no decisive piece of evidence for or against, but there are indications that point in both directions. Again, our ultimate assessment depends in some significant measure on issues of plausibility. My view is that the latter favor the conjecture.

The Notebook and the Entwurf Theory

If we accept that Einstein was aware of the use of coordinate conditions in the notebook and later, then we have several incongruities to explain. Why does he abandon their use so precipitously? Why does his later correspondence discount a perfectly serviceable extraction of the Newtonian limit from the candidate gravitation tensor T_{il}^{x}? Why is his discussion of the *Entwurf* theory, prior to his discovery of general arguments against general covariance so indifferent to the recovery of the generally

---

79 "Es gibt einen prinzipiellen Unterschied zwischen derjenigen Spezialisierung des Bezugssystems, welche die klassische Mechanik bezw. die Relativitätstheorie einführt und zwischen derjenigen, welche ich in der Gravitationstheorie anwende. Die letztere kann man nämlich stets einführen, wie auch die $g_{\mu\nu}$ gewählt werden mögen. Diese durch das Prinzip der Konstanz der Lichtgeschwindigkeit eingeführte Spezialisierung dagegen setzt Differenzialbeziehungen zwischen den $g_{\mu\nu}$ voraus, und zwar Beziehungen, deren physikalische Interpretation sehr schwierig sein dürfte. Das Erfülltsein dieser Beziehungen kann nicht für jede gegebene Mannigfaltigkeit durch passende Wahl des Bezugssystems erzwungen werden. Es gibt nach letzterer Auffassung zwei heterogene Bedingungen für die $g_{\mu\nu}$
1) das Analogon der Poisson'schen Gleichung
2) die Bedingungen, welche die Einführung eines Systems von konstantem c ermöglichen." Einstein to Max Planck, 7 July 1914 (CPAE 8, Doc. 18).

covariant gravitational field equations he allowed must exist? The conjecture of this part supplies an explanation that answers all of these questions.

Before we embrace that explanation, however, we should note that there is no direct evidence that Einstein did accord an independent reality to coordinate systems in the relevant context of the Newtonian limit. That is, we do not have unequivocal remarks by Einstein announcing it or a calculation whose only reasonable interpretation is that independence. It is hard to know how seriously to take this omission. Since Einstein was not using coordinate conditions to recover the Newtonian limit in his *Entwurf* theory, he had no occasion to undertake calculations that would unequivocally display an independent reality accorded his limit coordinate systems. What Einstein does give us are the manipulations of the hole argument. It is quite evident that he does there accord independent reality to the coordinate systems and his later admissions affirm this. Similarly, there were few occasions for Einstein to discuss how coordinate conditions could be used to recover the Newtonian limit, for this was not the construction he used in the *Entwurf* theory. On the few occasions in which he discussed general principles surrounding specialization of the coordinate system (see section 3.6 above), he makes no mention of an independent reality of the specialized coordinate systems. But then we would not expect him to. In section 3.1 we saw Einstein's difficulty in making explicit just how the manipulation of the hole argument depended on the independent reality of the coordinate system. If Einstein had such difficulty describing that independent reality when it was the essential point of the calculation, why should we expect him to express it clearer elsewhere?

Einstein's Later Discussion

Once Einstein had discovered his errors and returned to general covariance, he again had the opportunity to admit that he had accorded an independent reality to his coordinate systems. There were two prime occasions for such admission: his paper of 4 November 1915, and his letter to Sommerfeld of 28 November, in which he explained his rejection of the candidate gravitation tensor T_{il}^{x}. In both places, however, he emphasized the $\{\}$ prejudice as the source of his mistake. What is odd about both sources is that neither seek to explain the most public conceptual error of his *Entwurf* theory, the hole argument. At the time of the 4 November paper, Einstein had not yet discovered his misconception about static fields. As far as we know, the hole argument was the only foundational error of principle in the *Entwurf* theory, short of the ultimate mistake of choosing the *Entwurf* equations of restricted covariance. Since the error of the hole argument and the conjectured misuse of coordinate conditions are closely related, hesitancy in discussing the one should be expected to accompany hesitancy in discussing the other. And there was great hesitancy.

There are early published remarks that amount to the briefest retraction of the hole argument. But they offer little to explain the error of the argument. They appear in Einstein's celebrated computation of the anomalous motion of Mercury, in a paper presented to the Berlin Academy on 18 November 1915. There Einstein considers the gravitational field of a point mass at the origin of spatial coordinates, which he takes to be the sun. Solving for this case, even in lower order approximation, involves a system analogous to the hole of the hole argument. The field is constrained by Minkowskian boundary conditions at spatial infinity, just as the field in the hole is constrained by the surrounding matter distribution. In addition the field of the sun is constrained by the requirements that it be static and spatially symmetric about the origin. These additional requirements do not preclude all transformations; a spatial radial coordinate r could be arbitrarily transformed as long as the transformation does not disturb the limit at spatial infinity and preserves unit modulus by, say, corresponding adjustments elsewhere. Einstein remarked:

> We should however bear in mind that for a given solar mass the that the $g_{\mu\nu}$ are still not completely determined mathematically by the equations (1) and (3).[80] This follows since these equations are covariant with respect to arbitrary transformations of determinant 1. We may assume, however, that all these solutions can be reduced to one another through such transformations, so that they differ from one another only formally but not physically (for given boundary conditions). As a result of this conviction, I am satisfied for the present to derive a solution without being drawn into the question of whether it is the only possible [solution].[81]

If the covariance of the field equations is to block determination of the field in this case, it must be through the hole construction, so we have many solutions mathematically in the one coordinate system. Einstein parries the threat by observing that these solutions "differ from one another only formally but not physically" and the same remark would serve as an escape from the hole argument. Only a quite attentive reader would see the connection and even then such a reader may well find the remark unconvincing. Certainly Ehrenfest needed a more elaborate account of the failure of the hole argument before he was satisfied.[82] Yet Einstein concluded by explicitly disavowing any further discussion This neglect is striking in comparison to the careful self diagnosis elaborated as the $\{\}$ prejudice.

Why might he be reluctant to discuss the error of the hole argument? He may just have been reluctant to relive a painful experience, especially if he saw no benefit from it. Or perhaps he had some difficulty formulating precisely what the error was, even after he knew of it. It was sufficient that he knew that the hole construction did not produce physically distinct fields. If he had suffered this difficulty it would explain why he delayed detailed discussion of the error of the hole argument for nearly two

80 Einstein's equations are $T^x_{il} = 0$ and $|g_{\mu\nu}| = 1$, which are covariant under unimodular transformations.

81 "Es ist indessen wohl zu bedenken, daß die $g_{\mu\nu}$ bei gegebener Sonnenmasse durch die Gleichungen (1) und (3) mathematisch noch nicht vollständig bestimmt sind. Es folgt dies daraus, daß diese Gleichungen bezüglich beliebiger Transformationen mit der Determinate 1 kovariant sind. Es dürfte indessen berechtigt sein, vorauszusetzen, daß alle diese Lösungen durch solche Transformationen aufeinander reduziert werden können, daß sie sich also (bei gegebenen Grenzbedingungen) nur formell, nicht aber physikalisch voneinander unterscheiden. Dieser Überzeugung folgend begnüge ich mich vorerst damit, hier eine Lösung abzuleiten, ohne mich auf die Frage einzulassen, ob es die einzig mögliche sei." (Einstein 1915b, 832)

82 See section 3.2 and (Norton 1987, §4).

months after his public announcement of his return to general covariance. As far as we know from documents available to us, the first detailed discussion comes in his letter of 26 December 1915, to Ehrenfest (see section 3.2).

Whatever may have underpinned his reluctance to discuss the error of the hole argument, the same reason would surely induce a similar reluctance to discuss the closely related error conjectured here.

Einstein's Letter to de Sitter

According to the conjecture of this part, there is a close connection between two of Einstein's errors: the notebook rejection of the candidate gravitation tensor T_{il}^{x} and the hole argument. We would hope to see some trace of that connection. Such a trace may appear in a letter Einstein wrote to de Sitter on 23 January 1917.

To see how this letter can be interpreted, we must recall Einstein's return to general covariance in the fall of 1915. In several places, Einstein listed the clues that forced him to accept the inadequacy of his *Entwurf* theory.[83] In particular, Einstein had erroneously convinced himself that the *Entwurf* theory was covariant under rotation transformation (18).[84] The discovery of this error cast Einstein into despair over his theory, as he confided to his astronomer colleague Erwin Freundlich in a letter of 30 September 1915 (CPAE 8, Doc 123). In it, he was reduced to a despondent plea for help. He was not frozen into inactivity, however. A little over a month later, on 4 November, he announced his return to general covariance and the adoption of T_{il}^{x} as his gravitation tensor.

That one discovery of the lack of rotational covariance of the *Entwurf* theory seems to have been a powerful stimulus. Two things followed rapidly after it. He returned to general covariance (and therefore rejected the hole argument) and he readmitted the gravitation tensor T_{il}^{x} as gravitation tensor. If the original rejection of T_{il}^{x} had been due to improperly according independent reality to coordinate systems, then we may readily conceive natural scenarios that connect the two. For example, lack of rotational covariance would be fatal to Einstein's hopes of generalizing the principle of relativity to acceleration. So if he now realized that his *Entwurf* theory could not supply it, he might well return to the last candidate gravitation tensors considered in the context of the rotation transformation (18). That would be T_{il}^{x} and the related proposals around page 22 of the notebook. Now wiser and desperate and suspicious of all his methods and presumptions, Einstein might just finally be able to see past his objection to the coordinate condition (11) to the recognition that there was something improper in the core of his objection, his interpretation of what I have called the hole construction. That realization would have simultaneously allowed him to see that the hole argument does not succeed in showing the inadmissibility of gen-

83 See (Norton 1984, §7).

84 Janssen (1999) supplies a fascinating chronicle of this episode. It includes display of calculations in Einstein's hand apparently from June 1913 in which Einstein erroneously affirms that $g_{\mu\nu}^{ROT}$ is a solution of the *Entwurf* gravitational field equations and then a repetition of the same calculation probably from late September 1915 in which Einstein finds the error.

erally covariant gravitational field equations. For it also depends on the same interpretation of the hole construction. Because of the close connection between the two errors, some such scenario among many obvious variants is credible.

As we saw in section 3.2, Einstein gave several accounts of the error of the hole argument. None mentioned above contain autobiographical remarks on how Einstein found the error. There is one exception, a recollection in a letter of 23 January 1917, to de Sitter concerning the errors of Einstein (Einstein 1914c)

> ...there were the following two errors of reasoning [in (Einstein 1914c)]:
>
> 1) The consideration of §12 [the hole argument] is incorrect, since occurrences can be uniquely determined without the same being true for the functions used for their description.
>
> 2) In §14 at the top of page 1073 is a defective consideration.
>
> I noticed my mistakes from that time when I calculated directly that my field equations of that time were *not* satisfied in a rotating system in a Galilean space. Hilbert also found the second error.[85]

Here Einstein assures us that he found the errors of his 1914 review article, with the hole argument listed as the first of the two errors, because he discovered the lack of rotational covariance of his *Entwurf* field equations.[86] Without the conjecture of this part, it is hard to see why Einstein would proceed without great detours from that lack of rotational covariance to the rejection of the hole argument.

What is More Plausible?

In the absence of decisive evidence, we once again ask after the plausibility of the conjecture. To my mind, the one factor that speaks against the conjecture is this very lack of evidence. Things might have transpired as conjectured without more decisive evidence surviving. Einstein was not obligated to annotate his private calculations or later recount every misstep, so as to save the labor of future historians. The resulting paucity of evidence, however, is also compatible with a simpler explanation: things just did not go as conjectured. One factor makes this case a little different from the earlier deliberations on Einstein's supposed unawareness of the use of coordinate conditions: the conjecture ties Einstein's misturnings to the error of the hole argu-

85 "… folgende zwei Denkfehler waren darin:
1) Die Betrachtung des §12 ist unrichtig, weil das Geschehen eindeutig bestimmt sein kann, ohne dass die zu seiner Beschreibung dienenden Funktionen es sind.
2) In §14 ist oben auf Seite 1073 eine fehlerhafte Überlegung.
Ich merkte meine damaligen Irrtümer daran, dass ich direkt ausrechnete, dass meine damaligen Feldgleichungen für ein in einem Galileischen Raume rotierendes System *nicht* erfüllt waren. Den zweiten Fehler hat auch Hilbert gefunden." (CPAE 8, Doc. 290)

86 By a "Galilean space," Einstein refers to a Minkowski spacetime in the coordinates of (5). The second error is presumably the one Einstein discusses with Hilbert in a letter of 30 March 1916, to Hilbert (CPAE 8, Doc. 207) and concerns the failure of a variation operator to commute with coordinate differentiation. For discussion, see (Norton 1984, end of §6).

ment. In that case we have no doubt of Einstein's reticence to leave later traces of his error and that reticence would carry over to the related rejection of the tensor T^x_{il}. But now we tread on dangerous ground. We offer an account that also predicts that it will be difficult to find evidence for that account. Such accounts can be correct. They can also be a signal that a defective account has been protected illegitimately from refutation. There are earnest accounts of how our small planet is routinely visited by aliens intent on abductions. They face a sustained lack of concrete evidence. So we are assured that no irrefutable evidence of the visits survives because of a massive government conspiracy or the ingenuity and thoroughness of the aliens in eradicating all such traces!

These serious hesitations should be weighed against the need for some account of Einstein's twisted path. Again we risk a pitfall. If we are willing to multiply the errors Einstein is supposed to have committed, there is scarcely any pathway that we could not explain. What is appealing about the conjecture is that it requires us to posit no new errors. Aside from outright blunders of calculation and self deception, as documented in (Janssen 1999), Einstein was led astray for nearly three years by two groups of misconceptions. The first surrounded his presumptions on the form of the static metric and the weak field equations. The second pertained to the hole argument and the independent reality of the coordinate systems.

To arrive at the second, we need only ask that Einstein was consistent and thorough in his support of the misconception the hole argument. Then just one error leads Einstein to reject the use of coordinate conditions, to acquiesce to the gravely restricted covariance of the *Entwurf* theory and not to pursue its generally covariant generalization. The recognition of that same error both frees Einstein from the hole argument late in 1915 and allows him to propose T^x_{il} as his gravitation tensor.[87] If I must choose an account, I find this one plausible.

CONCLUSION

Why did Einstein reject the candidate gravitation tensor T^x_{il} in the notebook? His own answer emphasized his "fateful prejudice," the $\{\}$ prejudice. He did not see that the Christoffel symbols are the natural expression for the components of the gravitational field. As a result he could not properly relate the gravitation tensor to the requirement of energy conservation. Instead he was tempted to multiply out the Christoffel symbols to recover expressions explicitly in the metric tensor that would prove unwieldy.

That may well have been all that it took to convince Einstein to abandon the proposal. We must then discount as unrelated his anomalous concern with questions of

87 For comparison, consider the alternative account in which Einstein is just unaware of the use of coordinate conditions. This awareness must come if T^x_{il} is to be admissible as a gravitation tensor. So the preparation for the new proposals of November 1915 must include recognition of two independent errors, that of the hole argument and the neglect of coordinate conditions.

covariance on the pages surrounding page 22R on which the gravitation tensor is analyzed. While Einstein had clearly mastered the mathematical manipulations needed to apply a coordinate condition to expressions of general or near general covariance, his treatment of them suggests that his interpretation of the conditions was idiosyncratic. His concern for their covariance properties cannot be reconciled with his later attitude to them. So we have presumed that his treatment and interpretation of these coordinate condition supplied a further fateful prejudice that precluded admission of the candidate gravitation tensor T_{il}^{x} by somehow obstructing his extraction of the Newtonian limit. The supposition of this additional fateful prejudice makes Einstein appear far less capricious. In finding the gravitation tensor T_{il}^{x} he had circumvented the tangled cluster of problems he had imagined facing the Ricci tensor as gravitation tensor. We suppose that he abandoned the new proposal not just because the calculation looked complicated but because deeper matters of principle also seemed to speak against it.

Just how did Einstein's treatment of coordinate conditions defeat him? There is clear evidence in the notebook that Einstein used the requirements as what we call "coordinate restrictions": they are not just applied in the case of the Newtonian limit but universally. That alone does not explain why Einstein would think his candidate gravitation tensor unable to yield the Newtonian limit in a satisfactory manner. We have found two additional hypotheses that would supply the explanation. The first supposes an obtuse Einstein, overlooking a natural option. It supposes he just persistently failed to see that coordinate conditions could be invoked selectively as part of the restriction on covariance imposed in recovery of the Newtonian limit. The second, which I favor, portrays an excessively acute Einstein, zealously consistent even in his errors. He would soon improperly accord an independent reality to coordinate systems in the hole argument and the conjecture is that he did the same thing earlier in applying coordinate conditions. Both hypotheses have the same outcome. Einstein would come to an impossible demand: the requirement that reduces the candidate gravitation tensor to a Newtonian form must have sufficient covariance to support a generalization of the principle of relativity to acceleration. The first is a dim Einstein, felled by overlooking a standard device in general relativity that he later used without apology. The second is an Einstein of Byzantine sophistication, pursuing his errors, even when only dimly aware of them, to their farthest catastrophe. Perhaps another Einstein, the real Einstein, neither dim nor Byzantine, still waits to be discovered.

REFERENCES

CPAE 4: Martin J. Klein, A. J. Kox, Jürgen Renn, and Robert Schulmann (eds.), *The Collected Papers of Albert Einstein*. Vol. 4. *The Swiss Years: Writings, 1912–1914*. Princeton: Princeton University Press, 1995.

CPAE 4E: *The Collected Papers of Albert Einstein*. Vol. 4. *The Swiss Years: Writings, 1912–1914*. English edition translated by Anna Beck. Princeton: Princeton University Press, 1996.

CPAE 5: Martin J. Klein, A. J. Kox, and Robert Schulmann (eds.), *The Collected Papers of Albert Einstein*. Vol. 5. *The Swiss Years: Correspondence, 1902–1914*. Princeton: Princeton University Press, 1993.

CPAE 8: Robert Schulmann, A. J. Kox, Michel Janssen, and József Illy (eds.), *The Collected Papers of Albert Einstein*. Vol. 8. *The Berlin Years: Correspondence, 1914–1918*. Princeton: Princeton University Press, 1998. (2 Vols., Part A: 1914–1917, pp. 1–590; Part B: 1918, pp. 591–1118.)

Einstein, Albert. 1913. "Zum gegenwärtigen Stande des Gravitationsproblems." *Physikalische Zeitschrift* 14: 1249–1262, (CPAE 4, Doc. 17). (English translation in vol. 3 of this series.)

———. 1914a. "Prinzipielles zur verallgemeinerten Relativitätstheorie." *Physikalische Zeitschrift* 15: 176–180, (CPAE 4, Doc. 25).

———. 1914b. "Physikalische Grundlagen einer Gravitationstheorie." *Naturforschende Gesellschaft in Zürich. Vierteljahrsschrift* 58: 284–290, (CPAE 4, Doc. 16).

———. 1914c. "Die formale Grundlage der allgemeinen Relativitätstheorie." *Königlich Preussische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte* 1030–1085, (CPAE 6, Doc. 9).

———. 1915a. "Zur allgemeinen Relativitätstheorie," *Königlich Preussische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte* 778–786, (CPAE 6, Doc. 21).

———. 1915b. "Erklärung der Perihelbewegung des Merkur aus der allgemeinen Relativitätstheorie." *Königlich Preussische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte* 831–839, (CPAE 6, Doc. 24).

———. 1916. "Die Grundlage der allgemeinen Relativitätstheorie," *Annalen der Physik* 49: 769–822, (CPAE 6, Doc. 30).

———. 1917. *Relativity: the Special and the General Theory.* Translated by R. W. Lawson. London: Methuen, [1977].

———. 1922. *The Meaning of Relativity.* London: Methuen. [Fifth ed. Princeton: Princeton University Press, 1974.]

Einstein, Albert and Adriaan D Fokker. 1914. "Die Nordströmsche Gravitationstheorie vom Standpunkt des absoluten Differentialkalküls." *Annalen der Physik* 44: 321–328, (CPAE 4, Doc. 28).

Einstein, Albert and Marcel Grossmann. 1913. *Entwurf einer verallgemeinerten Relativitätstheorie und einer Theorie der Gravitation*. Leipzig: B.G.Teubner (separatum); with addendum by Einstein in *Zeitschrift für Mathematik und Physik* (1914) 63: 225–261, (CPAE 4, Docs. 13, 26).

———. 1914. "Kovarianzeigenschaften der Feldgleichungen der auf die verallgemeinerte Relativitätstheorie gegründeten Gravitationstheorie." *Zeitschrift für Mathematik und Physik* 63: 215–225, (CPAE 6, Doc. 2).

Fock, Vladimir. 1959. *The Theory of Space Time and Gravitation*. Translated by N. Kemmer. London: Pergamon.

Havas, Peter. 1964. "Four-dimensional Formulations of Newtonian Mechanics and their Relation to the Special and General Theory of Relativity." *Reviews of Modern Physics* 36: 938–965.

Howard, Don and John D. Norton. 1993. "Out of the Labyrinth? Einstein, Hertz and the Göttingen Answer to the Hole Argument." In J. Earman, M. Janssen, J. D. Norton (eds.), *The Attraction of Gravitation: New Studies in History of General Relativity.* Boston: Birkhäuser, 30–62.

Janssen, Michel. 1999."Rotation as the Nemesis of Einstein's 'Entwurf' Theory." In H. Goenner, J. Renn and T. Sauer (eds.), *The Expanding Worlds of General Relativity* (*Einstein Studies,* Vol. 7). Boston: Birkhäuser, 127–157.

Norton, John D. 1984 (1989). "How Einstein found his Field Equations: 1912–1915." *Historical Studies in the Physical Sciences* 14: 253–316. Reprinted 1989 in D. Howard and J. Stachel (eds.), *Einstein and the History of General Relativity* (*Einstein Studies,* Vol. 1). Boston: Birkhäuser, 101–159.

———. 1985. "What was Einstein's Principle of Equivalence?" *Studies in History and Philosophy of Science* 16: 203–246. Reprinted 1989 in D. Howard and J. Stachel (eds.), *Einstein and the History of General Relativity* (*Einstein Studies,* Vol. 1). Boston: Birkhäuser, 5–47.

———. 1987. "Einstein, the Hole Argument and the Reality of Space." In J. Forge (ed.), *Measurement, Realism and Objectivity.* Reidel, 153–188.

———. 1992. "Einstein, Nordström and the Early Demise of Scalar, Lorentz-Covariant Theories of Gravitation." *Archive for the History of Exact Sciences* 45: 17–94.

———. 1993. "Einstein and Nordström: Some Lesser-Known Thought Experiments in Gravitation." In J. Earman, M. Janssen and J. D. Norton (eds.), *The Attraction of Gravitation: New Studies in the History of General Relativity* (*Einstein Studies,* Vol. 5). Boston: Birkhäuser, 3–29.

Stachel, John. 1980. "Einstein's Search for General Covariance." A paper read at the Ninth International Conference on General Relativity and Gravitation, Jena. Printed 1989 in D. Howard and J. Stachel (eds.), *Einstein and the History of General Relativity* (*Einstein Studies,* Vol. 1). Boston: Birkhäuser, 63–100.

MICHEL JANSSEN

WHAT DID EINSTEIN KNOW AND WHEN DID HE KNOW IT? A BESSO MEMO DATED AUGUST 1913

1. THE CHALLENGE OF THE BESSO MEMO

The *Nachlass* of Einstein's close friend and confidant Michele Besso (1873–1955) contains four pages, written on a folded sheet, with what appear to be Besso's notes of discussions with Einstein about a preliminary version of general relativity known in the historical literature as the "*Entwurf*" ("outline") theory.[1] The first two pages of this Besso memo are reproduced in facsimile in Figs. 1 and 2.[2] Of the various points recorded in the memo two in particular are bound to catch the eye of a modern historian of relativity.

First, under point "b) 2." on the first page, Besso writes:

> If through rotation of a hollow sphere one produces a Coriolis field inside of it, then a centrifugal field is produced [...] that is not the same as the one that would occur in a rotating rigid system with the same Coriolis field. One can therefore not think of rotational forces as produced by the rotation of the fixed stars …[3]

The basic result that Besso is alluding to in the first sentence (as will be explained in detail in sec. 3) is that the 'rotation metric'[4] is not a vacuum solution of the *Entwurf* field equations, the field equations of severely restricted covariance of the *Entwurf* theory. And, as Besso indicates in the second sentence, if the rotation metric is not a solution, then the theory fails to relativize rotation along Machian lines.

Second, Einstein and Besso seem to have been on the brink of resolving what is essentially the well-known "hole argument."[5] This argument was first mentioned in

---

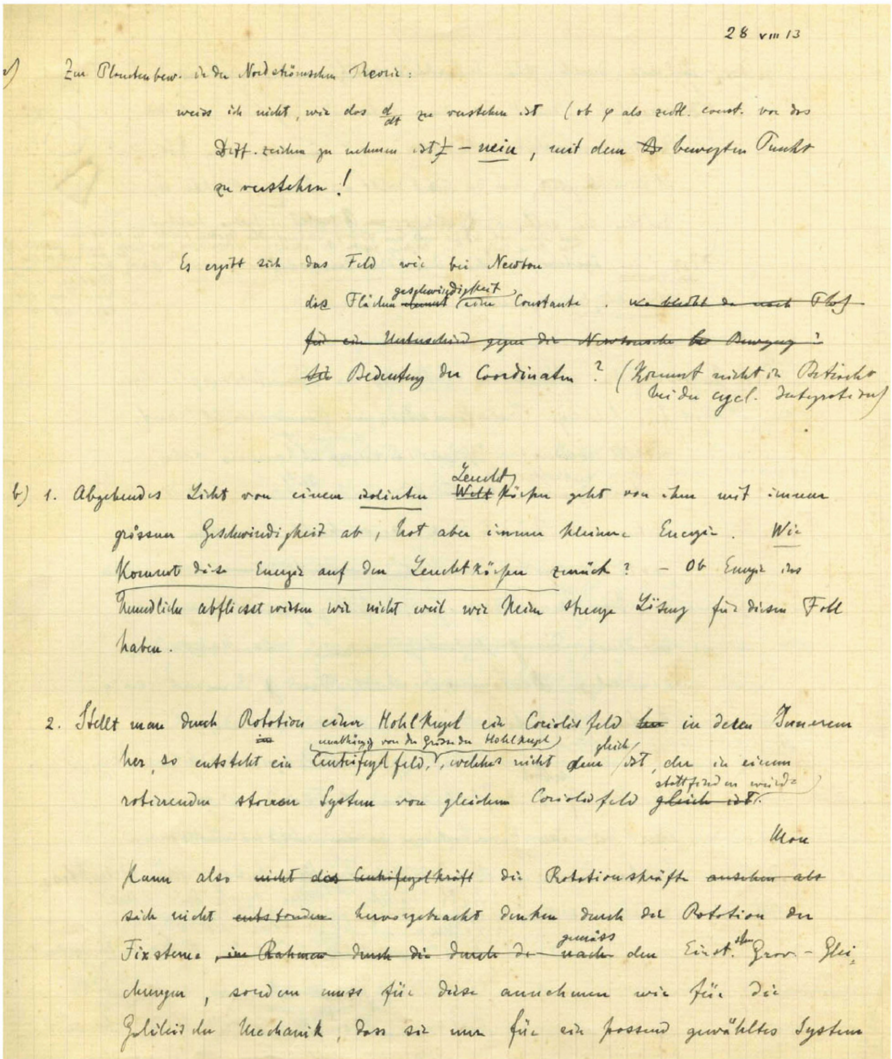
1 The name derives from the title of (Einstein and Grossmann 1913), the paper in which Einstein and his former classmate Marcel Grossmann first presented a theory of gravitation based on the metric tensor.

2 For a facsimile reproduction of all four pages of the Besso memo, see (Renn 2005, 127–130).

3 "Stellt man durch Rotation einer Hohlkugel ein Coriolisfeld in deren Innerem her, so entsteht ein Centrifugalfeld [...] welches nicht dem gleich ist, der in einem rotierenden starren System von gleichem Coriolisfeld statt finden würde. Man kann also die Rotationskräfte sich nicht hervorgebracht denken durch die Rotation der Fixsterne …"

4 I use this phrase as short-hand for the metric describing the geometry of Minkowski spacetime viewed from a uniformly rotating Cartesian coordinate system. See note 67 below for the explicit form of this metric.

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 2
Einstein's Zurich Notebook: Commentary and Essays.


28 VIII 13

a) Zur Planetenbew. in der Nordströmschen Theorie:

weiss ich nicht, wie das $\frac{d}{dt}$ zu verstehen ist (ob φ als zeitl. const. vor das Diff.zeichen zu nehmen ist?) — nein, mit dem bewegten Punkt zu verstehen!

Es ergibt sich das Feld wie bei Newton

die Flächengeschwindigkeit eine Constante. ~~Wo bleibt da noch Platz für ein Unterschied gegen die Newtonsche Bewegung?~~

Bedeutung der Coordinaten? (Kommt nicht in Betracht bei den cycl. Integrationen)

b) 1. Abgehendes Licht von einem isolierten Leuchtkörper geht von ihm mit immer grösserer Geschwindigkeit ab, hat aber immer kleinere Energie. Wie kommt diese Energie auf den Leuchtkörper zurück? — Ob Energie ins Unendliche abfliessen würde wissen wir nicht weil wir keine strenge Lösung für diesen Fall haben.

2. Stellt man durch Rotation einer Hohlkugel ein Coriolisfeld in deren Innerem her, so entsteht ein Centrifugalfeld (unabhängig von der Grösse der Hohlkugel), welches nicht gleich dem ist, der in einem rotierenden starren System von gleichem Coriolisfeld stattfinden würde.

Man kann also die Rotationskräfte nicht hervorgebracht denken durch die Rotation der Fixsterne gemäss den Einst.schen Grav.-Gleichungen, sondern muss für diese annehmen wie für die Gleichungen der Mechanik, dass sie nur für ein passend gewähltes System

Figure 1: Besso Memo, p. 1.
Reproduced with permission of the Besso Family Trust, Geneva, Switzerland

5 The classic historical discussions of Einstein's hole argument are (Stachel 1989, secs. 3–4), and (Norton 1984, sec. 5).

correspondence in November 1913 and first published in January 1914.[6] The argument was supposed to show that field equations for the metric field cannot be generally covariant, if the metric field is to be uniquely determined by its sources. At the bottom of the second page of the memo, Besso offers an escape from the hole argument:

> It is, however, not necessary that the [components of the metric] g themselves are determined uniquely, only the observable phenomena in the gravitation space, e.g., the motion of a material point, must be.[7]

This escape is then rejected in a comment appended to this passage.

Without any further information, I would not have hesitated dating Besso's memo to the fall of 1915. The two points singled out above strongly suggest that it belongs to the period between late September, when Einstein discovered the problem of rotation in his *Entwurf* theory,[8] and late October, when Einstein replaced the *Entwurf* field equations by equations of much broader covariance than the hole argument would seem to allow.[9]

Dating the memo to October 1915, however, is completely at odds with Besso's own dating of the document. In the top-right corner of the first page of the memo, Besso wrote: "28 VIII 13." With the apparent exception of the two passages quoted above, the available evidence supports Besso's own dating, as I will show in detail in sec. 2. I can understand the temptation to chalk it up to a slip on Besso's part. Acceptance of Besso's date poses some serious challenges for the reconstruction of Einstein's path to general relativity. With regard to both the problem of rotation and the hole argument the Besso memo raises the troubling question usually associated with politicians rather than scientists: what did he know and when did he know it?

Consider some of the other evidence we have pertaining to the problem of rotation. Writing to Freundlich in September 1915, Einstein made it sound as if he had just discovered that the rotation metric is not a vacuum solution of the *Entwurf* field equations. And reflecting on the tumultuous events of November 1915 in oft-quoted letters written shortly afterwards, he also talked about the problem of rotation as if it were a recent discovery that had precipitated the demise of the *Entwurf* theory.[10] Until late September 1915, it seems, Einstein had been under the impression that the rotation metric *is* a vacuum solution of the *Entwurf* field equations. The 1913 portion of the so-called Einstein-Besso manuscript contains a calculation, erroneous as it turns out, with which Einstein explicitly confirmed that this is the case.[11] Further

---

6 See Einstein to Ludwig Hopf, 2 November 1913 (CPAE 5, Doc. 480) and the "Comments" ("Bemerkungen") added to the journal version of the *Entwurf* paper (Einstein and Grossmann 1914a).

7 "Es ist allerdings nicht nötig, dass die g selbst eindeutig bestimmt sind, sondern nur die im Gravitationsraum beobachtbaren Erscheinungen, z.B. die Bewegung des materiellen Punktes, müssen es sein."

8 Einstein first reported the problem in a letter to Erwin Freundlich of 30 September 1915 (CPAE 8, Doc. 123).

9 Einstein published these equations in (Einstein 1915a), submitted to the Prussian Academy of Sciences on 4 November 1915.

10 See Einstein to Arnold Sommerfeld, 28 November 1915 (CPAE 8, Doc. 153) and Einstein to H. A. Lorentz, 1 January 1916 (CPAE 8, Doc. 177).

elaboration of the *Entwurf* theory in 1914 (and possibly another erroneous calculation) reinforced this belief.[12] The Machian account of rotation based in part on this (specious) result is hailed as one of the great triumphs of the *Entwurf* theory in the introduction of the lengthy exposition of the theory published in November 1914 (Einstein 1914e, 1031–1032). But if we accept the date of the Besso memo, Einstein already knew in August 1913 that the rotation metric is *not* a vacuum solution of the *Entwurf* field equations! How can this be reconciled with the other available evidence?

Similar questions can be raised about the hole argument. If the Besso memo is indeed from August 1913, the argument discussed in these pages is the earliest extant version of the hole argument. As we saw, Besso explicitly recognized that the field equations need not determine the metric field uniquely, only such things as particle trajectories. It may seem but a small step from particle trajectories to *intersections of* particle trajectories. With this modification we arrive at the core of the so-called "point-coincidence argument" with which Einstein explained the failure of the hole argument in correspondence of late 1915 and early 1916.[13] But if Einstein and Besso came this close to the resolution of the hole argument in August 1913, why did Einstein proceed to publish the argument no less than four times in 1914?

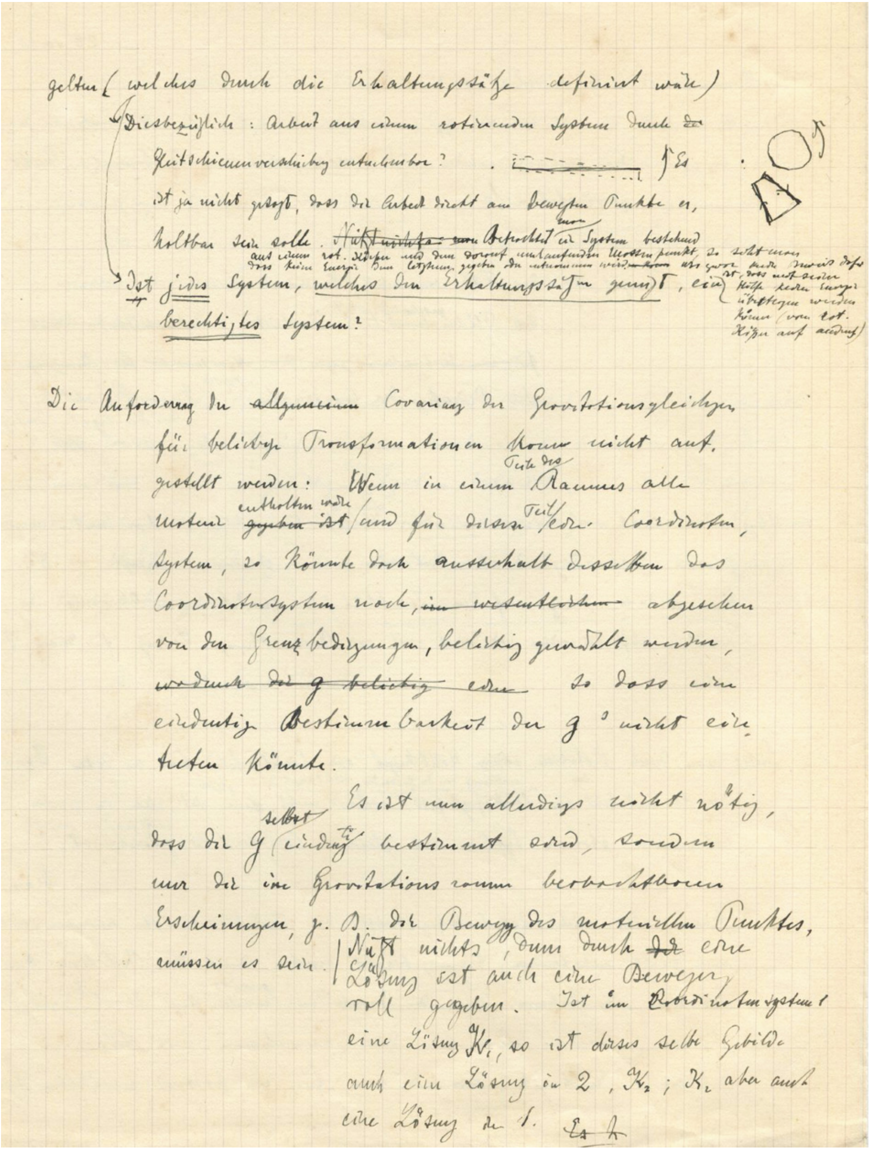
After making the case for accepting the date on the Besso memo in sec. 2, I try to answer these questions, for the problem of rotation in sec. 3, for the hole argument in sec. 4. I argue that, despite appearances to the contrary, there is a plausible reconstruction of events that can accommodate the Besso memo quite naturally. This reconstruction, however, does require that we invoke a certain element of opportunism—or expediency, to use a somewhat more neutral term—in Einstein's *modus operandi*. By this I mean that Einstein had a tendency to believe the results he wanted to believe and that he was quite adept at cooking up arguments to suit his needs, arguments he did not necessarily scrutinize all that closely. I want to make it clear right away that I do not use the term "opportunistic" in its pejorative sense. On the contrary, I strongly suspect that the way in which Einstein responded to the problems he ran into with his *Entwurf* theory is quite typical of creative work in theoretical physics and probably in other sciences as well. The reason I want to emphasize Einstein's opportunistic streak is that it has somehow been lost in the dominant tradition in recent Einstein studies, exemplified by the work of John Stachel and John Norton. To a very large extent my paper is squarely within this tradition, but I see it as an important weakness that, at least so far, it has failed to take into account this aspect of Einstein's general *MO*.

---

11 See (CPAE 4, Doc. 14, [pp. 41–42]) for the calculation, and (Janssen 1999) for a detailed analysis.

12 See Einstein to Michele Besso, ca. 10 March 1914 (CPAE 5, Doc. 514); Einstein to Joseph Petzoldt, 16 April 1914 (CPAE 8, Doc. 5); Einstein to Wilhelm Wien, 15 June 1914 (CPAE 8, Doc. 14).

13 See Einstein to Paul Ehrenfest, 26 December 1915 and 5 January 1916 (CPAE 8, Docs. 173 and 180) and Einstein to Michele Besso, 3 January 1916 (CPAE 8, Doc. 178). The argument was published in (Einstein 1916, 776–77). For historical discussion of the point-coincidence argument, see (Stachel 1989, sec. 6; Norton 1987; Kox 1989; Howard and Norton 1993; and Howard 1999).



gelten (welches durch die Erhaltungssätze definiert wäre)

Diesbezüglich: Arbeit aus einem rotierenden System durch die Zeitschiemenverstellung entnehmbar? Es ist ja nicht gesagt, dass die Arbeit direkt am bewegten Punkte ersichtbar sein solle. ~~Nützt nichts~~ Betrachtet man ein System bestehend aus einem rot. Körper und dem darauf umlaufenden Massenpunkt, so sieht man dass keine Energie dem letzteren gegeben oder entnommen werden kann [illegible] ist, dass mit dessen Hilfe keine Energie übertragen werden kann (vom rot. Körper auf andere)

Ist jedes System, welches den Erhaltungssätzen genügt, ein berechtigtes System?

Die Anforderung der ~~allgemeinen~~ Covarianz der Gravitationsgleichungen für beliebige Transformationen kann nicht aufgestellt werden: Wenn in einem Teile des Raumes alle Materie ~~gegeben ist~~ enthalten wäre / und für diesen Teil / ein Coordinaten-System, so könnte doch ausserhalb desselben das Coordinatensystem noch, ~~im wesentlichen~~ abgesehen von den Grenzbedingungen, beliebig gewählt werden, ~~wodurch die g beliebig ein~~ so dass eine eindeutige Bestimmbarkeit der g's nicht eintreten könnte.

Es ist nun allerdings nicht nötig, dass die g selbst (eindeutig) bestimmt sind, sondern nur die im Gravitationsraum beobachtbaren Erscheinungen, z. B. die Bewegung des materiellen Punktes, müssen es sein. | Nützt nichts, denn durch ~~die~~ eine Lösung ist auch eine Bewegung voll gegeben. Ist im Koordinatensystem 1 eine Lösung K_1, so ist dieses selbe Gebilde auch eine Lösung in 2, K_2; K_2 aber auch eine Lösung in 1. ~~Es h~~

Figure 2: Besso Memo, p. 2.
Reproduced with permission of the Besso Family Trust, Geneva, Switzerland

2. CAN BESSO'S DATING OF HIS MEMO BE TRUSTED?

2.1 Survey of the New Besso Material Containing the Besso Memo

The Besso memo is part of an assortment of manuscript pages that Laurent Besso showed to Robert Schulmann, then director of the Einstein Papers Project at Boston University, in Lausanne in 1998. The material consists of loose sheets of paper of various sorts and sizes carrying notes and calculations, all in Besso's hand. General relativity is the topic of most—but not all—pages. Leaving aside the Besso memo for a moment, one can divide the pages on general relativity into two groups of about ten pages each, one from (late) 1913 and possibly 1914, and one from 1916. The Besso memo, or so I will argue, belongs to the first group. But let me say a few words about the second group first.

2.1.1 The Second Group of Pages (1916)

There is strong evidence that the second group of new Besso pages related to general relativity is from 1916. There are several references to Einstein's review article on general relativity published in March 1916 (Einstein 1916) and no references to anything published later. The centerpiece of this group of pages is a partial draft of an unpublished essay entitled "The Relativity Principle in an Epistemological Formulation" ("Das Relativitätsprinzip in erkenntnistheoretischer Formulierung"). The remaining pages contain notes, mainly on (Einstein 1916), which Besso probably made while preparing this essay. The essay in turn is probably related to a lecture on Einstein's new theory that Besso was asked to give for the *Physikalische Gesellschaft* in Zurich.[14] This is suggested by the following remark jotted down at the top of the fourth page of the essay:

> Orechiante—whistling Beethoven symphony—the three professional musicians should straighten me out[15]

Besso used the exact same imagery in the draft of a letter to Einstein of late June 1916. The relevant passage explains what Besso meant by the cryptic remark above. At the same time, Besso's self-deprecating comment nicely illustrates how he perceived his relationship with his friend Einstein when it came to matters of science:

> … the devil has gotten into my friends in the Physical Society and they want a talk from me on your latest papers: even though there are at least three people here—Abraham, Grossmann, and Weyl—who know a hundred times more about the topic than I do. I feel like someone for whom Beethoven has whistled his symphony and who now on the basis of that has to whistle after him—someone with the score in front of his eyes, but only being able to read it the way I read sheet music …[16]

---

14 In that case, these pages could be the manuscript referred to in the context of this lecture in (Speziali 1972, 73, fn. 1).

15 "Orechiante—Beethoven symphonie pfeifen—die 3 Fachmusiker sollen mich zurecht weisen"

It seems safe to assume that letter and essay were drafted around the same time. We are thus led to date this whole group of pages to 1916.

Could the troublesome Besso memo be from 1916 as well? More specifically, could the memo have been prepared for Besso's talk to the Zurich Physical Society? One can well imagine that Besso wanted to go over some of the stumbling blocks on Einstein's path to the latest version of the theory. He would have been in an ideal position to do so. The date on the memo could then either simply be a slip or refer to a particularly memorable discussion with Einstein, which Besso, known for his absent-mindedness,[17] nonetheless confused with one of the many other memorable discussions they must have had over the years. This is an extremely unlikely scenario. There is strong evidence that the Besso memo is from 1913. An important part of this evidence is that the memo touches on many of the same topics that are addressed in the first of the two groups of pages that I distinguished above, pages that are almost certainly from 1913–1914.

2.1.2 The First Group of Pages, the Einstein-Besso Manuscript, and the Besso Memo

The contents of the first group of new Besso pages related to general relativity strongly suggest that they belong to the Einstein-Besso manuscript on the perihelion motion of Mercury (CPAE 4, Doc. 14).

When the original 53 pages of the Einstein-Besso manuscript were published, the editors of CPAE 4 divided them into two parts. Part One, which makes up the bulk of the manuscript (about 43 pages), was thought to have been produced in close collaboration during a visit by Besso to Einstein in Zurich in the spring of 1913 (with possible additions by Einstein alone during the remainder of 1913). Part Two, comprising the remaining ten pages, was thought to have been produced by Besso alone some-

16 "… der Teufel ist in die Freunde der physikalischen Gesellschaft gefahren und sie wollen von mir einen Vortrag über deine neuesten Arbeiten: obwohl mindestens drei da sind, Abraham, Grossmann und Weyl, die die Sache hundert mal besser kennen als ich. Ich komme mir vor, wie einer dem Beethoven seine Symphonie vorgepfiffen hätte und der nun daraus nachpfeifen soll—die Partitur zwar vor Augen hat, aber sie eben so lesen kann, wie ich eine Partitur …" Michele Besso to Einstein, draft, 28 June 1916 (CPAE 8, Doc. 229).

17 In a letter of March 1901, Einstein relates a typical example: "Once again, Michele had nothing to do. So his principal sends him to the Casale power station to inspect and check the newly installed lines. Our hero decides to leave in the evening, to save valuable time, of course, but unfortunately he missed the train. The next day he remembered the commission too late. On the third day he went to the train on time, but realized, to his horror, that he no longer knew what he had been told to do; so he immediately wrote a postcard to the office, asking that they should wire him what he was supposed to do! I think he has a screw loose" ("Hatte der Michele einst wieder einmal nichts zu thun. Da schickt ihn sein Prinzipal in die Zentrale Casale, damit er die neu gemachte Leitungen inspiziere und prüfe. Unser Held entschließt sich, abends zu fahren, natürlich um kostbare Zeit zu sparen, versäumte aber leider den Zug. Am nächsten Tag dachte er zu spät an seinen Auftrag. Am dritten Tag ging er zeitig an die Bahn, merkte aber zu seinem Schrecken, daß er nicht mehr wußte, was man ihm aufgetragen hatte; er schrieb also sofort eine Karte ins Bureau, man solle ihm hintelegrafieren, was er zu thun hätte! Ich glaube, der ist nicht normal." Einstein to Mileva Maric, 27 March 1901 [CPAE 1, Doc. 94]).

time between the beginning of 1914 and the demise of the *Entwurf* theory in 1915.[18] The new pages appear to belong to Part Two.

Clear-cut evidence that the original 53 pages of the Einstein-Besso manuscript could not all have been produced in the spring of 1913 was provided by one page (CPAE 4, Doc. 14, [p. 53]) with various references to the article that Einstein prepared for his lecture of 23 September 1913 to the 85th meeting of the *Gesellschaft Deutscher Naturforscher und Ärzte* (*GDNA*) in Vienna. Einstein gave Besso page proofs of the Vienna lecture at some point,[19] but these were certainly not available before late August 1913.[20]

The decision of the editors of CPAE 4 to make January 1914 rather than August 1913 the lower limit for the date of Part Two was based on a letter from Einstein to Besso of early January 1914, which begins:

> Here you finally have your manuscript package. It is really a shame if you do not bring the matter to completion.[21]

The Einstein-Besso manuscript was eventually discovered in the Besso *Nachlass*. It was thus conjectured that Besso left the perihelion calculations with Einstein when he departed Zurich in June 1913 and that early in 1914 Einstein sent them to Besso (perhaps along with material added in the meantime), whereupon Besso added the pages that constitute Part Two of the manuscript. In view of the Besso memo, it is very likely, as I will argue below, that Besso and Einstein met again in late August 1913. This means that some pages—both among the new ones and among both parts of the published portion of the manuscript—may have been produced during this visit.

Most of the material in the first group of the new Besso pages appears to have been intended for a paper on the perihelion problem. First of all, I want to point out that this provides a perfectly respectable explanation for why Einstein never published the result that he and Besso had presumably already obtained in the spring of 1913, viz. that the *Entwurf* theory predicts a secular advance of $18''$ for the perihelion motion of Mercury on top of the Newtonian prediction.[22] The cynical explanation (Earman and Janssen 1993, p. 136) is that Einstein wanted to suppress this result because the

---

18 For a more complete statement than will be given here of the considerations that went into the dating of (what at the time of publication of CPAE 4 was available of) the Einstein–Besso manuscript, see sec. III of the editorial note, "The Einstein-Besso Manuscript on the Motion of Mercury's Perihelion," CPAE 4, 344–359.

19 In the draft of the letter of June 1916 quoted on p. 790 above, Besso referred to "galley proofs" ("Druckbögen") of the Vienna lecture. The reference occurs in a deletion that was silently omitted in the transcription of this document in CPAE 8. The relevant sentence reads: "Secondly, although I have as a paradigm your Vienna lecture ⟨in the old presentation⟩, equations 1d) und 7e′) page 19, ⟨*of the page proofs I have*⟩ I am not able to develop the corresponding relations according to the new gravitational equations" ("Zweitens: trotzdem ich als Paradigma ⟨in der alten Darstellung⟩ deinen wiener Vortrags habe, Gleichungen 1d) und 7e′) Seite 19, ⟨der mir vorliegenden Druckbögen⟩ bringe ich es nicht fertig, die entsprechenden Beziehungen nach den neuen Gravitationsgleichungen zu entwikkeln." My emphasis). In the printed versions of the lecture, (Einstein 1913a, 1914a), equations 1d) und 7e′) occur on p. 1261 and p. 23, respectively. The deleted clause "in the old presentation" presumably refers to the old version of Einstein's gravitational theory, i.e., the *Entwurf* theory.

Entwurf theory could not (yet) account for the full discrepancy of 43″ between the Newtonian prediction and the astronomical observations. A more charitable explanation is that Einstein and Besso agreed that the latter would continue to look for additional effects contributing to the perihelion motion and write up the results. Despite some prodding from Einstein (see the letter of early 1914 quoted above), Besso never completed this project. In December 1914, they were then scooped by Lorentz's student Johannes Droste, who independently found and published the basic result of 18″ (Droste 1914).[23] It seems reasonable to assume that no further additions to the Einstein-Besso manuscript were made after Droste's publication.

The new pages suggest that Besso actually had given up much earlier. The scope of the treatise on the perihelion problem that he envisioned was simply too ambitious. Consider the following table of contents which is among the new pages:

1. General Introduction
2. The equations of motion of ordinary relativity theory
3. Application of the Einsteinian equations of gravitation—they only lend themselves (initially) to iterative approximation.
   First approximation (for the rotation of the sun and for pressure forces)
   Second approximation: for the deviations from the Newtonian law.

	Equations of motion for spherical symmetry
	" " " axial symmetry
3a. Calculation for the case of rest	α. Deviations from Newtonian law etc.
	β. Pressure forces […]
3b. Calculation for the case of rotation	α. Rotation of the sun
	β. Perturbations by Jupiter

4. Scalar theories

20 Einstein submitted the manuscript of his lecture together with a short letter to Alexander Witting, the editor of the *Verhandlungen* of the *GDNA* (CPAE 5, Doc. 464). He did not date this letter and the first digit of the postmark ([-]1.VIII.13) is illegible. In CPAE 5, the date is given as 11 August 1913. Jürgen Renn, however, has argued persuasively that the manuscript was probably not finished until after 15 August 1913. The main text includes an argument against generally-covariant field equations (Einstein 1913a, 1258) that Einstein only found that day. (In a letter to H. A. Lorentz of 16 August 1913 [CPAE 5, Doc. 470], Einstein wrote that he found the argument the day before.) If the manuscript had been completed and sent off earlier, it is puzzling that this argument occurs in the main text whereas the hole argument is only alluded to in a footnote that appears to have been added later (see p. 801 below for the text and the dating of this footnote). This suggests that the illegible first digit of the postmark is a 2 or a 3 rather than a 1. Given that Einstein asked Witting for expedited processing of his manuscript so that he would have page proofs by September 5, it is unlikely that he sent the manuscript on August 31. In short, the most plausible date is 21 August 1913.

21 "Hier erhältst Du endlich Dein Manuskriptbündel. Es ist sehr schade wenn Du die Sache nicht zu Ende führst" Einstein to Michele Besso, after 1 January 1914 (CPAE 5, Doc. 499).

22 The value given in the Einstein-Besso manuscript is actually 1821″ (CPAE 4, Doc. 14, [p. 28]) but there are several indications that Einstein and Besso realized that this result was off by a factor 100 due to an error of a factor 10 in the value they used for the mass of the sun (see CPAE 4, Doc. 14, [p. 30], note 136, and [p. 35], note 161).

23 One might even go as far as to speculate that it was partly due to some lingering resentment over this missed opportunity that Einstein did not acknowledge Besso in his perihelion paper of November 1915 (Einstein 1915b).

5. Four-vector theories
6. Six-vector theories
7. Theory of Abraham
8. Comparison with observations. Introduction of solar oblateness. Mass of Venus out of corrected motion of Mercury.[24]

Calculations for most of the effects listed in sec. 3 (on the *Entwurf* theory) can be found in the published portion of the Einstein-Besso manuscript. Calculations for secs. “3a α” (which give the 18″ mentioned above) and “3b α” take up most of Part One;[25] calculations for secs. “3a β” and “3b β” can be found in Part Two.[26] A discussion, in Besso’s hand, of the iterative approximation procedure mentioned at the beginning of sec. 3 can be found on the verso of a letter from C.-E. Guye to Einstein of 31 May 1913 (CPAE 5, Doc. 443, and CPAE 4, Doc. 14, [p. 16]). Among the newly discovered Besso material, there is a page with a similar discussion of the approximation procedure.[27]

However, of the various alternative theories to be covered in secs. 4–7 in Besso’s table of contents, the Nordström theory is the only one for which any calculations have been preserved.[28] Among the new material there is one page of calculations related to p. 53 of the published portion of the Einstein-Besso manuscript, which summarizes Besso’s analysis of the perihelion motion predicted by the Nordström theory (this is the page with references to the Vienna lecture mentioned on p. 792).[29] Finally, there are no calculations related to solar oblateness[30] or the mass of Venus, two topics to be covered in sec. 8.

---

24 “1. Allgemeine Einleitung
2. Die gewöhnliche Relativitätstheoretische Bew.gl.
3. Anwendung d. Einst. Grav.-Gl.—Sie eignen sich (zunächst) nur zur stufenweisen Näherung
Erste Annäherung (für die Sonnenrotation und f. die Druckkräfte)
Zweite Annäherung: f. d. Abw[eichungen] vom Newtonschen Ges.
Bewegungsgleichungen f.Punktsymmetrie
" " " Axensymmetrie
3a. Ausrechnung f. d. Fall der Ruhe — α. Abweichungen vom Newtonschen Ges. etc.
β. Druckkräfte […]
3b. Ausrechnung f. d. Fall d. Rotation — α. Sonnenrotation
β. Jupiterstörungen
4. Skalare Th.
5. Vierer-V. Th.
6. Sechser-V. Th.
7. Abraham. Th.
8. Vergleich m. der Erfahrung. Einführung der Abplattung der Sonne
Venusmasse aus der korr. Merkurbewegung”

25 For an outline of these calculations, see secs. II.1 and II.2 of the editorial note, “The Einstein-Besso Manuscript on the Motion of Mercury’s Perihelion” (CPAE 4, 344–359).

26 See (CPAE 4, Doc. 14, [p. 51]) for calculations of the effect of solar pressure, and [p. 50] and [p. 46] for (badly flawed) calculations of the effect of Jupiter.

27 Following the notation of the *Entwurf* paper, Besso wrote the contravariant components of the metric as $\gamma_{\mu\nu}$ on this page. This supports dating the page to 1913–1914. Einstein adopted the modern notation $g^{\mu\nu}$ in (Einstein 1914e) published in November 1914.

The overly ambitious character of Besso's paper becomes even clearer if we look at a table on another one of the new pages. The table has rows for scalar theories, four-vector theories, and six-vector theories (cf. secs. 4–6 in the table of contents above). Nordström, Laue (with a question mark[31]) and Mie are listed under scalar theories and get one row each. The final row, labeled "tensor theory," is for the *Entwurf* theory. The table has eight columns with the following headers:

> characteristic assumptions | bending of light rays | effect of uniformly moving masses | effect of accelerated masses | perihelion motion | motion of nodes | effect of perturbations on perihelion motion | correction of Leverrier's value for the mass of Venus calculated on the basis of perturbations of Mercury[32]

Besso did not even begin the task of filling out this table.[33] He did, however, produce a draft, replete with deletions as seems to have been his style, of an introductory paragraph leading up to this table. There is another page with what appear to be drafts of yet another introductory paragraph, this time for a somewhat less ambitious version of the paper, focusing more narrowly on the *Entwurf* theory.

How much of the work for Besso's *magnum opus* remained to be done, is further illustrated by two pages with a numbered list of sixteen problem areas that Besso was still planning to address. One of the entries in this list is related to the problem of rotation:

> 15. Final result about centrifugal forces without Coriolis forces? Would that not have astronomical consequences as well?[34]

---

28 There is, however, a page with a calculation of the perihelion advance predicted by a theory proposed in (Hall 1894), in which Newton's $1/r^2$ -law is replaced by a $1/r^{2+\delta}$ -law. Besso's source for this hypothesis was probably (Newcomb 1895), which is explicitly cited on p. 31 of the published portion of the Einstein-Besso manuscript and on one of the new pages. For a brief discussion of the predictions of various alternative gravitational theories for the perihelion motion of Mercury, see (Earman and Janssen 1993, sec. 3).

29 The new page contains a derivation of the expression for the angle between perihelion and aphelion predicted by the Nordström theory given on [p. 53] of the Einstein-Besso manuscript. The perihelion advance (in radians per half a revolution) is given by the deviation of this angle from π . There is a sign error in the relevant expression on [p. 53], see (CPAE 4, Doc. 14, [eq. 372] and note 248 for an analysis of the calculation). The expression on this new page also contains this error.

30 Given the controversy decades later over whether solar oblateness throws off the general relativistic prediction for the advance of Mercury's perihelion by a few seconds of arc (Dicke and Goldenberg 1967; 1974), it is interesting that Besso also considered solar oblateness in 1913–1914.

31 I know of no gravitational theory by Laue, so I can understand Besso's question mark. Perhaps Besso confused Laue with Abraham. There is no row for Abraham's theory, despite its inclusion in Besso's table of contents (sec. 7).

32 "Karakteristische Annahmen | Lichtstrahlen Krüng | Einfluss gleichförmig bewegter Massen | Einfluss beschl. Massen | Perihelbew. | Knotenbew. | Peri Störungsbeeinfl. | Correktion d. Venusmasse von Leverriersch Wert, ermittelt aus den Störgen des Merkur."

33 Part Two of the published portion of the Einstein-Besso manuscript does contain calculations for the motion of nodes, the phenomenon listed in the sixth column, on the basis the *Entwurf* theory (CPAE 4, Doc. 14, [pp. 45–49], and related material on [p. 31] and [pp. 41–42]). For an outline of these calculations, see sec. II.3 of the editorial note, "The Einstein-Besso Manuscript on the Motion of Mercury's Perihelion," (CPAE 4, 344–359).

In the margin, Besso wrote down his answer to the second question:

> Need not be taken into consideration, if one does not want to consider the Oppolzer curvature.[35]

Oppolzer is the astronomer Theodor Egon von Oppolzer (1841–1886). Other than that, I have no idea what Besso meant by this comment. The Besso memo, however, contains a passage in which Besso appears to be making a similar point. In the last full paragraph of the memo, he writes:

> Since, in Einstein's gravitational theory, our noticing the absence of Coriolis and/or centrifugal forces still does not prove that we find ourselves in an allowed frame of reference, we need to take into account, in dealing with the astronomical problem, such a[n allowed?] system of forces imposed from the outside (Schwarzschild-Oppolzer motion of the fixed stars)[36]

I have only the vaguest of notions of what Besso might have meant by this convoluted sentence. I nonetheless find the similarity between these two passages quite striking.

The issue of the relation between centrifugal forces and Coriolis forces apparently continued to exercise Besso. Both in 1914 and in 1916 it crops up in his correspondence with Einstein.[37] As I argued in the final section of (Janssen 1999), the relevant letter from Einstein of 1916 is particularly revealing in this context.[38] Here I want to draw attention to the draft of a letter from Besso to Einstein of March 1914. In early March 1914, Einstein had written to Besso (CPAE 5, Doc. 514) about the analysis of the covariance properties of the *Entwurf* field equations to be published in (Einstein and Grossmann 1914b). In reply, Besso raised the question:

> Does the result you obtained also give a clue perhaps for a more complete treatment of the rotation problem, so that one can get the correct value for the centrifugal force? Unfortunately, my brain, the way it happens to have developed, is much too feeble to answer this question myself, or even to guess from what side it could be attacked. For reasons already discussed, it seems to me (?) of importance for the astronomical problem

34 "15. Definitives Resultat über Centrifugalkräfte ohne Corioliskräfte? Würde das nicht auch astronomisch in Betracht kommen können?"

35 "Kommt nicht in Betracht, wenn man nicht die Oppelzersche Krümmung betrachten will."

36 "Da bei der Einst. Gravitationstheorie durch Constatierung der Abwesenheit von Coriolis oder/und Centrifugalkräften noch nicht bewiesen ist, dass man sich in einem zulässigen Bezugssystem befindet, ist bei der astronomischen Aufgabe ein von aussen aufgeprägtes solches System von Kräften mit zu berücksichtigen (Schwarzschild-Oppolzerschen Fixsternbewegung)."

37 For the sake of completeness, I note that the 1916 portion of the new Besso material also contains a reference to the inertial forces of rotation. On the fourth page of the draft of his essay, Besso wrote sideways in the margin: "That a rotating disc produces a centrifugal and a Coriolis field in its vicinity, and an accelerated mass an acceleration field, are direct consequences of the *principle* of general relativity: *how big* the effects are is given by the *theory* of general relativity" ("Dass eine rot. Scheibe in der Nähe ein Zentrifugal und Coriolisfeld erzeugt, eine beschl. Masse ein Beschleunigungsfeld, sind direkte Konsequenzen des Allg. Relativitäts*prinzips*: *wie gross* die Wirkungen sind ergibt die allgemeine Relativitäts*theorie*.")

38 Einstein to Michele Besso, 31 July 1916 (CPAE 8, Doc. 245).

> as well (since earlier at least it looked as if a system in which there are no Coriolis forces could still be the seat of centrifugal forces or the other way around).[39]

This otherwise maddening passage does seem to make it clear that the comments on Coriolis forces and centrifugal forces in the new Besso material were written earlier. There is no discussion of these issues in any of the extant correspondence between Einstein and Besso prior to this letter. Besso's "for reasons already discussed" must therefore either refer to correspondence that is now lost or, more likely, to discussions he and Einstein had in person.

The new Besso material also contains a page with a derivation of the equations of motion for a point mass in the metric field of the form generated by a rotating mass distribution.[40] The calculation, under the heading "field corresponding to a Coriolis force" ("Feld, welches einer Corioliskraft entspricht"), starts from the general equation of motion of a point mass in a metric field as given in equations (6)–(7) of (Einstein and Grossmann 1913), just as similar calculations in the Einstein-Besso manuscript.[41]

The final page that we need to examine is the verso of the page with the calculation on the Nordström theory discussed earlier (see note 29). The passage written on this page once again touches on the problem of rotation as well as on other themes addressed in the Besso memo. It starts with equation (9b) of the Vienna lecture (Einstein 1913b, 1259). This equation expresses the vanishing of the divergence of the sum of a mixed tensor density describing the energy-momentum density of matter[42] and a similar quantity for the gravitational field. Besso writes:

39 "Gibt das erreichte Resultat vielleicht auch einen Wink für eine vollständigere Behandlung des Drehungsproblems, so dass man den richtigen Wert der Centrifugalkraft bekommen kann? Leider ist mein Kopf, wenigstens so wie er einmal erzogen ist, viel zu schwach, um mir die Frage selbst zu beantworten, oder auch nur zu ahnen, wo man sie angreifen könnte. Aus schon besprochenen Gründen scheint sie (?) mir auch für das astronomische Problem von Bedeutung (weil es früher wenigstens so aussah, dass ein System in welchem keine Corioliskräfte huschen, doch Sitz von Centrifugalkräften sein können oder umgekehrt)" Michele Besso to Einstein, draft, 20 March 1914 (CPAE 5, Doc. 516).

40 This calculation is done on a sheet of paper with the name and address of a Zurich hospital and sanatorium ("'Paracelsus' Neues Privat-Krankenhaus und Augenheilanstalt") printed at the top and the name and address of a nearby pharmacy ("Apotheke Th. Vogel") at the bottom. Following the header there is a printed "Rp.," which suggests that it is a (blank) prescription ("Rezept") form. Besso's calculation appears on the verso of the sheet. On the recto Besso wrote down a list of three short items, the third of which is: "Departure Wednesday evening 11:40" ("Abreise Mittwochabends 11:40"). We know from an entry in Ehrenfest's diaries that Besso's visit with Einstein in the spring of 1913 ended 18 June (CPAE 4, 357, note 57). For what it is worth, this was a Wednesday. In a letter to Jost Winteler, Besso's father-in-law, of 23 June 1913 (CPAE 5, Doc. 447), Einstein mentioned that Besso had suffered from "persistent diarrhea" ("einem hartnäckigen Darmkatharrh") during his visit, which could explain the prescription form.

41 See (CPAE 4, Doc. 14, p. 8, especially [eq. 48] and note 35, and [p. 20], especially [eqs. 133–137]).

42 In (Einstein 1913a, 1258), this quantity is defined as $\mathfrak{T}_{\sigma\nu} \equiv \sqrt{-g}g_{\sigma\mu}\Theta_{\mu\nu}$, where $\Theta_{\mu\nu}$ is Einstein's notation during this period for the contravariant components of the energy-momentum tensor.

Equation (9b) $\sum \frac{\partial}{\partial x_\nu}(\mathfrak{T}_{\sigma\nu} + t_{\sigma\nu}) = 0$

which expresses the conservation laws is of such character that it is only covariant under linear transformations in any case. In this way the linearity of possible transformations of the [field] equations of gravitation has been "proven" and has at the same time been reduced to the conservation laws. *Only* such frames of reference are justified for which the conservation laws hold. It remains an open question, however, whether *all* frames of reference for which the conservation laws hold are justified in the sense of the equations of gravitation. There remains of course the possibility that, through using higher than second-order derivatives, one arrives at different equations of gravitation; if everything that happens, however, is to be determined uniquely by them, there still has to be a criterion for the admissibility of the coordinate system anyway (cf. the one-sided deformation in empty space), for which purpose in Einstein's theory the conservation laws can consistently be used (which definitely was not automatic given the tensorial nature of the gravitational quantities and the full determination of the kinematics by them).

Do the equations of gravitation tell us something about the kinematics of the rotating system? Probably not, because one cannot transform it into a Galilean [system].[43]

Unfortunately, I do not understand Besso's intriguing parenthetical remark "cf. the one-sided deformation in empty space" ("vgl. die einseitige Deformation im leeren Raume"). Most of rest of the passage, however, sounds very familiar. The comment about the possibility of field equations containing higher-order derivatives of the metric echoes a similar comment in the *Entwurf* paper. Einstein had drawn attention to this possibility to argue that such higher-order equations might be generally covariant.[44] The hole argument, to which Besso alludes in the next sentence, would obviously restrict the covariance of such higher-order equations as well. Whereas the hole argument is only alluded to, the other argument against general covariance that Ein-

43 "Gleichung (9b) $\sum \frac{\partial}{\partial x_\nu}(\mathfrak{T}_{\sigma\nu} + t_{\sigma\nu}) = 0$

die die Erhaltungssätze ausdrückt, hat einen solchen Charakter, dass sie jedenfalls nur f. lineare Transf. covariant ist. Damit ist die Linearität der möglichen Transformationen bei den Gravitationsgl. «erwiesen» und gleichzeitig auf die Erhaltungssätze zurückgeführt. Es sind nur solche Bezugssysteme berechtigt, für welche die Erhaltungssätze gelten. Es bleibt aber die Frage offen ob alle Bez. syst. für welche die Erh. s. gelten berechtigte sind im Sinne der Gravitationsgleichungen. Dabei bleibt natürlich unbenommen, dass man durch Mitnahme höheren als die zweiten Ableitungen, andere Gravitationsgl. erhalten kann; soll aber durch dieselben das Geschehen eindeutig bestimmt sein, so muss immerhin noch ein Criterium für die Zulässigkeit des Coordinatensystems vorliegen (vgl. die einseitige Deformation im leeren Raume) als welches Criterium bei der Einst. Theorie die Erhaltungssätze widerspruchsfrei benutzt werden konnten (was durchaus nicht selbstverständlich war bei der Tensornatur der Gravitationsgrössen und der alleinigen Besti[mmu]ng der Kinematik durch dieselbe).

Lehren uns die Gravitationsgl. etwas über die Kinematik des rot. Systems? Wohl nein, da man es nicht auf ein Galileisches transformieren kann."

44 Einstein writes: "A priori one cannot deny the possibility that the final exact equations of gravitation could be of higher than second order. There is still the possibility therefore that the fully exact differential equations of gravitation could be covariant under arbitrary substitutions." ("A priori kann allerdings nicht in abrede gestellt werden, daß die endgültigen, genauen Gleichungen der Gravitation von höherer als zweiter Ordnung sein könnten. Es besteht daher immer noch die Möglichkeit, daß die vollkommen exakten Differentialgleichungen der Gravitation beliebigen Substitutionen gegenüber kovariant sein könnten" [Einstein and Grossmann 1913, 11–12]).

stein advanced in late 1913 is given in some detail at the beginning of this passage. The form of equation (9b) for the energy-momentum conservation law, the argument goes, restricts the covariance of the field equations to linear transformations.[45]

We encounter both these arguments in the Besso memo as well, although this time the argument from energy-momentum conservation is only referred to implicitly while (a prototype of) the hole argument is stated explicitly. At the bottom of the first and the top of the second page of the memo (see Figs. 1 and 2), immediately after the sentence quoted in the introduction about the problem of rotation, Besso writes:

> One can therefore not think of rotational forces as produced by the rotation of the fixed stars according to the Einsteinian [field] equations of gravitation, but one has to assume for them as for Galilean mechanics, that they only hold for an appropriately chosen [coordinate] system (which would be defined through the conservation laws)[46]

To distinguish the crude argument from energy-momentum conservation of late 1913 from a more sophisticated argument from energy-momentum conservation of early 1914, it is important to establish that Besso was under the impression at this point that energy-momentum conservation requires the field equations to be covariant under *linear* transformations only. Fortunately, this can be done with the help of a passage that starts at the bottom of the third and continues at the top of the fourth page of the memo:

> From the gravitational theory … it is then inferred that the position and acceleration (but not the velocity) of masses in each other's vicinity affect the line element; and (recently) that the gravitational field cannot be transformed away anywhere since an accelerated frame of reference (because the conservation laws are not satisfied) is not a justified frame[47]

Since Besso claims that energy-momentum conservation rules out *all* accelerated frames of reference, it is clear that he is referring to the crude rather than to the sophisticated argument. As was pointed out above (see footnote 20), we know that Einstein hit upon the crude argument on 15 August 1913. This strongly supports the date on the Besso memo. On 28 August 1913, Besso had every reason to refer to this argument as something that had only been found "recently" ("neuerdings").

On the second page of the memo, Besso indicated (with two arrows; see Fig. 2) two comments to be appended to the clause "which would be defined through the conservation laws." The second comment once again raises the open question that

---

45 For a concise discussion of this argument and its flaws, see (Norton 1984, sec. 5). This is the argument that was referred to in footnote 20.

46 "Man kann also die Rotationskräfte sich nicht hervorgebracht denken durch die Rotation der Fixsterne gemäss den Einst. schen Grav.-Gleichungen, sondern muss für diese annehmen wie für die Galileische Mechanik, dass sie nur für ein passend gewähltes System gelten (welches durch die Erhaltungssätze definiert wäre)"

47 "Aus der … Grav.-theorie wird dann geschlossen, dass die Lage und die Beschleunigung benachbarter Massen auf das Linienelement von Einfluss ist (nicht aber deren Geschwindigkeit); und (neuerdings) dass an keiner Stelle ein Gravitationsfeld wegtransformiert werden kann, da ein beschleunigtes Bezugssystem (wegen Nichterfüllung der Erhaltungssätze) kein berechtigtes System ist."

Besso had identified in the passage quoted on p. 798: “Is *every* system *that satisfies the conservation laws a justified* system?”[48] Then, as in the passage quoted on p. 798, he goes into what is essentially the hole argument, this time in some detail. Discussion of this passage will be postponed until sec. 4.

2.2 Piecing Together the Case for Accepting the Date on the Besso Memo

What conclusions can be drawn from the preceding survey of the new material about the date of the Besso memo? I think a strong case can be made for accepting Besso’s own dating of the document, even though this does require two additional assumptions: first, that an important footnote alluding to the hole argument was added to the text of the Vienna lecture in early September a few weeks before Einstein actually gave the lecture; second, that Besso visited Einstein in Zurich in late August 1913.

Perhaps the strongest evidence in support of Besso’s date of 28 August 1913 is that he refers to Einstein’s argument against general covariance from energy-momentum conservation of 15 August 1913 as something that was shown “recently” (“neuerdings”). In any case, as we shall see below, the reference to this argument rules out any date after 20 March 1914, when Besso expressed his satisfaction that Einstein had meanwhile replaced this crude argument by a more sophisticated one.[49] Let me emphasize that the memo would still present the puzzles described in the introduction even if it had been written in March 1914 instead of in August 1913.

The combination of the two arguments against general covariance referred to in the memo, the argument from energy-momentum conservation and the hole argument, is characteristic of the period late 1913–early 1914. This provides another reason for accepting the date on the Besso memo, even though 28 August 1913 seems to be a little early given the dates of other texts documenting this state of affairs. Consider the following passage from the printed summary of Einstein’s lecture to the 96th annual meeting of the *Schweizerische Naturforschende Gesellschaft* in Frauenfeld on 9 September 1913, less than two weeks after the date on the Besso memo. Einstein writes:

> On the other hand, it turns out to be logically impossible to formulate equations to determine the gravitational field (i.e., the $g_{\mu\nu}$) that are covariant with respect to arbitrary substitutions. Starting from the conservation laws of momentum and energy, we are led to choose the frame of reference […] in such a way that *only* linear transformations—but contrary to the situation in ordinary relativity theory *arbitrary* linear transformations—leave the equations covariant.[50]

48 “Ist *jedes* System, *welches den Erhaltungssätzen genügt*, ein *berechtigtes* System?”

49 See the quotation on p. 805 below from Besso to Einstein, draft, 20 March 1914.

50 “Dagegen erweist es sich als logisch unmöglich, Gleichungen zur Bestimmung des Gravitationsfeldes (d. h. der $g_{\mu\nu}$) aufzustellen, die bezüglich beliebigen Substitutionen kovariant sind. Wir gelangen, ausgehend von den Erhaltungssätzen des Impulses und der Energie, dazu, das Bezugssystem […] derart zu wählen, dass nur mehr lineare, aber im Gegensatz zur gewöhnlichen Relativitätstheorie beliebige lineare Substitutionen die Gleichungen kovariant lassen” (Einstein 1913b, 138).

Very similar comments can be found in the more extended version of the Frauenfeld lecture first published 15 December 1913 and then reprinted as (Einstein 1914d, see pp. 288–289). Unfortunately, it is not clear for either of these documents how close the printed texts are to what Einstein actually said in his lecture. What makes one suspicious is that the texts must have been completed after Einstein completed the printed text of the Vienna lecture, delivered on 23 September 1913.[51] In the latter, the hole argument is alluded to only in a footnote:

> Over the last few days, I found the proof that such a generally-covariant solution to the problem [of finding suitable gravitational field equations] cannot exist at all.[52]

If this footnote were added *after* the Vienna lecture was delivered, Einstein could not have alluded to the hole argument in Frauenfeld in early September even though such allusions do occur in the printed texts. And, more importantly for my purposes, Besso could not have known about the hole argument on 28 August and the date on the Besso memo would have to be incorrect. If we accept the date on the Besso memo, we therefore have to assume that Einstein added the footnote alluding to the hole argument to the text of his Vienna lecture sometime in early September, a few weeks *before* he delivered the lecture. This assumption is much more natural than it appears to be at first glance. Recall that Einstein submitted the manuscript for the lecture in late August and that he requested that page proofs be ready by September 5 (see note 20).

Unfortunately, the earliest extant correspondence documenting the state of affairs transpiring from these published texts and from the Besso memo (i.e., the combination of the two arguments against general covariance) is from November 1913, two months after the date on the memo. The letter to Hopf, already mentioned in the introduction, is from 2 November 1913. In addition there are two undated letters to Ehrenfest (CPAE 5, Docs. 481 and 484) that can both be dated to November 1913, one to early November, the other to the second half of the month. In this second letter, Einstein put the matter very concisely:

> A unique determination of the $g_{\mu\nu}$ by the [components of the energy-momentum tensor] $T_{\mu\nu}$ is only possible if special coordinate systems are chosen (rigorously provable). Energy-momentum conservation only allows *linear* substitutions.[53]

---

51 In CPAE 4, the texts related to the Frauenfeld lecture (Einstein 1913b; 1914d) are presented as Docs. 15 and 16, respectively, whereas the printed text of the Vienna lecture (Einstein 1913c) is presented as Doc. 17. Since the order of presentation in the Collected Papers is determined by the date of completion of the texts, the order should have been reversed.

52 "In den letzten Tagen fand ich den Beweis dafür, daß eine derartige allgemein kovariante Lösung überhaupt nicht existieren kann" (Einstein 1913a, 1257).

53 "Eine eindeutige Bestimmung der $g_{\mu\nu}$ aus den $T_{\mu\nu}$ ist nur bei Wahl spezieller Koordinatensysteme möglich (streng beweisbar). Die Impuls-[E]nergie-Erhaltung lässt nur lineare Substitutionen zu." Einstein to Paul Ehrenfest, second half of November 1913 (CPAE 5, Doc. 484). In both these letters to Ehrenfest, Einstein briefly explored the consequences of the restriction of the covariance of the field equations to linear transformations in geometrical terms. Einstein also refers to these ideas in the printed text of the Frauenfeld lecture (Einstein 1914d, 289).

Neither in this letter nor in the letters to Ehrenfest and Hopf from early November does Einstein say exactly when he found these results. It is thus perfectly conceivable that he already had them two months earlier, as we are forced to accept if we accept the date on the Besso memo, and that he actually did present them in his lecture in Frauenfeld in early September. Given all available evidence, I would say that this is not only conceivable but actually quite plausible.

A third consideration in support of accepting the date on the Besso memo is that Besso's date is consistent with the date one would assign to the memo on the basis of various connections with the group of new Besso pages belonging to the Einstein-Besso manuscript. These pages can firmly be dated to the period 1913–1914, and most of them can plausibly be dated more narrowly to late 1913. For instance, in the drafts (mentioned on p. 795) of an introductory paragraph for a paper on various astronomical predictions of the *Entwurf* theory, Besso refers to the *Entwurf* paper as "Z. f. M. & P. … *1913*" (my emphasis). This is the reprint of the *Entwurf* paper in the *Zeitschrift für Mathematik und Physik* (Einstein and Grossmann 1914a). In the first part of the Einstein-Besso manuscript of June 1913 (CPAE 4, Doc. 14, [p. 8]), Besso used the page numbers of the original separatum (Einstein and Grossmann 1913). The journal version was not published until 30 January 1914. This suggests that Besso wrote this page in late 1913, expecting the journal version to become available before the end of the year.

There is a particularly striking connection between the memo and one of the new pages of the Einstein-Besso manuscript with calculations for the Nordström theory on the recto (see note 29) and a discussion of the covariance of the field equations on the verso (see p. 798). The calculations on the recto are related both to the page on the Nordström theory in the published part of the Einstein-Besso manuscript (CPAE 4, Doc. 14, [p. 53]) and to the first entry of the Besso memo (see Fig. 1):[54]

> a) On the motion of planets in the Nordström theory:
> I do not know how to interpret d/ds (whether φ can be put in front of the differentiation since it is constant in time)—*no*, to be interpreted as co-moving with the point!
> The field comes out as in Newton[:] the area velocity [is] a constant. ⟨Where does that leave room for deviation from the Newtonian motion?⟩ Meaning of the coordinates? (Does not play a role given the cyclical integration)[55]

The discussion of the covariance of the field equations on the verso of this new page closely matches the corresponding discussion in the Besso memo.[56] First comes the (crude) argument from energy-momentum conservation, then the question whether

54 It may seem that the connection between the memo and the Vienna lecture through p. 53 of the Einstein-Besso manuscript amounts to an argument *against* accepting the date on the memo. 28 August, one could argue, is too late for Besso to have had access to the manuscript of the lecture, which Einstein probably submitted 21 August, and too early to have had access to page proofs, which Einstein probably did not receive until early September (see note 20 for the argument in support of these dates). It could well be, however, that the entry on the Nordström theory is somewhat earlier than the two pages of calculations that have been preserved. The questions raised in the memo are rather elementary and could refer to an earlier attempt to do these calculations.

all frames of reference in which the conservation laws hold are justified, and then the hole argument. Moreover, in both cases it is mentioned in this context that a rotating frame is not a justified frame.

I already drew attention to the striking similarity between a comment on the problem of rotation on the last page of the Besso memo and an entry in a list of open questions to be addressed in Besso's ambitious paper on the perihelion problem (see p. 795). In both cases, Besso comments on the problematic relation between centrifugal forces and Coriolis forces in Einstein's theory; in both cases, he emphasizes the astronomical relevance of the problem; and, for what it is worth, in both cases he refers to the astronomer Oppolzer.

There is one question though that still needs to be answered before we can feel comfortable accepting the date on the Besso memo and move on to the discussion of the resulting problems for reconstructing Einstein's path to general relativity. The question is this: how did Besso find out about Einstein's two arguments against general covariance in late August 1913.[57] Neither argument had been published yet. The argument from energy-momentum conservation was probably included in the page proofs of the Vienna lecture, but it is extremely unlikely that Besso would have read those by 28 August 1913. And, as we have seen (see the footnote quoted on p. 801), even in the final text of the Vienna lecture there is no more than an allusion to the hole argument.[58] There are also no letters from Einstein to Besso during this period explaining these arguments. But then no letters survive from the period between March 1912 and the end of 1913. There are two letters from Einstein to Besso of early 1914 and a draft of Besso's reply to the second and then there is another gap from late March 1914 to February 1915. So it is possible that there were more letters that are now lost, in one of which Einstein explained these arguments. It is much

55 "a) Zur Planetenbew. in der Nordströmschen Theorie
weiss ich nicht wie das d/ds zu verstehen ist (ob φ als zeitl. const. vor das Diff. zeichen zu nehmen ist)—nein, mit dem bewegten Punkt zu verstehen!
Es ergibt sich das Feld wie bei Newton[:] die Flächengeschwindigkeit eine Constante. ⟨Wo bleibt da noch Platz für ein Unterschied gegen die Newtonsche Bewegung?⟩ Bedeutung der Coordinaten? (Kommt nicht in Betracht bei der cycl. Integration[)]."
The quantity φ represents the scalar gravitational potential in the Nordström theory. In the Einstein-Besso manuscript, the perihelion motion is derived in part from a quantity called the "area velocity" (see CPAE 4, Doc. 14, [p. 8], note 38). If the area velocity is constant, Kepler's area law holds and there is no perihelion motion.

56 For reasons similar to those given in note 54, the passage in the memo is presumably earlier than the other passage, which starts with equation (9b) from the Vienna lecture.

57 Given everything we know about the interaction between Einstein and Besso, we can safely rule out the possibility that Einstein got these arguments from Besso.

58 Given that the first three of the four published versions of the hole argument are notoriously cryptic, the Besso memo reflects a remarkably thorough understanding of how the argument is supposed to work (as will become clear in sec. 4). It is also remarkable, that Besso, as we have seen (see p. 798), realized that the hole argument undercuts Einstein's claim in the *Entwurf* paper that higher-order gravitational field equations might be generally covariant.

more likely, however, that Einstein explained them to Besso in person during a visit by the latter to Zurich around 28 August 1913.[59]

This would provide a natural explanation for a striking feature of the Besso memo that I did not emphasize so far. At several points the memo records a certain back and forth: a question is raised to which an answer is appended (see, e.g., the entry on the Nordström theory quoted on p. 802) or an argument is put forward to which a counter-argument is appended. The most striking example in the latter category is the passage on the hole argument (see Fig. 2 for a facsimile reproduction and sec. 4 for a transcription and discussion). First, the argument is stated. In the next paragraph a counter-argument is given. Appended to this paragraph is a comment purporting to refute the counter-argument, which begins: "Of no use" ("Nützt nichts"). This same phrase is appended to a passage at the top of the page (although in that case it was subsequently deleted and replaced by another comment). A natural explanation for this back and forth is that Besso recorded Einstein's responses to some of the points he raised.

There is no direct independent evidence that Besso visited Einstein in Zurich in late August 1913,[60] but it is not at all implausible that such a visit took place. And all three extant letters between Einstein and Besso of 1914 yield important indirect evidence for this conjecture. The strongest evidence comes from the second of these letters. Einstein writes:

> The strict equivalence of inertial and gravitational mass, also of the gravitational field, I had, I believe, already proven by the time of your visit.[61]

The editors of CPAE 5 identify this visit as Besso's well-documented visit of June 1913, but they refer to an addendum to the published version of the Vienna lecture as the place where Einstein first published the result he mentions (Einstein 1914b). The result clearly belongs to the period of the Vienna lecture and not to the period shortly after the completion of the *Entwurf* paper. Hence, it is much more likely that Einstein is referring to a visit by Besso in late August than to the visit in June.

A second visit by Besso in August 1913 also fits better with Einstein's first letter to Besso of 1914, the one with which he presumably sent the Einstein-Besso manuscript. After exhorting Besso to finish the work on the perihelion problem, he writes:

---

59 The other possibility—of a visit during this period by Einstein to Besso in Görz, Austria (now Gorizia, Italy)—can be ruled out. In a letter to Freundlich received on 26 August 1913, Einstein wrote from Zurich: "I will be here until the middle of September, with the exception of September 7 & 8 when I have to give a talk in Frauenfeld." ("Ich bin bis Mitte September hier, ausgenommen den 7. & 8. September, wo ich in Frauenfeld einen Vortrag zu halten habe." Einstein to Erwin Freundlich, before 26 August 1913 [CPAE 5, Doc. 472]).

60 As I pointed out above, one of the new Besso pages is written on letterhead of a Zurich hospital, which would lead one to think that this page was written in Zurich. It may have been written, however, during Besso's visit of June 1913 (see note 40).

61 "Die strenge Aequivalenz der trägen und schweren Masse, auch des Gravitationsfeldes, hatte ich, glaube ich, schon zur Zeit deines Besuches bewiesen." Einstein to Michele Besso, ca. 10 March 1914 (CPAE 5, Doc. 514).

> I have not found out much since. But I have shown rigorously that the measure for the inertial as well as the gravitational mass of closed systems is given by the total rest energy of the system including its gravitational energy.[62]

The result mentioned in this passage is the same as the result mentioned in the letter of March 1914. This is obviously not the only result since Besso's visit of June 1913 that is worth mentioning. One need only think of the arguments against general covariance. If Einstein and Besso got together again in August 1913, it would presumably be in reference to this later visit, that Einstein writes that he has "not found out much since," which makes much more sense.[63]

I already pointed out that the draft of Besso's reply to Einstein's letter of March 1914 (see p. 796) requires us to assume either that some correspondence between Einstein and Besso is missing or that they had the opportunity to discuss the problem of rotation in person. A visit by Besso to Zurich in August 1913 would take care of that problem as well. It would also explain Besso's remarkably sharp grasp of the new stage of the *Entwurf* theory that was reached with Einstein's new treatment of the theory's covariance properties in (Einstein and Grossmann 1914b). In his letter of March 1914, Einstein briefly explained these developments to Besso. In the draft of his reply, Besso wrote (in a passage immediately preceding the passage quoted on p. 796):

> You already had the fundamental insight that the conservation laws represent the condition for positing an admissible coordinate system; but it did not appear to be ruled out that a restriction to Lorentz transformations was thereby essentially already given, so that nothing interesting epistemologically comes out of it. Now everything is fundamentally completely satisfactory.[64]

Besso's clear appreciation of the situation becomes readily understandable when we assume that he and Einstein had the opportunity in August 1913 to discuss the earlier stage of the theory in person.

The upshot of the discussion in this section then is that we have very good reason to believe (a) that the date on the Besso memo is accurate, and (b) that the memo reflects discussions between Einstein and Besso during a visit by the latter to Zurich around 28 August 1913. Of course, I largely avoided the two troublesome passages in

---

62 "Ich habe seitdem wenig herausgefunden. Strenge bewiesen habe ich aber, dass sowohl für die schwere wie für die träge Masse abgeschlossener Systeme die gesamte Ruheenergie des Systems mit Einschluss der Schwereenergie massgebend ist." Einstein to Michele Besso, after 1 January 1914 (CPAE 5, Doc. 499).

63 A slight discrepancy remains: the letter of January 1914 suggests that Einstein only found the result in question *after* Besso's last visit, whereas the letter of March 1914 suggests that he had already found the result *by the time of* this visit. In the letter of March 1914, however, Einstein also indicates that he is not entirely sure.

64 "Du hattest schon principiell eingesehen, dass die Erhaltungssätze die Bedingung für die Aufstellung eines zulässigen Coordinatensystems darstellen; aber es schien nicht ausgeschlossen, dass schon dadurch, im Wesentlichen, die Beschränkung auf die Lorentztransformationen gegeben sei, so dass nichts erkenntnistheoretisch besonders interessantes herauskam. Nun ist alles principiell vollkommen befriedigend." Michele Besso to Einstein, draft, 20 March 1914 (CPAE 5, Doc. 516).

the memo quoted in the introduction. In the remainder of this paper, I will focus on those two passages.

3. THE BESSO MEMO AND THE PROBLEM OF ROTATION

Consider, once again, the comments on the problem of rotation on the first page of the Besso memo (see Fig. 1 for a facsimile reproduction):

> If through rotation of a hollow sphere one produces a Coriolis field inside of it, then a centrifugal field, independent of the size of the hollow sphere, is produced that is not the same as the one that would occur in a rotating rigid system with the same Coriolis field. One can therefore not think of rotational forces as produced by the rotation of the fixed stars according to the Einsteinian gravitational [field] equations[65]

As will become clear below, this passage requires a serious modification of the reconstruction I gave in (Janssen 1999) of Einstein's struggles with rotation in the period 1913–1915. At the same time, however, it confirms a conjecture in my paper about the Machian motivation behind the specific way in which Einstein checked, both in 1913 and in 1915, whether or not the rotation metric is a vacuum solution of the *Entwurf* field equations. The strongest evidence for this conjecture offered in my paper came from a letter written after the fact.[66] The Besso memo provides contemporary evidence.

Let me first reiterate briefly how Einstein checked whether the rotation metric is a vacuum solution of the *Entwurf* field equations (see Janssen 1999, secs. 5–9, for details). The rotation metric has several components proportional to ω, the constant angular frequency with which the coordinate system rotates in Minkowski spacetime, and (in its covariant form) only one component with a term proportional to ω^2.[67] For a slowly rotating frame, the terms proportional to ω can be looked upon as small first-order deviations from the standard diagonal Minkowski metric; the term proportional to ω^2 as a second-order deviation. It is clear upon inspection that,

65 "Stellt man durch Rotation einer Hohlkugel ein Coriolisfeld in deren Innerem her, so entsteht ein Centrifugalfeld, unabhängig von der Grösse der Hohlkugel, welches nicht dem gleich ist, der in einem rotierenden starren System von gleichem Coriolisfeld statt finden würde. Man kann also die Rotationskräfte sich nicht hervorgebracht denken durch die Rotation der Fixsterne gemäss den Einst.schen Grav.-Gleichungen"

66 Einstein to Michele Besso, 31 July 1916 (CPAE 8, Doc. 245), quoted and analyzed in (Janssen 1999, sec. 11).

67 In a Cartesian coordinate system $x^\mu = (x, y, z, t)$ (with $\mu = 1, 2, 3, 4$ and with units such that the velocity of light $c = 1$), rotating counterclockwise around its z-axis with respect to some inertial frame in Minkowski spacetime, the Minkowski line element is given by:

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu = -dx^2 - dy^2 - dz^2 + (1-\omega^2 r^2)dt^2 + 2\omega y dx dt - 2\omega x dy dt$$

(where $r^2 \equiv x^2 + y^2 + z^2$). For a quick derivation of this expression, see p. 42R of the Zurich Notebook (discussed in "Commentary …" [in this volume], sec. 5.5.7). The non-vanishing components of the metric can easily be read off from the line element: $g_{11} = g_{22} = g_{33} = -1$, $g_{44} = 1 - \omega^2 r^2$, $g_{14} = g_{41} = \omega y$, and $g_{24} = g_{42} = -\omega x$.

to first order in ω, the rotation metric is a vacuum solution of the *Entwurf* field equations (Janssen 1999, 141–142). To check whether this is also true to second order in ω, Einstein used the same iterative approximation procedure that he used in his perihelion calculations (both in the Einstein-Besso manuscript and in Einstein 1915b). The second-order terms in the vacuum *Entwurf* field equations are of two kinds: terms with second-order derivatives of second-order terms in the metric, and terms with products of first-order derivatives of first-order terms in the metric. Inserting the ω-terms of the rotation metric into the latter and solving for the ω^2-term in the one component of the metric that has an ω^2-term in the case of the rotation metric, Einstein checked whether this ω^2-term is the same as the ω^2-term in the rotation metric. In 1913, he convinced himself that it is.[68] In 1915, he redid the calculation and discovered that it is not.[69] He reported the result in a letter to Erwin Freundlich of 31 September 1915 (CPAE 8, Doc. 123), calling it "*a blatant contradiction*" ("*ein flagranter Widerspruch*"). Rotation thus became the nemesis of the *Entwurf* theory.

I now turn to the Machian motivation behind Einstein's use of his iterative approximation procedure in this context. Inserting the rotation metric into the geodesic equation, one sees that the ω-terms give the Coriolis force, while the ω^2-term gives the centrifugal force (Janssen 1999, sec. 11). The "Coriolis field" mentioned in the quoted passage of the Besso memo accordingly refers to the ω-terms in a metric of the form of the rotation metric, and the "centrifugal field" refers to the ω^2-term in such a metric. In the Einstein-Besso manuscript, Einstein had, in first-order approximation and using Minkowskian boundary conditions at infinity, already solved the *Entwurf* field equations for a material source consisting of a hollow thin shell rotating with angular frequency ω (CPAE 4, Doc. 14, [pp. 36–37]). He referred to these calculations in a letter to Mach of June 1913 (to which I will return below), so it is clear that this calculation was done either during or shortly after Besso's visit to Zurich of June 1913. Einstein found that, to first order in ω, the metric field inside the rotating shell has the same form as the rotation metric. Contrary to what Besso claims in his memo, this metric is not independent of the size of the shell, but by choosing the

---

68 The relevant calculations can be found in the Einstein-Besso manuscript (CPAE 4, Doc. 14, [pp. 41–42]).

69 (A version of) this calculation has been preserved on a sheet of paper that Einstein subsequently used for the draft of a letter to Otto Naumann (CPAE 8, Doc. 124). The draft is from early October 1915, the calculation probably from late September 1915. On the verso of this document there is an aborted attempt to circumvent the problem. The calculation on the recto shows that the standard transformation to rotating coordinates in Minkowski spacetime leads to a metric field that is *not* a vacuum solution of the *Entwurf* field equations. On the verso, Einstein modified this transformation, presumably to check whether this would lead to a metric field that is. It is possible that Einstein had tried this strategy before, even though no records of such attempts survive. Given how quickly he gave up on it in 1915, it is unlikely, I would say, that he ever found an acceptable escape from the problem of rotation along these lines. One other calculation that might be related to the problem of rotation should be mentioned here. This calculation can be found in the 1913 portion of the Einstein-Besso manuscript (CPAE 4, Doc. 14, [pp. 43–44]). Unfortunately, the purpose of this calculation remains unclear.

radius and the mass of the shell appropriately, its metric field near the center can nonetheless be used to mimic the part of the rotation metric giving rise to the Coriolis force. The *Entwurf* theory thus seemed to provide the resources necessary to produce a Machian account of Newton's rotating bucket experiment.[70] If the inertial forces in a rotating frame in Minkowski spacetime can be interpreted as gravitational forces in a frame at rest, the rotation with respect to absolute space in Newton's explanation of the bucket experiment can be replaced by the relative rotation of the bucket with respect to the rest of the universe. Einstein's first-order calculations for a rotating hollow shell showed that the Coriolis force can be interpreted as a gravitational force due to rotating distant masses. Inserting the Coriolis field, the ω-terms in the rotation metric, into the second-order vacuum field equations and solving for the ω^2-terms, Einstein hoped to show that the centrifugal force can be interpreted as due to the gravitational field produced by this first-order field. This, I conjectured, is what lies behind Einstein's use of his iterative approximation procedure to check whether the rotation metric is a vacuum solution of the *Entwurf* field equations.

Rephrasing the quoted passage from the Besso memo using the terminology introduced above, one sees both why it provides strong evidence in support of this conjecture and why it is hard to square with everything else we know about Einstein's struggles with rotation. Starting from the ω-terms of the rotation metric (more precisely: from a Coriolis field inside a rotating shell mimicking the Coriolis field in a rotating coordinate system in Minkowski spacetime) and using the vacuum *Entwurf* field equations to calculate the ω^2-term, one finds an ω^2-term (the centrifugal field inside the rotating shell) that is *not* the same as the ω^2-term of the rotation metric (the centrifugal field in a rotating coordinate system in Minkowski spacetime). This is precisely the "blatant contradiction" that Einstein reported to Freundlich over two years later!

What was Einstein's response to Besso's claim? That is the question that will occupy us for most of the remainder of this section. One possibility that can be ruled out immediately is that Einstein *carefully* redid the calculation himself. If he had, he would have found that Besso was right. And the assumption that Einstein in August 1913 knew *for a fact* that the rotation metric is not a vacuum solution of the *Entwurf* field equations would render the developments of 1914–1915 completely incomprehensible. I believe that Einstein was quite capable of fooling himself, but not that he was *in denial* for two years about the problem of rotation.

So Einstein did not check Besso's claim, at least not carefully. The interesting question then becomes: did he accept it or not? Before I address that question, I want to examine (and reject) one concrete scenario in which Einstein made a half-hearted attempt to check Besso's claim and satisfied himself that it was false alarm.

70 At this point I need to add the usual disclaimer that I am concerned only with Einstein's reading of Mach, not with what Mach actually may or may not have said. See (Barbour and Pfister 1995) for further discussion.

In this scenario, the calculations in the Einstein-Besso manuscript with which Einstein confirmed that the rotation metric is a vacuum solution of the *Entwurf* field equations were done not in June 1913, as has so far been assumed, but in late 1913 *in response* to the claim in the Besso memo. As I emphasized in sec. 2, these pages could in principle have been written any time between June 1913 and January 1914 (when Einstein presumably sent the whole package to Besso). There is, however, a good argument against redating these particular pages. Examining Einstein's calculations on these pages, one gets the impression that he was so convinced that the rotation metric is a vacuum solution of the *Entwurf* field equations that he made the right errors in the right places to arrive at the result he expected all along (Janssen 1999, sec. 8). The analysis of the Zurich Notebook in this volume provides a good explanation for why *in June 1913* Einstein would have been so confident. When Einstein first derived and decided to publish the *Entwurf* field equations, he felt that these were the *only* equations that agree with Newtonian theory in the static weak-field limit while at the same time being compatible with energy-momentum conservation. Convinced of the existence of field equations satisfying *all* his heuristic requirements, Einstein must have had the strong expectation that the rotation metric would be a vacuum solution of the *Entwurf* field equations. This requirement had played a particularly important role in Einstein's search for such equations.[71] It is therefore much more likely that the calculations on [pp. 41–42] of the Einstein-Besso manuscript were done in June 1913, when Einstein was supremely confident that the rotation metric was a vacuum solution, than that they were done in August 1913, when Besso's calculations called this into question.

It is, of course, *conceivable* that Einstein in August 1913 was still sufficiently convinced that the rotation metric would be a vacuum solution to dismiss Besso's claim out of hand, trusting his own calculation of June 1913 over Besso's. In and of itself this would not even be that unreasonable. Besso's contributions to the Einstein-Besso manuscript contain several egregious mathematical errors.[72] If Einstein simply dismissed Besso's claim, the story I told in (Janssen 1999) does not have to be changed much. All that needs to be added is that Einstein missed yet another opportunity to discover that the rotation metric is not a vacuum solution of the *Entwurf* field equations.

---

71 See the Zurich Notebook, pp. 42L–43L, p. 24R, and p. 25R, discussed in "Commentary …" (in this volume), secs. 5.5.6–5.5.9, sec. 5.6.1, and sec. 5.6.4, respectively. On p. 24R, Einstein missed his first opportunity to discover that the rotation metric is not a vacuum solution of the *Entwurf* field equations. Since the *Entwurf* field equations are constructed so as to guarantee energy-momentum conservation, the divergence of $t_{\mu\nu}$, the mixed tensor density representing gravitational energy-momentum density in the *Entwurf* theory, vanishes for every vacuum solution of these field equations. The divergence of $t_{\mu\nu}$ does *not* vanish for the rotation metric, which implies that the rotation metric cannot be a vacuum solution of the *Entwurf* field equations. At the top of p. 24R, however, Einstein wrote that an expression that can be identified as the divergence of $t_{\mu\nu}$ *does* vanish for the rotation metric.

72 Compare, for instance, Einstein's impeccable derivation of the metric inside a rotating shell on pp. 36–37 of the manuscript to Besso's bungled derivation of the metric inside a rotating ring on p. 50 (see CPAE 4, Doc. 14, note 234).

The most plausible scenario, however, if we take into account all available source material, is that Einstein *accepted* Besso's claim. Recall that on 15 August 1913, less than two weeks before the discussions recorded in the Besso memo, Einstein had found the (fallacious) argument from energy-momentum conservation[73] that convinced him that it was a desirable rather than a problematic feature of the *Entwurf* theory that its field equations seemed to be invariant under linear transformations only. On the basis of this argument, Einstein was prepared to accept that the rotation metric is not a solution of these equations. This sounds like a perfectly natural scenario. Einstein clearly believed at this point that the *Entwurf* field equations are not invariant under the non-linear transformation to a rotating coordinate system. Does it not simply follow from this that the rotation metric cannot be a vacuum solution of these equations?

What complicates matters is that this does *not* follow, as I emphasized in (Janssen 1999, sec. 2). The statement that some field equations are invariant under rotation is much stronger than the statement that the rotation metric is a solution of these equations. The former is a statement about an *arbitrary* metric expressed in two coordinate systems rotating with respect to one another, the latter is a statement about a *specific* metric, the Minkowski metric, in two such coordinate systems. In late 1913–early 1914, Einstein could therefore consistently have held both the (mistaken) belief that the rotation metric is a vacuum solution of the *Entwurf* field equations and the (correct) belief that these equations are not invariant under rotation.

Long before the discovery of the Besso memo, Jürgen Renn (private communication) already rejected this analysis of the situation as "logical hair-splitting." The Besso memo, I think, has proven Renn right. The most plausible scenario, I now believe, is that Einstein, without careful consideration, accepted Besso's claim that the rotation metric is not a vacuum solution of the *Entwurf* field equations, and that he accepted it largely *because* he believed that the equations were not invariant under rotation anyway.

There is clear evidence, however, that less than two weeks earlier Einstein had been keenly aware of the logical picture I painted above. My analysis of the situation essentially turns on the distinction between "autonomous" ("selbständige") and "non-autonomous" ("unselbständige") transformations that Einstein himself made in a letter to Lorentz of 14 August 1913. "Autonomous transformations" are the usual type of transformations from old to new coordinates, in which the latter are simply functions of the former. "Non-autonomous transformations" are coordinate transformations in which the new coordinates depend not only on the old coordinates but also on the components of a specific metric field expressed in terms of the old coordinates. Although Einstein had not used the term "non-autonomous transformations" before, the notion already played an important role in the Zurich Notebook.[74] It was further developed in 1914 when Einstein began to analyze the covariance properties of the *Entwurf* field

73 See p. 798 for Besso's statement of the argument.

equations in terms of "justified" ("berechtigte") transformations between "adapted" ("angepasste") coordinate systems (Einstein and Grossmann 1914b, 221).

When Einstein first explicitly introduced the distinction between "autonomous" and "non-autonomous" transformations in his letter to Lorentz, the lack of covariance of the *Entwurf* field equations had become something of an embarrassment for him. He wrote:

> *But the gravitational* [field] *equations themselves unfortunately do not have the property of general covariance*. Only their covariance with respect to *linear* transformations is guaranteed. The entire confidence one has in the theory, however, rests on the conviction that acceleration of the frame of reference is equivalent to a gravitational field. Hence, if not all sets of equations of the theory, hence also equations (18) [one of the forms in which the "Entwurf" field equations are given in *(*Einstein and Grossmann 1913)], allow other transformations besides linear ones, then the theory refutes its own starting point and is left hanging in the air.[75]

After explaining the distinction between "autonomous" and "non-autonomous" transformations in the next paragraph, he continued:

> The existence of "autonomous" non-linear transformations [that preserve the form of the "Entwurf" field equations] is the simpler possibility; but that possibility does not seem to obtain, although I would not know how to prove this. The existence of "non-autonomous" non-linear transformations, however, already suffices to avoid a conflict a posteriori with the equivalence hypothesis.[76]

This suggests that Einstein wanted to do for the *Entwurf* field equations what he had done for several candidate field equations in the Zurich Notebook, viz. to check whether they allow transformations to uniformly rectilinearly accelerated frames as well as to uniformly rotating frames in the special case of Minkowski spacetime. Hence, on 14 August 1913, Einstein was fully aware of the possibility that the *Entwurf* field equations might not be invariant under *autonomous* transformations to rotating coordinates, while still being invariant under *non-autonomous* transformations to rotating coordinates in the special case that the metric in the old coordinates is the standard diagonal Minkowski metric.

---

74 See, e.g., p. 22L and p. 23R, discussed in "Commentary …" (in this volume), secs. 5.5.3 and 5.5.5, respectively). For general discussion of "non-autonomous transformations" see "Commentary …" (in this volume), the introduction to sec. 4.3.

75 "*Aber die Gravitationsgleichungen selbst haben die Eigenschaft der allgemeinen Kovarianz leider nicht*. Nur deren Kovarianz *linearen* Transformationen gegenüber ist gesichert. Nun beruht aber das ganze Vertrauen auf die Theorie auf der Überzeugung, dass Beschleunigung des Bezugssystems einem Schwerefeld äquivalent sei. Wenn also nicht alle Gleichungssysteme der Theorie, also auch Gleichungen (18), ausser den linearen noch andere Transformationen zulassen, so widerlegt die Theorie ihren eigenen Ausgangspunkt; sie steht dann in der Luft." Einstein to Lorentz, 14 August 1913 (CPAE 5, Doc. 467).

76 "Die Existenz "selbständiger" nicht linearer Transformationen ist die einfachere Möglichkeit; dies scheint aber nicht zuzutreffen, ohne dass ich dies zu beweisen wüsste. Es genügt aber schon die Existenz "unselbständiger" nicht linearer Transformationen, um mit der Aequivalenzhypothese nicht nachträglich in Konflikt zu geraten." Ibid.

Two days later, Einstein wrote to Lorentz again, telling him that the previous day he had hit upon the argument from energy-momentum conservation that, he felt, made it fully respectable that the *Entwurf* field equations appear to be invariant only under (autonomous) linear transformations. The tone of this second letter is markedly different from that of the first. The first letter, more than any other document I know of, shows Einstein deeply troubled by the lack of general covariance of his field equations. The second shows him greatly relieved. This second letter concludes:

> Only now does the theory please me, after this ugly dark spot seems to have been removed.[77]

The darkest hour had been right before the dawn.

Einstein was probably quite happy to abandon his quest for non-autonomous non-linear transformations leaving the *Entwurf* field equations invariant. What he conveniently neglected to mention in the first letter to Lorentz was that he had already tried this strategy during his search for suitable field equations recorded in the Zurich Notebook, and that despite extensive efforts it had never gotten him anywhere (see, e.g., "Commentary …" (in this volume), sec. 4.3.1–4.3.2[78] and secs. 5.5.5–5.5.9). He had now found an argument that seemed to show that invariance under autonomous linear transformations was enough. The *Entwurf* field equations, he felt, no longer needed to be invariant under non-autonomous transformations to rotating systems.[79]

77 "Erst jetzt macht mir die Theorie Vergnügen, nachdem dieser hässliche dunkle Fleck beseitigt zu sein scheint." Einstein to H. A. Lorentz, 16 August 1913 (CPAE 5, Doc. 470). Translation from (Norton 1984, sec. 5).

78 Einstein was quite happy to switch from looking for non-autonomous transformations under which the expression $\frac{\partial}{\partial x_\alpha}\left(\gamma_{\alpha\beta}\frac{\partial\gamma_{\mu\nu}}{\partial x_\beta}\right)$ transforms as a tensor to finding a suitable generalization of this expression that would transform as a tensor under ordinary autonomous transformations.

79 I can think of one other consideration that indicates that Einstein did not want to bother anymore with non-autonomous transformations once he had found his argument from energy-momentum conservation. The argument supposedly restricts the covariance of the field equations to those transformations preserving the form $\frac{\partial}{\partial x_\nu}(\mathfrak{T}_{\sigma\nu} + t_{\sigma\nu}) = 0$ of the law of energy-momentum conservation. Under Einstein's (erroneous) assumption that $t_{\sigma\nu}$ has the same transformation character as $\mathfrak{T}_{\sigma\nu}$, this law is indeed invariant only under linear transformations as far as *autonomous* transformations are concerned. But it might be invariant under *non*-linear *non-autonomous* transformations. In fact, it is precisely this possibility that Einstein exploited in (Einstein and Grossmann 1914b) and (Einstein 1914e), where the covariance properties of the *Entwurf* field equations are analyzed in terms of "justified" transformations between "adapted" coordinates. If we assume that Einstein in late 1913 still remembered and believed the (erroneous) result he had recorded in the Zurich Notebook, viz. that $\partial t_{\sigma\nu}/\partial x_\nu = 0$ for the rotation metric (see note 71 above), he would even have had an example of such a non-linear non-autonomous transformation preserving the form of the equation $\frac{\partial}{\partial x_\nu}(\mathfrak{T}_{\sigma\nu} + t_{\sigma\nu}) = 0$, viz. the transformation to rotating coordinates in the important special case where the old metric in the old coordinates is the Minkowski metric in its standard diagonal form.

The rotation metric no longer *needed* to be a solution. If Besso now told him that it *was* not a solution, so be it.

There still seems to be one weighty objection to this reconstruction of events. If the rotation metric was not a vacuum solution of the *Entwurf* field equations, then Einstein's hopes for a Machian account of Newton's bucket experiment were dashed. The Besso memo shows that even Besso clearly saw that implication. There is ample contemporary evidence for Einstein's Machian hopes for the *Entwurf* theory. And Einstein's recollections a few years later indicate that these Machian hopes became especially important once he had convinced himself that generally-covariant field equations were not to be had. In November 1916, Einstein wrote to Willem de Sitter:

> Psychologically, this conception [of a full relativity of inertia] played an important role for me, since it gave me the courage to continue to work on the problem when I absolutely could not find covariant field equations.[80]

In view of this letter, it is hard to believe that Einstein would have given up the hope that his theory would provide a Machian account of the bucket experiment just when he had come to accept that the covariance of the *Entwurf* field equations is extremely limited. These considerations then would seem to rule out that, in August 1913, Einstein would have accepted, even if only temporarily, that the rotation metric is not a vacuum solution of the *Entwurf* field equations. Comparison of two letters to Mach, however, one from June 1913 and one from December 1913, shows a remarkable change in Einstein's conception of what made his theory Machian. This change fits exactly with the reconstruction given above of Einstein's reaction to Besso's claim in the memo of August 1913.

First consider Einstein's letter to Mach of June 1913. In this letter, as in the letter from December 1913, Einstein acknowledged the inspiration he had drawn from Mach's work in formulating the *Entwurf* theory. If the theory's prediction of the bending of light could be confirmed during the solar eclipse of 1914,[81] Einstein enthusiastically told Mach, then

> your brilliant studies about the foundations of mechanics receive […] beautiful confirmation. For it necessarily follows [from the theory] that *inertia* has its origin in a form of *interaction* between bodies, completely in the sense of your considerations about Newton's bucket experiment.[82]

---

80 "Psychologisch hat diese Auffassung bei mir eine bedeutende Rolle gespielt; denn sie gab mir den Mut, an dem Problem weiterzuarbeiten, als es mir absolut nicht gelingen wollte, kovariante Feldgleichungen zu erlangen." Einstein to Willem de Sitter, 4 November 1916 (CPAE 8, Doc. 273). The letter is Einstein's first contribution to the famous Einstein–De Sitter debate. See the editorial note, "The Einstein–De Sitter–Weyl–Klein Debate," (CPAE 8, 351–357), for further discussion.

81 Einstein's astronomer friend Erwin Freundlich was planning an expedition to the Crimea for this purpose.

82 "erfahren Ihre geniale Untersuchungen über die Grundlagen der Mechanik […] eine glänzende Bestätigung. Denn es ergibt sich mit notwendigkeit, dass die *Trägheit* in einer Art *Wechselwirkung* der Körper ihren Ursprung hat, ganz im Sinne Ihrer Überlegungen zum Newton'sche Eimer-Versuch." Einstein to Ernst Mach, 25 June 1913 (CPAE 5, Doc. 448).

Einstein went on to describe two concrete examples of effects illustrating the relativity of inertia, and thereby the Machian character of his theory. I already discussed the second example, which is that of a rotating shell producing a Coriolis field. These same examples are mentioned in the first of the two letters to Lorentz of August 1913 discussed above. In both letters, Einstein emphasized that the effects will be very small. The relativity of inertia is also mentioned prominently in the two lectures of fall 1913, the Frauenfeld lecture and the Vienna lecture (see Einstein 1913a, sec. 9; 1913b, p. 138; and 1914d, p. 290). In the Vienna lecture, though not in the Frauenfeld lecture, Einstein once again emphasized that the effects will be small:

> unfortunately, the effect one expects is so small that we cannot hope to detect it in terrestrial experiments or in astronomy.[83]

My conjecture is that Einstein accepted in late 1913 that the relativity of inertia effects predicted by the *Entwurf* theory, for all their importance from a conceptual point of view, were too small to produce a full Machian account of the bucket experiment. I want to suggest that he could accept this *because* of the argument from energy-momentum conservation of August 1913. That argument, or so Einstein thought, explains why the frame in which the bucket is rotating is privileged over the frame in which the bucket is at rest.

This may sound like wild speculation but surprisingly strong evidence for this conjecture comes from a second letter from Einstein to Mach of December 1913. Einstein begins this letter in much the same way as he began his letter half a year earlier:

> It pleases me enormously that the development of the theory reveals the depth and the importance of your studies about the foundation of classical mechanics.[84]

Gone, however, are the references to the relativity of inertia "in the sense of [Mach's] considerations about Newton's bucket experiment." Einstein's identification of what makes his theory Machian has changed drastically:

> For me it is an absurdity to ascribe physical properties to "space." The totality of masses generates a $g_{\mu\nu}$-field (gravitational field), which in turn governs the unfolding of all events, including the propagation of light rays and the behavior of measuring rods and clocks. Everything that happens is initially described in terms of four *completely arbitrary* spatio-temporal variables. If the conservation laws for momentum and energy are to be satisfied, these variables then need to be specialized in such a way that only (fully) [Einstein probably meant 'arbitrary'; MJ] *linear* substitutions connect one justified frame of reference to another. The frame of reference is, in a manner of speaking, tailored to the existing world with the help of the energy law and loses its nebulous a priori existence.[85]

The important point is no longer that inertia comes out as an interaction between masses and that Newton's bucket experiment can be accounted for without invoking

83 "leider ist der zu erwartende Effekt so gering, daß wir nicht hoffen dürfen, ihn durch terrestrische Versuche oder in der Astronomie zu konstatieren" (Einstein 1913a, 1261–1262).

84 "Es freut mich ausserordentlich, dass bei der Entwickelung der Theorie die Tiefe und die Wichtigkeit Ihrer Untersuchungen über das Fundament der klassischen Mechanik offenkundig wird." Einstein to Ernst Mach, second half of December 1913 (CPAE 5, Doc. 495).

absolute motion. Rather it is that the law of energy-momentum conservation now explains why certain frames of reference are privileged. In other words, the argument from energy-momentum conservation of 15 August 1913 not only made it possible for Einstein to give up the Machian requirement that the rotation metric be a solution of the gravitational field equations, it also supplied what Einstein saw as a perfectly good alternative Machian feature of his theory!

I confess that it is not clear to me why Einstein ever felt that this was a satisfactory solution to the problem of absolute space.[86] Einstein's answer to the question why certain frames of reference are privileged—if I understand the passage quoted above correctly—is that in those frames the law of energy-momentum conservation has a particularly simple form. Is this really an answer or is it just a way of restating the question? How does Einstein's answer differ from the following "answer"? Inertial frames are privileged because in those frames Newton's second law takes the form $F = ma$. That just side-steps the question what it is about those frames that is responsible for Newton's second law taking this form. In the end, of course, none of Einstein's attempts to make his theory Machian ever panned out, but in the case of the other attempts it is at least clear how they could in principle have met the challenge of Newton's bucket experiment. In this case, not even that much is clear, at least not to me.

The important point, however, in this context is that Einstein believed that with his simple argument from energy-momentum conservation he had somehow solved the thorny problem of absolute space. And this removes what seemed to be the one strong objection to the scenario in which Einstein accepted in late 1913 that the rotation metric is not a solution of the *Entwurf* equations.

Einstein's satisfaction with this state of affairs was relatively short-lived. The simple argument from energy-momentum conservation of August 1913 was retracted in print in (Einstein and Grossmann 1914b, 218, footnote). This paper was not published until May 1914, but it was almost certainly completed by the time Einstein left Zurich for Berlin in March 1914.[87] As I pointed out in sec. 2, Einstein briefly outlined the new treatment of the covariance properties of the *Entwurf* field equations of (Einstein

---

85 "Für mich ist es absurd, dem "Raum" physikalische Eigenschaften zuzuschreiben. Die Gesamtheit der Massen erzeugt ein $g_{\mu\nu}$-Feld (Gravitationsfeld), das seinerseits den Ablauf aller Vorgänge, auch die Ausbreitung der Lichtstrahlen und das Verhalten der Massstäbe und Uhren regiert. Das Geschehen wird zunächst auf vier *ganz willkürliche* raum-zeitliche Variable[n] bezogen. Die müssen dann, wenn den Erhaltungssätzen des Impulses und der Energie Genüge geleistet werden soll, derart spezialisiert werden, dass nur (ganz) *lineare* Substitutionen von einem berechtigten Bezugssystem zu einem andern führen. Das Bezugssystem ist der bestehenden Welt mit Hilfe des Energiesatzes sozusagen angemessen und verliert seine nebulose apriorische Existenz." Ibid. One might be tempted to read the second sentence of this passage as an early statement of what later became "Mach's principle" (Einstein 1918), the requirement that the metric field be fully determined by matter. However, since the Minkowski metric in its standard diagonal form is a vacuum solution of the *Entwurf* field equations, the *Entwurf* theory certainly does not qualify as Machian by this criterion.

86 Einstein, like everybody else at the time, did not systematically make the distinction, at the heart of the modern analysis of these issues (see Earman 1989), between the problem of motion (absolute versus relative) and the problem of space (substantival versus relational).

and Grossmann 1914b) in a letter to Besso of early March 1914.[88] In particular, Einstein claimed that the rotation metric satisfies the condition formulated in that paper for "adapted" coordinates. As a matter of fact, it does not (see Janssen 1999, fn. 47). If Einstein explicitly checked this, he must have botched this calculation too. It is possible, however, that he reached his conclusion on the basis of more general considerations. Such considerations can be found in a letter to Lorentz at the beginning of the following year, January 1915. In this letter, Einstein claimed that the condition for "adapted" coordinates puts no restrictions whatsoever on the state of motion of allowed coordinate systems. Referring to secs. 12–14 of (Einstein 1914e), published in November 1914, he explained how the condition for "adapted" coordinates uniquely determines the coordinates inside some matter-free spacetime region Σ once the coordinates on the region's borders are specified. The letter then continues:

> From what has been said it is then also clear that linear transformations belong to the "justified" transformations. It is also clear that the *state of motion* of justified systems can be chosen arbitrarily since the coordinates at the regions' borders can be freely chosen; one can easily verify this for special cases.[89]

Unfortunately, I do not understand Einstein's argument, so I cannot pinpoint where it fails. *That* it fails is clear from the fact that the rotation metric does not satisfy the condition for "adapted" coordinates. The results in (Einstein 1914e) on which this argument of January 1915 turns are a natural extension of the results reported in (Einstein and Grossmann 1914b). So Einstein may well have hit upon this argument in early 1914. Regardless, however, of how exactly Einstein convinced himself in early 1914 that the rotation metric is a vacuum solution of the *Entwurf* field equations after all, it is clear that in doing so he essentially reverted to the position of the first of the two letters to Lorentz of August 1913 discussed above. He believed that the *Entwurf* field equations were invariant under a broad enough class of *non-autonomous* transformations to satisfy the demands of the equivalence principle and the generalized relativity principle.

At this juncture, an assumption I made earlier becomes crucial, viz. that Einstein accepted the claim about rotation in the Besso memo *without carefully checking it*. When in early 1914 his new approach indicated that the rotation metric is a vacuum solution of the *Entwurf* field equations, Einstein either had forgotten about Besso's

87 In June 1914, referring to (Einstein and Grossmann 1914b), Einstein wrote: "While still in Zurich I found the proof of the covariance of the gravitational [field] equations. Now the relativity theory has really been extended to arbitrarily moving systems" ("In Zürich fand ich noch den Nachweis der Kovarianz der Gravitationsgleichungen. Nun ist die Relat[ivitäts]theorie wirklich auf beliebig bewegte Systeme ausgedehnt." Einstein to Wilhelm Wien, 15 June 1914 [CPAE 8, Doc. 14]).

88 The uncharacteristically lucid part of the draft of Besso's reply is quoted on p. 805; the more typical muddle-headed part on p. 796.

89 "Aus dem Gesagten ist dann auch klar, dass die linearen T[ra]nsformationen zu den "berechtigten" gehören. Ebenso ist klar, dass der *Bewegungszustand* berechtigter Systeme willkürlich wählbar ist, da die Koordinatenwahl an den Gebietsgrenzen frei ist; man kann sich dies an Spezialfällen leicht vergegenwärtigen." Einstein to H. A. Lorentz, 23 January 1915 (CPAE 8, Doc. 47).

earlier calculations or assumed they must have been in error. In his letter to Besso of March 1914, he did not refer to these calculations at all. Besso accepted the verdict implied by this omission. In the draft of his reply, he modestly inquired whether the new treatment gives the right answer for the centrifugal force.[90] Theirs was not a partnership of equals.

By convincing himself that the *Entwurf* field equations allow non-autonomous transformations to rotating coordinates in Minkowski spacetime, Einstein also convinced himself that the *Entwurf* theory vindicates the Machian account of Newton's bucket experiment. The latter claim is featured prominently in what was clearly meant to be the definitive exposition of the *Entwurf* theory submitted to the Prussian Academy on 29 October 1914. In the introduction, Einstein wrote:

> Let … K be a justified coordinate system in the Galilean-Newtonian sense, and let K' be a coordinate system rotating uniformly with respect K. In that case there will be centrifugal forces acting on masses at rest with respect to K', whereas there will be no such forces on masses at rest with respect to K. Newton already considered this proof that one has to look upon the rotation of K' as an "absolute" rotation, and that one is not equally justified in treating K' as being "at rest" as one is with K. This argument, however—as has been shown in particular by E. Mach—is not valid. We do not necessarily have to attribute the existence of centrifugal forces to the motion of K'; we can just as well attribute them to the average rotational motion of the ponderable distant masses of the surroundings with respect to K', where we treat K' as being "at rest."[91]

So, by the end of 1914, any doubts Einstein might have had concerning rotation had completely evaporated.[92]

In March 1915, the Italian mathematician Tullio Levi-Civita produced a coordinate transformation that turned the Minkowski metric in its standard diagonal form into a metric field that is no longer a vacuum solution of the *Entwurf* field equations, even though the transformation was specifically constructed in such a way that it was "justified" ("berechtigt") according to Einstein's own criterion for "adapted" ("angepasste") coordinate systems.[93] Einstein had claimed that this condition is both necessary and sufficient for coordinates to be adapted to a given metric field. Einstein tried to get around Levi-Civita's counter-example, but Levi-Civita remained uncon-

---

90 See the quotation on p. 796 from Besso to Einstein, draft, 20 March 1914.

91 "Es sei … K ein im Galilei-Newtonschen Sinne berechtigtes Koordinatensystem, K' ein relativ zu K gleichförmig rotierendes Koordinatensystem. Dann wirken auf relativ zu K' ruhende Massen Zentrifugalkräfte, während auf relativ zu K ruhende Massen solche nicht wirken. Hierin sah bereits Newton einen Beweis dafür, daß man die Rotation von K' als eine »absolute« aufzufassen habe, daß man also K' nicht mit demselben Rechte wie K als »ruhend« behandeln könne. Dies Argument ist aber—wie insbesondere E. Mach ausgeführt hat—nicht stichhaltig. Die Existenz jener Zentrifugalkräfte brauchen wir nämlich nicht notwendig auf eine Bewegung von K' zurückzuführen; wir können sie vielmehr ebensogut zurückführen auf die durchschnittliche Rotationsbewegung der ponderablen fernen Massen der Umgebung in bezug auf K', wobei wir K' als »ruhend« behandeln" (Einstein 1914e, p. 1031).

92 It is very telling that the title of (Einstein 1914e) proudly announces a "general" ("allgemeinen") theory of relativity, whereas the titles of (Einstein and Grossmann 1913; 1914b; and Einstein 1914c) had more modestly announced only a "generalized" ("verallgemeinerten") theory of relativity.

vinced, and the exchange between the two men about the covariance properties of the *Entwurf* field equations ended in a stalemate (see Einstein to Tullio Levi-Civita, 5 May 1915 [CPAE 8, Doc. 80]). It appears that Einstein remained supremely confident of his theory, but perhaps Levi-Civita's criticism got under his skin more than he cared to admit. There is no indication, however, that Einstein expressed any doubts about his theory in his Wolfskehl lectures in Göttingen in the summer of 1915. And although he preferred not to include any papers on the *Entwurf* theory in a proposed new edition of the Teubner anthology on the principle of relativity, this does not seem to have been because of any doubts concerning the theory (see Einstein to Arnold Sommerfeld, 15 July 1915 [CPAE 8, Doc. 96]).

Still, at some point during the late summer of 1915, Einstein must have gotten sufficiently worried about his theory to subject it to a test the outcome of which, be it positive or negative, he had essentially taken for granted for two years: is the rotation metric a vacuum solution of the *Entwurf* field equations or not? It could very well be, as Jürgen Renn has suggested to me, that Besso brought up the problem of rotation again during Einstein's visit to Zurich in September 1915 and that this is what finally made Einstein decide to repeat the calculation carefully. Despite the early warning signs recorded in the Besso memo two years earlier, one can understand the shock that comes through in the letter to Freundlich in which Einstein reported that the rotation metric is definitely *not* a solution.

This is the most convincing reconstruction of Einstein's struggles with rotation in the period 1913–1915 that I can come up with given all available source material. In (Janssen 1999), I argued that Einstein probably believed throughout the life span of the *Entwurf* theory that the rotation metric is a vacuum solution of its field equations. I now believe that there is strong evidence that there was a period in late 1913–early 1914 when Einstein believed that the rotation metric is actually *not* a solution. The comments about rotation in the Besso memo constitute the most important part of this evidence, but striking corroborating evidence comes from the comparison of Einstein's letters to Mach of June and December 1913.

In the reconstruction given in this section, one feature of the reconstruction I gave in (Janssen 1999) gets strongly amplified. One clearly sees a certain opportunistic streak in Einstein's *modus operandi*. In (Janssen 1999), I argued that Einstein on at least two occasions[94] missed the problem of rotation in his *Entwurf* theory because he was too convinced that his calculations would bear out his *stable* expectation that the rotation metric would be a vacuum solution of the *Entwurf* field equations. In the new reconstruction, things get worse in two (rather different) ways.

First, Einstein repeatedly failed to check carefully whether his theory fulfilled his expectations even though these expectations *changed* over time. In June 1913 and in

93 Tullio Levi-Civita to Einstein, 28 March 1915 (CPAE 8, Doc. 67). For a discussion of this exchange between Einstein and Levi-Civita, see (Cattani and De Maria 1989). This specific result is discussed on pp. 189–190.

94 In June 1913, when he did the calculations on pp. 41–42 of the Einstein-Besso manuscript, and in March 1914, when he checked whether the rotation metric satisfies the condition for adapted coordinates.

March 1914 he expected the rotation metric to be a vacuum solution of the *Entwurf* field equations, but in August 1913 he expected it not to be. Yet, in none of these cases did he take the trouble to do a relatively simple calculation to establish once and for all whether the rotation metric is a solution or not.

Secondly, Einstein drastically changed the formulation of what would make his theory Machian according to what he thought the *Entwurf* theory could deliver. When he was under the impression that the rotation metric was a vacuum solution of the field equations, he claimed that rotating and non-rotating frames of reference are equivalent in his theory. When he was under the impression that it was not, he claimed that the theory explained why rotating and non-rotating frames of reference are not equivalent. I am not saying that this was an unreasonable thing to do. On the contrary, it would have been foolish for Einstein to hold on stubbornly to the letter of his heuristic requirements if he felt that an otherwise attractive theory simply lacked the resources to meet this or that requirement. Creative scientists may need a healthy dose of opportunism. We shall see more manifestations of this trait of Einstein when we turn to the hole argument in the next section.[95]

4. THE BESSO MEMO AND THE HOLE ARGUMENT

We now finally get to the most intriguing passage of the Besso memo, the bottom half of the second page, where in three short paragraphs we find, first, the earliest extant version of the hole argument, second, a promising proposal for an escape from the argument, and, third, a brusque rejection of this escape. A facsimile reproduction of this passage can be found in Fig. 2. The first of the three paragraphs reads:

> The requirement of ⟨general⟩ covariance of the gravitational equations under arbitrary transformations cannot be imposed: if all matter ⟨is given⟩ were contained in one part of space and for this part of space a coordinate system [is given], then outside of it the coordinate system could still ⟨essentially⟩ except for boundary conditions be chosen arbitrarily, ⟨through which the g arbitrarily⟩ so that a unique determinability of the g's cannot be obtained.[96]

In essence, this is (a qualitative description of) the hole argument: if the field equations for the metric field are generally covariant, then a given matter distribution does not uniquely determine the metric field in matter-free regions. It is not essential to the

95 The two opportunistic moves distinguished above nicely illustrate an insightful distinction introduced in (Earman and Eisenstaedt 1999, 230) and amplified in (Kennefick 2005, 119–120). There are (at least) two kinds of opportunism: the garden variety—what Earman and Eisenstaedt call "unscrupulous opportunism"—and a much rarer form, which they call, using a beautiful oxymoron, "principled opportunism." Einstein, they argue and Kennefick concurs, was the master of the latter. Opportunism comes into play when there is a conflict between a principle (e.g., relativity of motion) and some concrete result (e.g., the rotation metric is not a solution of the vacuum *Entwurf* field equations). Faced with such a conflict, the unscrupulous opportunist changes the principle (as Einstein did when he redefined what makes a theory Machian), whereas the principled opportunist finds ways to ignore the result (as Einstein did with the troublesome rotation metric). I am grateful to Dan Kennefick for making me appreciate this point.

argument which regions contain matter and which regions do not. It is also not essential whether the argument is phrased in terms of regions of four-dimensional spacetime or in terms of regions of three-dimensional space. In the published versions of the hole argument, there is a finite matter-free region of spacetime—the "hole" from which the argument derives its name[97]—while there can be matter everywhere else.[98] In the version of the Besso memo it is just the other way around. There is a finite region of space in which all matter is concentrated while there is no matter anywhere else. This, of course, also divides spacetime into regions with and without matter. So, what Besso describes could be called an inverted hole argument or a hole argument without a hole. It should not be surprising that the hole argument is not yet stated in its canonical form given that Einstein only found it somewhere between 16 August 1913, when he reported the argument against general covariance from energy-momentum conservation to Lorentz, and 28 August 1913, the date of the Besso memo.

The formulation of the hole argument in the Besso memo suggests that the argument originated in concerns about the uniqueness of the metric field of the sun, which Einstein and Besso calculated in their attempt to account for the perihelion anomaly on the basis of the *Entwurf* theory. In his summary of the iterative approximation procedure used to find the static spherically symmetric solution of the *Entwurf* field equations to represent the metric field of the sun, Besso explicitly raised the issue of the solution's uniqueness: "Is the static gravitational field [. . .] a particular solution? Or is it the general solution expressed in particular coordinates?" (CPAE 4, Doc. 14, [p. 16]).[99]

The second paragraph of the passage from the Besso memo is the one quoted in the introduction. It proposes an escape from the hole argument:

> It is, however, not necessary that the g themselves are determined uniquely, only the observable phenomena in the gravitation space, e.g., the motion of a material point, must be.[100]

The more cryptic third paragraph, appended to the second, purports to show why the escape fails:

---

96 "Die Anforderung der ⟨allgemeinen⟩ Covarianz der Gravitationsgleichungen für beliebige Transformationen kann nicht aufgestellt werden: wenn in einem Teile des Raumes alle Materie ⟨gegeben ist⟩ enthalten wäre und für diesen Teil ein Coordinatensystem, so könnte doch ausserhalb desselben das Coordinatensystem noch, ⟨im wesentlichen⟩ abgesehen von den Grenzbedingungen, beliebig gewählt werden, ⟨wodurch die g beliebig eine⟩ so dass eine eindeutige Bestimmbarkeit der g s nicht eintreten könne."

97 I know of only one place where Einstein explicitly used the phrase "hole argument" ("Lochbetrachtung") in writing and that is in his letter to Besso of 3 January 1916 (CPAE 8, Doc. 178). In the first two published versions of the argument, the letter '*L*,' which presumably stands for "Loch" (Stachel 1989, 71), is used to designate the matter-free region (Einstein and Grossmann 1914a, 260; 1914b, 217–218). In the last two published versions, the matter-free region is designated by 'Φ' (Einstein 1914c, 178) and by 'Σ' (Einstein 1914e, 1067), respectively.

> Of no use, since with ⟨the⟩ a solution a motion is also fully given. If in coordinate system 1, there is a solution K_1, then this same construct is also a solution in 2, K_2; K_2, however, also a solution in 1.[101]

As I explained in sec. 2 (see p. 803–805), I am assuming that the Besso memo is the record of discussions between Einstein and Besso during a visit by the latter to Zurich in late August 1913. On this assumption, the first of the three paragraphs is most naturally understood as Besso's formulation of an embryonic version of the hole argument that Einstein had just told him about; the second as giving Besso's own proposal for an escape from the argument; and the third as giving (Besso's recollection of) Einstein's negative response to this proposal. As I pointed out in the introduction, Besso's proposal does not seem to be all that different from the point-coincidence argument with which Einstein himself explained the failure of the hole argument over two years later in letters to Ehrenfest *and Besso*. The main task of this section therefore will be to make it plausible that Einstein was aware of the escape proposed by Besso as early as August 1913 and nonetheless continued to trot out the hole argument for the next two years.

Before I get to this task, I want to point out that the passage on the hole argument in the Besso memo, like the passage on rotation discussed in sec. 3, supports the existing reconstruction of events in at least one important respect even though it may undermine it in others.[102] The last sentence of the third paragraph provides strong support for the interpretation of the first three published versions of the hole argument as cryptic statements of the fourth[103] (Norton 1984, 131).[104]

---

98 Einstein's hole argument is not quite the same as the argument advanced under the same name in (Earman and Norton 1987). Discussions of the latter account for a disproportionate fraction of the recent literature in philosophy of space and time. In one of the more interesting early contributions to this debate, Tim Maudlin gave a concise statement of how Einstein's argument differs from Earman and Norton's: "the question [in Einstein's case] is not whether the entire state of the universe outside the hole determines the state inside [as in Earman and Norton's case]. Rather the question is whether the stress-energy tensor […] defined everywhere determines the metric […] everywhere" (Maudlin 1990, 556). Maudlin goes on to suggest that Einstein was not so much worried about a conflict between general covariance and determinism (which is the focus of Norton and Earman's hole argument), but rather about a conflict between general covariance and Mach's principle, the requirement that matter fully determines the metric field. What militates against Maudlin's suggestion is that this requirement was not formulated until three and a half years later in Einstein to Willem de Sitter, 24 March 1917 (CPAE 8, Doc. 317) and was only published in (Einstein 1918). In late 1913, as we saw in sec. 3, Einstein still articulated his Machian intuitions in terms of the relativity of inertia. Einstein moreover explicitly introduced the final version of the hole argument in terms of a conflict between general covariance and the "law of causality" ("Kausalgesetz," Einstein 1914e, 1066). It is true that Maudlin only proposed his Machian interpretation for the first three published versions of the hole argument, which he sharply distinguished from the fourth, but the Besso memo, as we shall see, strongly suggests that there is no such distinction and that the first three versions are just cryptic formulations of the fourth. Despite these reservations, I find it an appealing suggestion that the worries about determinism and causality that are behind Einstein's hole argument have strong Machian overtones. This same suggestion was made in (Hoefer 1994).

99 "Ist das stat Schwerefeld [. . .] ein spezielles? Oder ist es das allgemeine, auf spec. Coordinaten zurück geführtes?" I owe this observation about the origin of the hole argument to Jürgen Renn.

To make good on this claim, I need to remind the reader of some of the fine print of the argument. It will be convenient to do so for the canonical form of the argument, rather than for the "inverted" form in the Besso memo. Consider two coordinate systems—one with coordinates x^μ, the other with coordinates x'^μ—that differ only in some finite matter-free region of spacetime ("the hole"). It follows that the components of the energy-momentum tensor, which describe the matter distribution, are given by the exact same functions of the coordinates in these two coordinate systems: $T_{\mu\nu}(x) = T'_{\mu\nu}(x')$.[105] Suppose we have some local field equations setting some differential operator acting on the metric tensor equal to the energy-momentum tensor.[106] Suppose $g_{\mu\nu}(x)$ is a solution of these field equations in the unprimed coordinate system for the matter distribution $T_{\mu\nu}(x)$. If the field equations are invariant under the transformation from x^μ to x'^μ, as would be case for generally-covariant field equations, then $g'_{\mu\nu}(x')$ will be a solution for $T'_{\mu\nu}(x')$ in the primed coordinate system. Inside the hole, $g_{\mu\nu}(x) \neq g'_{\mu\nu}(x')$[107] even though $T_{\mu\nu}(x) = T'_{\mu\nu}(x')$ everywhere.

The standard interpretation of the hole argument before the work of John Stachel and John Norton in the 1980s, was that Einstein thought that the inequality $g_{\mu\nu}(x) \neq g'_{\mu\nu}(x')$ amounted to indeterminism and that because of this the field equations could not be allowed to be invariant under such transformations as the one from x^μ to x'^μ (see, e.g., Pais 1982, 222). Needless to say, this interpretation is not very

---

100 "Es ist nun allerdings nicht nötig, dass die g selbst eindeutig bestimmt sind, sondern nur die im Gravitationsraum beobachtbaren Erscheinungen, z.B. die Bewegung des materiellen Punktes, müssen es sein."

101 "Nützt nichts, denn durch ⟨der⟩ eine Lösung ist auch eine Bewegung voll gegeben. Ist im Coordinatensystem 1 eine Lösung K_1, so ist dieses selbe Gebilde auch eine Lösung in 2, K_2; K_2 aber auch eine Lösung in 1." The deleted fragment at the bottom of the page ("⟨*Es h*⟩") is the beginning of the (deleted) first sentence of the third page: "⟨After all, it is only called *co*variance, not *in*variance of the gravitational [field] equations!?⟩" ("⟨Es heisst aber auch bloss *Co*varianz, nicht *In*varianz, der Grav. gl.!?⟩").

102 In the case of rotation, the Besso memo provided strong support for my earlier conjecture that Einstein used his iterative approximation procedure to check whether the rotation metric is a vacuum solution of the *Entwurf* field equations in order to interpret the inertial forces of rotation as gravitational forces due to distant rotating masses. At the same time, it seriously undermined my earlier assumption that from June 1913 till the end of September 1915 Einstein never wavered in his belief that the rotation metric to be a vacuum solution of the *Entwurf* field equations.

103 See note 97 for detailed references to all four versions.

104 Even in (Stachel 1989), the possibility is left open that the fourth version "may represent a significant evolution in Einstein's thinking about the "hole" argument" (ibid., 72).

105 Let me state this equality somewhat more precisely with the help of some modern terminology. At every point p of the manifold covered by two maps—one with coordinates x^μ, the other with coordinates x'^μ—the components $T_{\mu\nu}$ at p in x^μ-coordinates are equal to the components of $T'_{\mu\nu}$ at p in x'^μ-coordinates. A more explicit form of the equality would be $T_{\mu\nu}(x^\rho_p) = T'_{\mu\nu}(x'^\rho_p)$. For points p outside and on the edge of the hole, the equality holds because $x^\rho_p = x'^\rho_p$. For points p inside the hole, the equality holds, even though $x^\rho_p \neq x'^\rho_p$, simply because $T_{\mu\nu}(x^\rho_p) = T'_{\mu\nu}(x'^\rho_p) = 0$.

106 It is important for the hole argument to go through that the equations always set quantities evaluated *at the same point* equal to one another: hence the restriction to *local* field equations.

flattering to Einstein. It accuses him of mistaking different coordinate representations of one and the same field configuration for physically distinct field configurations. Only the last of the four published versions of the hole argument, however, unambiguously rules out this unkind reading.

In the fourth version, Einstein explicitly added an extra step, making the argument much more interesting. If $g'_{\mu\nu}(x')$ is a solution for $T'_{\mu\nu}(x')$ in the primed coordinates, then $g'_{\mu\nu}(x)$ is a solution for $T'_{\mu\nu}(x) = T_{\mu\nu}(x)$ in the unprimed coordinates. After all, the functions $g'_{\mu\nu}(x')$ will be a solution of the local field equations no matter whether we read its arguments as primed or as unprimed coordinates. Notice that Einstein has now generated two metric fields, $g_{\mu\nu}(x)$ and $g'_{\mu\nu}(x)$, expressed in the same coordinates, that differ inside the hole, even though they are both solutions of the field equations for the same matter distribution $T_{\mu\nu}(x)$.[108] In other words, the fourth version of the hole argument explicitly involves an active point transformation rather than just a passive coordinate transformation. It was in order to avoid the kind of indeterminism that seems to come out of these active point transformations—and *not* to avoid the completely innocuous non-uniqueness of the metric field's coordinate representation—that Einstein wanted to rule out field equations that are invariant under transformations such as the one from x^μ to x'^μ used in the hole construction.

The last sentence of the passage on the hole argument in the Besso memo shows that the crucial extra step of the fourth published version of the argument was there from the very beginning. Let me quote this sentence again, this time adding some of the notation introduced above (but suppressing all indices) to bring out the point more clearly:

> If in coordinate system 1 [with coordinates x], there is a solution K_1 [i.e., $g(x)$], then this same construct [modulo a coordinate transformation] is also a solution in [coordinate system] 2 [with coordinates x'], K_2 [i.e., $g'(x')$]; K_2, however, [is] also a solution in 1 [i.e., $g'(x)$]

So already in August 1913, Einstein went through the full sequence "$g(x) \rightarrow g'(x') \rightarrow g'(x)$," which only makes sense if he wanted to argue against generally-covariant field equations on the basis of the indeterminism lurking in the inequality $g(x) \neq g'(x)$. This makes it extremely unlikely that when he subsequently published the argument he would stop at $g'(x')$ and base his argument on the trivial inequality $g(x) \neq g'(x')$. The Besso memo therefore strongly suggests that the first three published versions of the hole argument are no different from the fourth, except that they were stated much more cryptically.

I now turn to the central question of this section. Why did Einstein reject Besso's escape from the hole argument in August 1913? Given the reconstruction in sec. 3 of

107 More explicitly, for points p in the overlap of the x^μ-map and the x'^μ-map, $g_{\mu\nu}(x^\rho_p) \neq g'_{\mu\nu}(x'^\rho_p)$ (cf. note 105).

108 It is helpful to think in somewhat more intuitive terms of how Einstein's two-step procedure changes the situation inside the hole. First, the values $g_{\mu\nu}$ are transformed to new values $g'_{\mu\nu}$ for every point inside the hole. Then these new values are redistributed over the points in the hole.

Einstein's struggles with rotation during this same period, an easy answer to this question readily suggests itself. We could simply invoke Einstein's opportunistic streak again. The story would then go something like this.

In the fall of 1913, about the last thing Einstein wanted was an escape from the hole argument. His letters to Lorentz of 14 August and 16 August 1913, discussed on pp. 810–813 above, show how in two short days he went from grave concern, thinking that any acceptable field equations had to be of broad covariance, to great relief, thinking that energy-momentum conservation restricts the covariance of the field equations to general linear transformations. The hole argument dovetailed very nicely with this argument from energy-momentum conservation.[109] As a result, Einstein was probably, shall we say, less than receptive to Besso's suggestion shortly afterwards that the hole argument might not rule out generally-covariant field equations after all.[110]

Two years later, the story continues, the situation was very different. In November 1915, Einstein replaced the *Entwurf* field equations by generally-covariant ones. Yet, in the relevant publications one searches in vain for an explanation of what is wrong with the hole argument. Einstein only offered such an explanation about a month later when pressed on the issue in correspondence with Ehrenfest and Besso. Given that Einstein at this point wanted to dispose of the hole argument, one can understand why he now endorsed essentially the same escape that he had rejected two years earlier, when the conclusion of the hole argument had been more congenial to him.

The basic presumption in this reconstruction, in which expediency is the sole determining factor, is that the escape proposed in the Besso memo is very close to the escape based on the point-coincidence argument. Besso argued that only wordlines need to be determined uniquely. Einstein's minor emendation was to replace 'worldlines' by 'intersections of worldlines.' The step from one to the other, however, is perhaps not as trivial as it may appear to be at first sight. Taking my inspiration from the debate over the hole argument in the recent philosophy of space and time literature (see note 98), I will argue that it may in fact have been a very significant step. The "Of no use"-comment with which Besso's escape is rejected in the memo can then be seen as a serious objection, which can be answered only when point coincidences, i.e., intersections of wordlines, are substituted for worldlines. There is strong evidence that Einstein got the notion of point coincidences from (Kretschmann 1915), which he probably read shortly after it appeared in late 1915 (Howard and Norton 1993, 52–55). On this reconstruction then, Einstein did not have the resources in August 1913 to overcome what in the unkind glare of hindsight looks like a lame objection to a perfectly viable escape from the hole argument. Pursuing this line of thought, one arrives at an answer to the question why Einstein rejected Besso's escape from the hole argument in 1913 that does not involve any opportunism at all.

---

109 Cf., e.g., the passage from a letter to Ehrenfest of November 1913 quoted on p. 801.

110 It would not be the last time that Einstein overhastily rejected a viable escape from the hole argument. The same thing happened in the fall of 1915 in correspondence with the Göttingen mathematician Paul Hertz. For a detailed reconstruction of this episode, see (Howard and Norton 1993).

Using the language of modern differential geometry,[111] one can say that a metric field, by assigning metrical properties to all points of the manifold, "dresses up" the "bare" manifold to become a spacetime. Before the manifold acquires it spatio-temporal properties from geometrical object fields such as the metric or the affine connection, it is not really a spacetime at all. Now two solutions, $g_{\mu\nu}(x)$ and $g'_{\mu\nu}(x)$, of some generally-covariant field equations—where $g'_{\mu\nu}(x)$ is generated from $g_{\mu\nu}(x)$ through Einstein's hole construction—dress up the bare manifold differently. Suppose $g_{\mu\nu}(x)$ dresses up the bare manifold points *p* and *q* inside the hole to become the spacetime points *P* and *Q* connected by a timelike geodesic. The metric field $g'_{\mu\nu}(x)$ will (in general) dress up *p* and *q* to become two different spacetime points that (in general) will not be connected by a timelike geodesic. So, if bare manifold points can be individuated independently of the metric field, $g_{\mu\nu}(x)$ and $g'_{\mu\nu}(x)$ give different geodesics.[112] The antecedent of this conditional expresses a strong form of spacetime substantivalism. Given his Leibnizian-Machian relationist leanings, Einstein would of course have rejected such substantivalism, had he recognized it. But that recognition did not come until late 1915. So the above does provide us with a way to make sense of his comment in 1913 that Besso's escape fails "since with a solution a motion is also fully given." Since, on the tacit substantivalist assumption spelled out above, $g_{\mu\nu}(x)$ and $g'_{\mu\nu}(x)$ give different geodesics, generally-covariant field equations do not seem to determine particle trajectories uniquely, just as they do not seem to determine the metric field uniquely.

The problem disappears the moment we accept that bare manifold points cannot be individuated independently of the metric field. Both $g_{\mu\nu}(x)$ and $g'_{\mu\nu}(x)$ then turn the bare manifold into the same spacetime. It does not matter that $g_{\mu\nu}(x)$ dresses up the bare manifold points *p* and *q* to become the spacetime points *P* and *Q* connected by a timelike geodesic, whereas $g'_{\mu\nu}(x)$ will dress up some other bare manifold points *r* and *s* to become those same spacetime points *P* and *Q*. The bare manifold points *p, q, r,* and *s* only get their identity by becoming the spacetime points *P* and *Q*. The last time I checked, it was still an open question among philosophers of space and time whether this is a philosophically coherent account of identity and individuation.[113] But philosophical qualms aside, the above does seem to capture Einstein's response to the hole argument in his letters to Ehrenfest and Besso in December 1915–January 1916.[114]

111 I will phrase the argument using the kind of language that is used in discussions of the hole argument in modern philosophy of space and time. I am relying in particular on chap. 9 of (Earman 1989) on the hole argument, and on—various incarnations of—(Stachel 1993).

112 For example, $g_{\mu\nu}(x)$ and $g'_{\mu\nu}(x)$ may have a timelike geodesic that is the same in both solutions outside the hole but then splits into two different ones inside the hole.

113 See, e.g., (Earman 1989, 196–199; Butterfield 1989, 21; and Maudlin 1990, 539). The basic problem is how to strike a balance between "thisness" and "suchness" in one's philosophical account of identity. In this area, Maudlin argued, Einstein could have learned a thing or two from Aristotle. Both Earman and Butterfield turn to more recent work, viz. (Adams 1979), for help with this difficult philosophical problem.

The point-coincidence argument plays an important role in these letters. It is easy to see why. Curves in the x^μ-coordinate system representing geodesics will be different inside the hole, depending on whether one uses $g_{\mu\nu}(x)$ or $g'_{\mu\nu}(x)$ to dress up the manifold with its spatio-temporal properties. By focusing on intersections of these geodesics (i.e., on point coincidences), however, one easily sees that there will be no observable differences. There is a one-one correspondence between the curves representing the geodesics in the picture based on $g_{\mu\nu}(x)$ and the curves representing the geodesics in the picture based on $g'_{\mu\nu}(x)$. One picture can be obtained from the other by a process of continuous deformation that preserves all point coincidences. Such point coincidences, Einstein suggested, exhaust the empirical content of the theory.[115] In an often reproduced diagram in his letter to Ehrenfest of 5 January 1916, Einstein illustrated this state of affairs with an example of null-geodesics originating from a star, going through an aperture, and then hitting a photographic plate. These considerations show that the metric fields $g_{\mu\nu}(x)$ and $g'_{\mu\nu}(x)$ are empirically fully equivalent. Ontologically one still has indeterminism, but of an epistemologically totally harmless variety. If one rejects the tacit assumption from which the hole argument ultimately derives its apparent force (i.e., the assumption that bare manifold points have their identities independently of the geometrical object fields responsible for their spatio-temporal properties), the ontological indeterminism also disappears.

Given that Einstein did not have the notion of point coincidences in August 1913, one can understand why it would not have been easy for him to answer his own objection to Besso's proposed escape from the hole argument, even if he had tried. Given that the conclusion of the hole argument suited him just fine at the time, he probably did not try very hard. I therefore believe that the "Of no use"-comment in the Besso memo reflects both a genuine difficulty that Einstein saw in Besso's proposal and a certain impatience with Besso's criticism of a new and promising way to justify the limited covariance of the *Entwurf* field equations. In other words, I believe that the most plausible reconstruction of Einstein's dismissal of Besso's proposed escape from the hole argument is obtained by recognizing both the opportunistic element and the important conceptual gap between Besso's suggestion of 1913 and Einstein's resolution of the hole argument in late 1915.

For the record, I note that the notion of point coincidences is completely absent from the 1913–1914 portion of the new Besso material, whereas it is mentioned prominently on one of the pages from 1916. On the third page of the draft of his essay "The Relativity Principle in an Epistemological Formulation" (cf. pp. 790–791), Besso wrote in the margin:

---

114 See note 13 for exact references. These letters are quoted and commented on *at length* in (Stachel 1989, 86–88; Norton 1987, 168–184; and Howard 1999, 467–471). There is no need to quote them again here. It suffices to say that I subscribe to the view, shared by these three authors, that the point-coincidence argument should be seen not as signalling a retreat on Einstein's part to crude verificationism but as a vehicle for individuating spacetime points.

115 The proper length of the sections of geodesics between corresponding point coincidences will, of course, also be the same in the two pictures.

> What happens at the same place and at the same time—the spatiotemporal coincidence⟨s⟩—is the basic element of this kinematical world. 'at the same time and place' has absolute meaning; 'at the same time' or 'at the same place' alone do not.[116]

The reconstruction given above suggests that the notion of point coincidences was crucial to Einstein's resolution of the hole argument in late 1915. Kretschmann's paper, from which he presumably got this notion, must have come as a godsend. As was pointed out in (Howard and Norton 1993, 54): "What is extremely suggestive is that Kretschmann's paper appeared in an issue of the *Annalen der Physik* that was distributed on December 21, 1915, five days before the earliest of the surviving letters in which Einstein articulates the point-coincidence argument, his letter to Ehrenfest of December 26." It is true that Einstein had already made statements, both in print and in correspondence, about how his new generally-covariant theory robbed space and time of the last remnant of objective reality.[117] I think it is a mistake though to read the denial of the sort of spacetime substantivalism that gives the hole argument its apparent force into these statements.[118] They are much more naturally understood, I believe, in the light of Einstein's Machian commitments. Einstein thought that general covariance automatically meant full relativity of motion, an illusion he was disabused of sometime in 1916.[119] As late as July 1916, he still believed, for instance, that general relativity, simply by virtue of its generally-covariant field equations, finally provided the Machian account of rotation that had eluded him in the end with the *Entwurf* theory.[120] In November 1915 Einstein may have had some vague ideas about how to resolve the hole argument, but I very much doubt that he had articulated these ideas before Ehrenfest and Besso started pressing him on the issue or that he would have been able to articulate them had not Kretschmann's paper fallen into his lap at exactly the right time. If Einstein already had a fully worked-out escape from the hole argument when he replaced the *Entwurf* field equations by generally-covariant ones, it becomes hard to understand why he did not even mention the existence of such an escape in his papers of November 1915.

So Einstein published generally-covariant field equation without knowing exactly what was wrong with an argument that he had meanwhile published four times and

---

116 "Was am gleichen Orte und zur gleichen Zeit eintrifft—⟨die⟩ der zeiträumliche Coinzidenz⟨en⟩—ist das Grundelement dieser kinematischen Welt. Gleichzeitortlich hat absoluten Sinn; gleichzeitig allein, oder gleichortig allein hat keinen"

117 In the introduction of the paper on the perihelion motion presented on 18 November 1915, Einstein wrote about the assumption of general covariance "by which time and space are robbed of the last trace of objective reality" ("durch welche Zeit und Raum der letzten Spur objektiver Realität beraubt werden," Einstein 1915b, 831). In a letter to Schlick, he again wrote about general covariance that "[t]hereby time and space lose the last vestige of physical reality" ("Dadurch verlieren Zeit & Raum den letzter Rest von physikalischer Realität." Einstein to Moritz Schlick, 14 December 1915 [CPAE 8, Doc. 165]).

118 This interpretation is implied in (Stachel 1989, sec. 6).

119 See the editorial note, "The Einstein–De Sitter–Weyl–Klein Debate" (CPAE 8, 351–357), for further discussion. Cf. also note 86 above.

120 See the letter cited in note 66.

according to which generally-covariant field equations are physically unacceptable. On the face of it, this looks like another blatantly opportunistic move. Given that Einstein knew that Hilbert was hot on his trail, such opportunism would even be understandable. As in the case of his response to Besso's proposed escape in 1913, however, opportunism is not the whole story. Einstein may in fact have had good reason to believe that the hole argument had to be wrong by the time he published (Einstein 1915a), even if he could not quite put his finger yet on exactly where it went wrong.

This possibility is suggested by the close connection between the hole argument and so-called "coordinate restrictions."[121] This connection was first made by John Norton, who conjectures that the same sort of spacetime substantivalism that Einstein tacitly assumed in the hole argument was also responsible for Einstein's use of coordinate restrictions rather than coordinate conditions in the Zurich Notebook.[122] It is my belief that Einstein used coordinate restrictions in the Zurich Notebook simply because he did not yet have the modern understanding of coordinate conditions. No further explanation is needed. Consequently, I am skeptical about Norton's conjecture. It does suggest, however, that Einstein may well have realized something had to be wrong with the hole argument once he made the transition from coordinate restrictions to coordinate conditions.

A coordinate condition in the modern sense picks out (at least) one representative of each equivalence class of metric field configurations. Two metric field configurations are equivalent if they are merely different coordinate representations of what physically is one and the same field configuration. I want to emphasize two features of such coordinate conditions. First, a good coordinate condition picks a representative of each equivalence class of metric field configurations, no matter whether that field configuration is allowed by the field equations or not. It is not the job of the coordinate condition to decide which field configurations are allowed and which ones are not. That is the job of the field equations. Secondly, different coordinate conditions can be used for different problems. For example, the coordinate condition we use to show that the field equations have a sensible Newtonian limit need not be the same as the coordinate condition we use in deriving the exact solution for the case of a point mass.

The Zurich Notebook contains many examples of conditions that, at first sight, look like coordinate conditions. The role of these "coordinate conditions," however, is not clearly separated from the role of the field equations. Moreover, the freedom to apply different coordinate conditions in different contexts is not recognized. In several cases, Einstein rejected some candidate field equations because the "coordinate condition" used in showing that the equations had a sensible Newtonian limit ruled out, for example, the rotation metric. As these two observations illustrate, Einstein

121 For discussion of the difference between such coordinate restrictions and coordinate conditions in the modern sense, see also "Commentary …" (in this volume), e.g., the conclusion of sec. 5.5.4.

122 See "What was Einstein's fateful prejudice?" (in this volume).

treated these conditions as integral parts of the theories he was examining, on a par with the field equations, and not simply as a way of choosing a representation of the theory that would be convenient in a given context. The term "coordinate restrictions" was therefore introduced to distinguish these conditions from coordinate conditions in the modern sense.

In November 1915, Einstein's understanding of the role of coordinate conditions had become much closer to the modern understanding.[123] On the last page of (Einstein 1915a), Einstein first imposed the condition $\partial g^{\alpha\beta}/\partial x_\beta = 0$ to show that his new field equations had a sensible Newtonian limit and then, in the next paragraph, wrote that the theory allows transformations to rotating coordinate systems because rotations are part of the covariance group of the field equations. This is no longer true after the condition above is imposed. As Einstein presumably came to realize in the course of the research documented in the Zurich Notebook, $\partial g^{\alpha\beta}/\partial x_\beta \neq 0$ for the rotation metric. By November 1915, Einstein had apparently understood that this is not a problem at all. This suggests that he had essentially arrived at the modern understanding of coordinate conditions.

If Einstein did indeed have the modern understanding of coordinate conditions in November 1915, then he faced a very serious difficulty for the hole argument. Remember that a good coordinate condition picks out a representative of *every* equivalence class of metric field configurations. Hence, if the hole argument is valid, a good coordinate condition should *not* be selective between representations of *different* field configurations in the *same* coordinate system (such as $g_{\mu\nu}(x)$ and $g'_{\mu\nu}(x)$ related to one another through the hole construction); but it should be selective between representations of the *same* field in *different* coordinate systems (such as $g_{\mu\nu}(x)$ and $g'_{\mu\nu}(x')$). There is, however, a one-one correspondence between $g'_{\mu\nu}(x)$ and $g'_{\mu\nu}(x')$. So, if the hole argument is valid, coordinate conditions must do the impossible. Despite the one-one correspondence between $g'_{\mu\nu}(x)$ and $g'_{\mu\nu}(x')$, it must be selective between $g_{\mu\nu}(x)$ and $g'_{\mu\nu}(x')$, but not between $g_{\mu\nu}(x)$ and $g'_{\mu\nu}(x')$! Einstein may have sensed this absurdity[124] and may have concluded from it that the hole argument had to be wrong. A month or so later, he read (Kretschmann 1915) and was able to articulate exactly why the hole argument failed.

In the meantime, he had charged ahead and had published generally-covariant field equations, making it clear by doing so that he no longer believed in the hole argument but not explaining why not. There still is an element of opportunism in this. But it would have been foolish on Einstein's part to wait until he had found a fully satisfactory resolution of the hole argument. Just how tricky the hole argument is was illustrated forcefully a few years ago by the discovery of page proofs of Hilbert's first paper on general relativity (Corry et al. 1997). These page proofs suggest that even the

---

123 See "Untying the Knot" (in this volume), sec. 1.5 and sec. 5, for a reconstruction of how Einstein went from using coordinate restrictions to using coordinate conditions.

124 In the Berlin group analyzing the Zurich Notebook, we used to refer to this absurdity as "the Norton elephant" and we had lengthy discussions about whether or not Einstein swallowed the elephant.

great Göttingen mathematician originally fell hook, line, and sinker for the hole argument. There is still no consensus among philosophers of space and time exactly what the resolution of the hole argument is supposed to be. Einstein had more pressing business to take care of. He recognized that the *Entwurf* field equations were unacceptable and had to find new ones, preferably before Hilbert would. Pondering the niceties of the hole argument could be left for another occasion. Physics is not philosophy.

The upshot of this section then is that both in the 1913 chapter and in the 1915 chapter of the hole story we see a mixture of grappling with serious difficulties and a certain opportunism, the overhasty rejection of the escape from the hole argument proposed in the Besso memo in 1913, and the overhasty rejection of the hole argument itself in 1915. It is fair to say, however, that opportunism played a much smaller part in the hole story than in the rotation saga examined in sec. 3. I argued that there is an important difference between worldlines and intersections of worldlines and that the escape from the hole argument in the Besso memo of 1913 is therefore significantly different from the escape based on the point-coincidence argument of 1915. I also argued that Einstein may have recognized that the hole argument had to be wrong once he had reached the modern understanding of coordinate conditions in the fall of 1915 even though this did not tell him exactly where the argument went wrong.

5. EINSTEIN'S OPPORTUNISTIC STREAK AND THE CHARITY PRINCIPLE

In closing, I want to draw attention one more time to what I consider to be the most important feature of the story told in this paper. The story of Einstein's struggle with the problem of rotation and with the hole argument highlights what I have called, for lack of a better term, Einstein's opportunistic streak. This element is missing from current accounts of the genesis of general relativity.

In attempts to reconstruct Einstein's route to general relativity following the pathbreaking papers of John Stachel and John Norton in the 1980s (especially Stachel 1989 and Norton 1984), there has been a tendency to adopt a strong form of the so-called "charity principle." In practice, what this means is that one tries to reconstruct the development of Einstein's work starting from the strong presumption that he had good reasons for every move he made. So, to put it somewhat bluntly, whenever one encounters a passage containing what on the face of it looks like an error on Einstein's part, the strategy is to look for an interpretation in which the apparent error is the manifestation of some deep conceptual difficulty that had to be overcome before general relativity as we know it could be formulated.[125]

The two most impressive results of this approach were new much more satisfactory answers to two questions central to any reconstruction of Einstein's search for

125 In his classic 1984 paper, John Norton, for instance, explicitly stated that one of his major goals was to show that "Einstein's difficulties were based on nontrivial misconceptions and that the path he followed was a thoroughly reasonable one" (Norton 1984, 102). Certainly not all authors writing on Einstein these days adopt the charity principle. In fact, some are quite uncharitable. See, for instance, (Earman and Eisenstaedt 1999).

field equations for the metric field. The results are well known (Norton 1984, 101–102), but I nonetheless want to describe them briefly in order to bring out both the appeal and the danger of adopting the charity principle in this context. The two (pairs of) questions are the following. First, why did Einstein originally reject field equations based on the Ricci tensor and what made him eventually change his mind? Second, what exactly was Einstein's hole argument against general covariance and how did he overcome this argument in the end? In the older literature (see, e.g., Pais 1982, 221–223, 244, 251–252), the answers to these questions essentially turn on ascribing an elementary blunder to Einstein. The big stumbling block, the story went, was that Einstein did not fully appreciate that one and the same metric field can be represented in different coordinates. He initially believed that such different coordinate representations correspond to physically different field configurations. This, the story continued, had two dire consequences. First, Einstein did not recognize the freedom to apply coordinate conditions and had to reject the Ricci tensor because it apparently did not reduce to the Laplace equation of Newtonian theory in the special case of weak static fields. Secondly, it led Einstein to believe that a generally-covariant theory would be indeterministic, as illustrated in the hole argument. After the big stumbling block was removed, the story concluded, Einstein realized that the Ricci tensor was acceptable after all and that the indeterminism in the hole argument was completely illusory.

Might there not be more charitable answers to the questions under consideration? Of course, we now know there are.[126] As John Stachel first suggested, Einstein did not mistake the existence of different coordinate representations of the same metric field configuration for indeterminism, he only used these different coordinate representations to construct what look like genuinely different field configurations in one and the same coordinate system (cf. the discussion on p. 822 above). Consequently, Stachel argued, Einstein had to make an important conceptual leap to remove the apparent indeterminism revealed in the argument: he had to recognize that spacetime points cannot be individuated independently of the metric field. Similarly, we now know that Einstein's rejection of and eventual return to the Ricci tensor can be given a much more charitable interpretation. As John Norton argued, following another suggestion of John Stachel, the problem with the Ricci tensor and the Newtonian limit was not that Einstein did not know about coordinate conditions; the problem was that the relevant coordinate condition ruled out the simple form that Einstein mistakenly but quite naturally expected the metric field to take in the case of weak static fields. Once the calculations on the perihelion motion of Mercury in November 1915 made it clear that this expectation was not warranted, Einstein promptly returned to the Ricci tensor.

Both Stachel and Norton made extensive use of textual evidence, both published and unpublished, to argue for their new answers and against the old answers to the various questions at issue. A close reading of the various published versions of the

---

126 The great virtue of these new answers that is emphasized in the introductions of both (Norton 1984) and (Stachel 1989) is, in fact, that one no longer has to attribute trivial errors to Einstein.

hole argument, for instance, shows that the final version is compatible only with Stachel's more charitable interpretation of Einstein's reasoning. In more spectacular fashion, John Norton conclusively disproved the old interpretation of Einstein's problem with the Ricci tensor and the Newtonian limit by producing a page of the Zurich Notebook (p. 19L) containing a calculation formally identical to the application of a coordinate condition reducing the Ricci tensor to the Laplace equation for weak static fields. When one adds to this the enormous impact that the historical work on the hole argument has had in the philosophy of space and time literature, one readily understands why several historians of relativity, myself included, adopted the charity principle without too much critical reflection in further work on this fascinating episode.

Why have I nonetheless grown suspicious of the charity principle? For two reasons. The first is that there is a real danger that, in searching for charitable interpretations of Einstein's writings, one reads modern results back into these texts. John Norton's use of p. 19L of the Zurich Notebook to demonstrate that Einstein knew perfectly well about coordinate conditions provides a case in point, a case all the more instructive for its subtlety. The further analysis of the Zurich Notebook reported in this volume strongly suggests that Einstein did not yet have the modern understanding of coordinate conditions when he did the calculation on p. 19L after all. It now seems plausible that this lack of understanding played a much more important a role in Einstein's rejection of the Ricci tensor than his preconceptions concerning the form of the metric for weak static fields.[127]

The story told in this paper brings out a second and, at least to my mind, more serious danger of adopting the charity principle in reconstructing Einstein's route to general relativity. When one is not careful, one easily ends up with a seriously distorted view of Einstein's *modus operandi*. It is fair to say, I think, that the charity principle, as used in the tradition of Stachel and Norton, rests in part on the conviction that scientists like Einstein working on fundamentally new theories proceed in accordance with very strict standards of rationality, the kind one would expect, for instance, of a modern philosopher of science. Obviously I am in no position to spell out these standards. Fortunately I do not have to. What I hope to have shown in this paper is that at various points on his path to general relativity, Einstein made moves that can clearly and noncontroversially be recognized as blatant violations of such standards. Einstein's handling of the problem of rotation probably provides the most clear-cut example. For most of the life-span of the *Entwurf* theory, it was crucial for Einstein's goal of giving a Machian account of rotation that the rotation metric be a vacuum solution of the *Entwurf* field equations. Yet, for about two years, he never bothered to do a simple calculation with sufficient care to determine once and for all whether this was the case or not. Instead, his beliefs on this score appear to have been guided by wishful thinking, strong enough to ignore warning signs that something was wrong. In a historically accurate and well-balanced picture of how Einstein arrived at general relativity, these aspects of Einstein's work cannot be ignored.

---

127 See "Untying the Knot" (in this volume), sec. 1.1.

What is at stake here is not just a matter of historical accuracy in this particular case. Presumably, an important motivation for studying the case of Einstein and general relativity is that it may tell us something more general about the way in which new scientific theories are produced. One of the morals of this paper then is that the model that implicitly underlies the charity principle of scientists having logically cogent reasons for every move they make needs to be modified. As I suggested at the end of sec. 3, creative science may need a healthy dose of opportunism.

Even those readers who share my worries about the charity principle may feel, given the impressive results achieved with the help of it, that the dangers do not outweigh the benefits. I therefore want to make one final observation, namely that the charity principle deserves only part of the credit for these results. The most important factor behind the successes of Stachel and Norton was undoubtedly their ability and willingness to examine in unprecedented detail all relevant source material available to them. Their findings confirmed their strong suspicions that earlier historians had been too quick to attribute basic errors to Einstein. A new picture emerged in which the development of general relativity became predominantly a matter of Einstein struggling with subtle conceptual issues. Further evaluation of the source material augmented by important new archival findings such as the Besso memo discussed in this paper have modified that picture. The charity principle has become somewhat of a hindrance at this stage of the game, and we can easily do without it. That way we shall be better prepared to recognize aspects of Einstein's *modus operandi* that do not fit the model of scientific theorizing it presupposes.

NOTE ADDED IN PROOF

Since I wrote this paper, one more document that bears on the problem of rotation has come to light. A. J. Kox has alerted me to a letter that Paul Ehrenfest wrote to H. A. Lorentz in August 1913, shortly before the Besso Memo discussed in this paper.[128] The letter contains the following intriguing passage:

> It would also be nice at some point to check the calculation of the gravitational effects connected to rotational motion (constant angular velocity around a fixed axis). Five or six times Einstein has done this now—calculational errors have produced a different result almost every time: the Coriolis force came out correctly, but not the centrifugal force. To this day he still does not know whether this is merely the result of a calculational error or of a fundamental impossibility.[129]

The problem alluded to by Ehrenfest is basically the same as that recorded in the Besso Memo: the rotation metric—the Minkowski metric in rotating coordinates—is

128 Ehrenfest dated this letter 10 August 1913. The letter, however, is in response to a letter from Lorentz of 14 August 1913. Lorentz, in turn, responded to Ehrenfest's letter on 24 August 1913. Hence, the letter must have been written sometime between the 15th and the 23rd. I am grateful to A. J. Kox for alerting me to this letter, for providing me with a transcription, and for the information about its dating.

not a vacuum solution of the *Entwurf* field equations (for a more detailed exegesis, see the beginning of sec. 3).

How does the Ehrenfest letter affect the argument of my paper? First of all, it provides further evidence that the date on the Besso Memo was not just a slip on Besso's part. The possible dates for the letter, 15–23 August 1913, are remarkably close to the date on the memo, 28 August 1913. Second, the letter shows that by the middle of August Einstein was already aware of the trouble with rotation. In the paper I suggested that it was Besso who alerted him to the problem. In view of this letter, however, it is a definite possibility that Besso simply recorded the problem as explained to him by Einstein, as I conjectured he did in the case of the hole argument (see sec. 4). Third, since the letter makes it clear that Einstein had spent considerable time and effort on the problem of rotation, the question arises what, if anything, he might have been doing besides performing "five or six times" the relatively simple calculation mentioned in the Besso memo and found both in the 1913 portion of the Einstein-Besso manuscript (CPAE 4, Doc. 14, [pp. 41–42]) and on the 1915 Naumann draft (CPAE 8, Doc. 124)? The only two items I am aware of that may contain clues to answer this question are the calculation on [pp. 43–44] of the Einstein-Besso manuscript (the purpose of which remains unclear) and the calculation on the verso of the Naumann draft (see note 69). Both calculations break off before Einstein reached any definite conclusions. There is thus no evidence of a concerted effort on Einstein's part to find an escape from the problem of rotation. This then leads me to the fourth, and for the purposes of this paper, most important conclusion. If, as Ehrenfest's remarks suggest, by August 1913 Einstein had checked "five or six times" whether the rotation metric is a vacuum solution of the *Entwurf* field equations, finding "a different result almost every time," it becomes all the more astonishing that he was able to convince himself in early 1914 that the rotation metric is a solution and that it took him until September 1915 to establish that it is not. The Ehrenfest letter thus provides striking additional evidence for the opportunistic streak in Einstein's *modus operandi*.

ACKNOWLEDGMENTS

First of all, I want to express my gratitude to the Besso family for granting me permission to quote from the pages in the Besso *Nachlass* that form the basis for this paper and to reproduce two of these pages in facsimile. I am indebted to Robert Schulmann, who alerted me to the existence of this material in the late summer of 1998, and to Urs Schoepflin, for his tireless efforts to make the material available.

---

129 "Schön wäre auch noch gelegentlich die Gravitationierung der rotierenden Bewegung (constante Winkelgeschw. um feste Axe) nachzurechnen. Fünf oder sechsmal hat es Einstein gemacht — Rechenfehler haben fast jedesmal ein anderes Resultat ergeben: die Corioliskraft kam gut heraus, nicht aber die Centrifugalkraft. Er weiss noch bis heute nicht ob ein blosser Rechenfehler vorliegt oder eine wesentliche Unmöglichkeit." Paul Ehrenfest to H. A. Lorentz, before 24 August 1913 (Kox Forthcoming, Letter 278).

John Norton is largely responsible for the format of my paper by relentlessly pressing me for answers to various questions raised by accepting the date on the Besso memo. I would not have had answers to these questions, had not Robert Schulmann opened my eyes to Einstein's opportunistic streak. I also benefited from numerous discussions over the years with my co-authors of this volume, John Norton, Jürgen Renn, Tilman Sauer, and John Stachel, especially about the relation between "coordinate restrictions" and the hole argument and about the notion of "non-autonomous transformations." I would like to thank Matthias Schemmel for making me see the importance of Einstein's letter to Mach of December 1913. My former colleagues at the Einstein Papers Project, Christoph Lehner and Robert Schulmann, patiently answered requests to check this or that document, and offered many helpful suggestions. An embryonic version of this paper was presented at the Fifth International Conference on the History and Foundations of General Relativity (HGR5) held at the University of Notre Dame in June 1999. Further work on it was supported by the Max Planck Institute for the History of Science in Berlin. *Dea mea invita, opus conficere non poteram.*

REFERENCES

Adams, Robert. 1979. "Primitive Thisness and Primitive Identity." *The Journal for Philosophy 76: 5–26.*

Barbour, Julian B., and Herbert Pfister (eds.). 1995. *Mach's Principle: from Newton's Bucket to Quantum Gravity (Einstein Studies,* Vol. 6*)*. Boston: Birkhäuser.

Butterfield, Jeremy. 1989. "The Hole Truth." *British Journal for the Philosophy of Science* 40: 1–28.

Cattani, Carlo, and Michelangelo De Maria. 1989. "The 1915 Epistolary Controversy Between Einstein and Tullio Levi-Civita." In (Howard and Stachel 1989, 175–200).

Corry, Leo, Jürgen Renn, and John Stachel. 1998. "Belated Decision in the Einstein–Hilbert Priority Dispute." *Science* 278 (1997): 1270–1273.

CPAE 1: John Stachel, David C. Cassidy, Robert Schulmann, and Jürgen Renn (eds.). *The Collected Papers of Albert Einstein*. Vol.1. *The Early Years, 1879–1902*. Princeton: Princeton University Press, 1987.

CPAE 4: Martin J. Klein, A. J. Kox, Jürgen Renn, and Robert Schulmann (eds.). *The Collected Papers of Albert Einstein*. Vol. 4. *The Swiss Years: Writings, 1912–1914*. Princeton: Princeton University Press, 1995.

CPAE 5: Martin J. Klein, A. J. Kox, and Robert Schulmann (eds.). *The Collected Papers of Albert Einstein*. Vol. 5. *The Swiss Years: Correspondence, 1902–1914*. Princeton: Princeton University Press, 1993.

CPAE 6: A. J. Kox, Martin J. Klein, and Robert Schulmann (eds.). *The Collected Papers of Albert Einstein*. Vol. 6. *The Berlin Years: Writings, 1914–1917*. Princeton: Princeton University Press, 1996.

CPAE 7: Michel Janssen, Robert Schulmann, József Illy, Christoph Lehner, and Diana Kormos Buchwald (eds.). *The Collected Papers of Albert Einstein*. Vol. 7. *The Berlin Years: Writings, 1918–1921*. Princeton: Princeton University Press, 2002.

CPAE 8: Robert Schulmann, A. J. Kox, Michel Janssen, and József Illy (eds.). *The Collected Papers of Albert Einstein*. Vol. 8. *The Berlin Years: Correspondence, 1914–1918*. Princeton: Princeton University Press, 1998.

Dicke, Robert H., and H. Mark Goldenberger. 1967. "Solar Oblateness and General Relativity." *Physical Review Letters* 18: 313–316.

———. 1974. "The Oblateness of the Sun." *Astrophysical Journal Supplement Series* 27 (1974): 131–182.

Droste, Johannes. 1914. "On the Field of a Single Centre in Einstein's Theory of Gravitation." *Koninklijke Akademie van Wetenschappen te Amsterdam. Section of Sciences. Proceedings* 17 (1914–1915): 998–1011.

Earman, John. 1989. *World Enough and Space-Time*. Cambridge, MA: MIT Press.

Earman, John, and Jean Eisenstaedt. 1999. "Einstein and Singularities." *Studies in History and Philosophy of Modern Physics* 30B: 185–235.

Earman, John, and Michel Janssen. 1993. "Einstein's Explanation of the Motion of Mercury's Perihelion." In (Earman et al. 1993, 129–172).

Earman, John, Michel Janssen, and John D. Norton (eds.). 1993. *The Attraction of Gravitation: New Studies in the History of General Relativity* (*Einstein Studies,* Vol. 5). Boston: Birkhäuser.
Earman, John, and John D. Norton. 1987. "What Price Space-Time Substantivalism? The Hole Story." *British Journal for the Philosophy of Science* 38: 515–525.
Einstein, Albert. 1913a. "Zum gegenwärtigen Stande des Gravitationsproblems." *Physikalische Zeitschrift* 14: 1249–1262, (CPAE 4, Doc. 17). Reprinted as (Einstein 1914a).
———. 1913b. "Gravitationstheorie." *Schweizerische Naturforschende Gesellschaft. Verhandlungen* 96, part 2: 137–138, (CPAE 4, Doc. 15).
———. 1914a. "Zum gegenwärtigen Stande des Gravitationsproblems." In *Verhandlungen der Gesellschaft deutscher Naturforscher und Ärzte. 85. Versammlung zu Wien vom 21. bis 28. September 1913.* Part 2, sec. 1. Alexander Witting (ed.). Leipzig: Vogel, 3–24. Reprint of (Einstein 1913a).
———. 1914b. "Nachträgliche Antwort auf eine Frage von Herrn Reißner." *Physikalische Zeitschrift* 14 (1913): 108–110, (CPAE 4, Doc. 24).
———. 1914c. "Prinzipielles zur verallgemeinerten Relativitätstheorie und Gravitationstheorie." *Physikalische Zeitschrift* 15: 176–180, (CPAE 4, Doc. 25).
———. 1914d. "Physikalische Grundlagen einer Gravitationstheorie." *Naturforschende Gesellschaft in Zürich. Vierteljahrsschrift* 58: 284–290, (CPAE 4, Doc. 16).
———. 1914e. "Die formale Grundlage der allgemeinen Relativitätstheorie." *Königlich Preußische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte*: 1030–1085, (CPAE 6, Doc. 9)
———. 1915a. "Zur allgemeinen Relativitätstheorie." *Königlich Preußische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte*: 778–786, (CPAE 6, Doc. 21).
———. 1915b. "Erklärung der Perihelbewegung des Merkur aus der allgemeinen Relativitätstheorie." *Königlich Preußische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte*: 831–839, (CPAE 6, Doc. 23).
———. 1916. "Die Grundlage der allgemeinen Relativitätstheorie." *Annalen der Physik* 49: 769–822, (CPAE 6, Doc. 30)
———. 1918. "Prinzipielles zur allgemeinen Relativitätstheorie." *Königlich Preußische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte*: 831–839, (CPAE 7, Doc. 4)
Einstein, Albert, and Marcel Grossmann. 1913. *Entwurf einer verallgemeinerten Relativitätstheorie und einer Theorie der Gravitation*. Leipzig: Teubner, (CPAE 4, Doc. 13). Reprinted with additional "Comments" ("Bemerkungen") (CPAE 4, Doc. 26) as (Einstein and Grossmann 1914a).
———. 1914a. "Entwurf einer verallgemeinerten Relativitätstheorie und einer Theorie der Gravitation." *Zeitschrift für Mathematik und Physik* 62 (1914): 225–259. Reprint of *Einstein and Grossmann 1913* with additional "Comments" ("Bemerkungen") (CPAE 4, Doc. 26).
———. 1914b. "Kovarianzeigenschaften der Feldgleichungen der auf die verallgemeinerte Relativitätstheorie gegründeten Gravitationstheorie." *Zeitschrift für Mathematik und Physik* 63: 215–225, (CPAE 6, Doc. 21)
Goenner, Hubert, Jürgen Renn, Jim Ritter, and Tilman Sauer (eds.). 1999. *The Expanding Worlds of General Relativity* (*Einstein Studies*, Vol. 7). Boston: Birkhäuser.
Hall, Asaph. 1894. "A Suggestion in the Theory of Mercury." *Astrophysical Journal* 14: 49–51.
Hoefer, Carl. 1994. "Einstein's Struggle for a Machian Gravitation Theory." *Studies in History and Philosophy of Science* 25: 287–335.
Howard, Don. 1999. "Point Coincidences and Pointer Coincidences: Einstein on the Invariant Content of Space-Time Theories." In (Goenner et al. 1999, 463–500).
Howard, Don, and John D. Norton. 1993. "Out of the Labyrinth? Einstein, Hertz, and the Göttingen Response to the Hole Argument." In (Earman et al. 1993, 30–62).
Howard, Don, and John Stachel (eds.). 1989. *Einstein and the History of General Relativity* (*Einstein Studies,* Vol. 1). Boston: Birkhäuser.
Janssen, Michel. 1999. "Rotation as the Nemesis of Einstein's *Entwurf* Theory." In (Goenner et al. 1999, 127–157).
Kennefick, Daniel. 2005. "Einstein and the Problem of Motion: A Small Clue." In (Kox and Eisenstaedt 2005, 109–124).
Kox, A. J. 1988. "Hendrik Antoon Lorentz, the Ether, and the General Theory of Relativity." *Archive for History of Exact Sciences* 38: 67–78. Reprinted in (Howard and Stachel 1989, 201–212).
——— (ed.). Forthcoming. *Hendrik Antoon Lorentz, Correspondence with Paul Ehrenfest*. New York: Springer.
Kox, A. J., and Jean Eisenstaedt (eds.). 2005. *The Universe of General Relativity* (*Einstein Studies*, Vol. 11). Boston: Birkhäuser.
Kretschmann, Erich. 1915. "Über die prinzipielle Bestimmbarkeit der berechtigten Bezugssysteme beliebiger Relativitätstheorien." *Annalen der Physik* 48: 907–942, 943–982.
Maudlin, Tim. 1990. "Substances and Space-Time: What Aristotle Would Have Said to Einstein." *Studies in History and Philosophy of Science* 21: 531–561.

Newcomb, Simon. 1895. *The Elements of the Four Inner Planets and the Fundamental Constants of Astronomy. Supplement to the American Ephemeris and Nautical Almanac for 1897*. Washington, D. C.: Government Printing Office.

Norton, John D. 1984. “How Einstein Found his Field Equations, 1912–1915.” *Historical Studies in the Physical Sciences* 14: 253–316. Reprinted in (Howard and Stachel 1989, 101–159). Page references are to this reprint.

———. 1987. “Einstein, the Hole Argument and the Reality of Space.” In J. Forge (ed.), *Measurement, Realism, and Objectivity*. Dordrecht: Reidel, 153–188.

Pais, Abraham. 1982. *‘Subtle is the Lord…’: The Science and the Life of Albert Einstein*. Oxford: Clarendon Press; New York: Oxford University Press.

Renn, Jürgen (ed.). 2005. *Dokumente eines Lebensweges/Documents of a Life’s Pathway*. Weinheim: Wiley–VCH.

Speziali, Pierre (transl. and ed.). 1972. *Albert Einstein/Michele Besso, Correspondance, 1903–1955*. Paris: Hermann.

Stachel, John. 1989. “Einstein’s Search for General Covariance, 1912–1915.” (Based on a talk given at the Ninth International Conference on General Relativity and Gravitation in Jena, 1980.) In (Howard and Stachel 1989, 62–100).

———. 1993. “The Meaning of General Covariance: the Hole Story.” In John Earman, Allen I. Janis, Gerald J. Massey, and Nicholas Rescher (eds), *Philosophical Problems of the Internal and External World: Essays on the Philosophy of Adolf Grünbaum*. Konstanz: Universitätsverlag/Pittsburgh: University of Pittsburgh Press, 129–160.

MICHEL JANSSEN AND JÜRGEN RENN

UNTYING THE KNOT: HOW EINSTEIN FOUND HIS WAY BACK TO FIELD EQUATIONS DISCARDED IN THE ZURICH NOTEBOOK

"She bent down to tie the laces of my shoes. Tangled up in blue."
—Bob Dylan

1. INTRODUCTION: NEW ANSWERS TO OLD QUESTIONS

Sometimes the most obvious questions are the most fruitful ones. The Zurich Notebook is a case in point. The notebook shows that Einstein already considered the field equations of general relativity about three years before he published them in November 1915. In the spring of 1913 he settled on different equations, known as the "*Entwurf*" field equations after the title of the paper in which they were first published (Einstein and Grossmann 1913). By Einstein's own lights, this move compromised one of the fundamental principles of his theory, the extension of the principle of relativity from uniform to arbitrary motion. Einstein had sought to implement this principle by constructing field equations out of generally-covariant expressions.[1] The *Entwurf* field equations are not generally covariant. When Einstein published the equations, only their covariance under general linear transformations was assured. This raises two obvious questions. Why did Einstein reject equations of much broader covariance in 1912-1913? And why did he return to them in November 1915?

1 Throughout the period covered by this paper, Einstein thought that general covariance automatically extends the principle of relativity from uniform to arbitrary motion. In part, this was because he did not distinguish carefully, for reasons laid out in (Norton 1999), between the roles of Lorentz invariance in special relativity and of general covariance in general relativity. In part, as is argued in (Janssen 2005), it was just a matter of misleading terminology. Einstein chose to describe a key feature of his new theory—i.e., observers in arbitrary motion with respect to one another can both maintain to be at rest as long as they agree to disagree about whether or not there is a gravitational field—in terms of general relativity of motion, whereas what is relative is not so much the motion but the split of inertio-gravitational effects into inertial and gravitational components. As Einstein put it in 1920: "Like an induced electric field, the gravitational field at a particular point only has a relative existence" ("Das Gravitationsfeld hat an einem betrachteten [Punkte] in ähnlicher Weise nur eine relative Existenz wie das durch magnetelektrische Induktion erzeugte elektrische Feld." CPAE 7, Doc. 31, [p. 21]).

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 2
Einstein's Zurich Notebook: Commentary and Essays.

A new answer to the first question has emerged from the analysis of the Zurich Notebook presented in this volume. This calls for a reassessment of Einstein's subsequent elaboration of the *Entwurf* theory and of the transition to the theory of November 1915. On the basis of a reexamination of Einstein's papers and correspondence of this period, we propose a new answer to the second question.

For the discussion of these matters, it is important to distinguish between two strategies for finding suitable gravitational field equations, a 'physical strategy' and a 'mathematical strategy'. Following the physical strategy, one constructs field equations in analogy with Maxwell's equations, making sure from the start that energy-momentum conservation is satisfied and that the Poisson equation of Newtonian theory is recovered in the case of weak static fields. This is the approach that originally led Einstein to the *Entwurf* field equations. Following the mathematical strategy, one picks candidate field equations based largely on considerations of mathematical elegance and then investigates whether they make sense from a physical point of view.[2] With hindsight, one easily recognizes that the latter approach provides a royal road to the generally-covariant field equations of November 1915. Einstein himself used a combination of the two strategies. In the Zurich Notebook, he tried the mathematical strategy first, ran into what appeared to be insurmountable difficulties, switched to the physical strategy, and ended up with the *Entwurf* field equations. On that much all scholars working in this area agree. The question is what Einstein did in late 1915. The currently standard answer is that he abandoned the physical strategy, went back to the mathematical strategy prematurely abandoned in the Zurich Notebook, and in short order produced the happy results of November 1915.[3] With very few exceptions, Einstein's pronouncements—both at the time and in retrospect years later—fit very well with this answer.

As the title of our paper suggests, however, we see no abrupt change of strategy in 1915. Our metaphor is not "cutting the knot" but "untying the knot." We argue that Einstein found the field equations of general relativity by changing one element in a formalism he had developed in 1914 encoding the various physical considerations that had gone into the derivation of the *Entwurf* field equations. He picked a new mathematical object, known as the Christoffel symbols, to represent the gravitational field. This one modification, it turned out, untangled the knot of conditions and definitions that his theory had become in 1914–1915. We thus argue that the field equations of general relativity were the fruit of Einstein's relentless pursuit of the physical strategy. That is not to say that the mathematical strategy did not play any role at all. Without it Einstein would not have recognized that his new definition of the gravitational field was the key to the solution of his problem of finding suitable field equations. What happened in 1915 was that the physical strategy led Einstein back to field

2 For more careful discussion of the distinction between the 'mathematical strategy' and the 'physical strategy', see secs. 1.1 and 5.1 of "Commentary …" (in this volume).

3 See, e.g., (Norton 1984, 142), (Janssen 1999, 151), (Van Dongen 2002, 30). The most explicit version of this account is given in (Norton 2000). We shall have occasion to quote some typical passages from this paper in sec. 10.

equations to which the mathematical strategy had already led him in the Zurich Notebook but which he had then been forced to reject since he could not find a satisfactory physical interpretation for them. With two routes to the same field equations, Einstein had the luxury of a choice in how to present them to the Berlin Academy. He went with the mathematical considerations, which he turned into a simple and effective argument for his new field equations. It would have been much more complicated and less persuasive to opt for an exposition faithful to the arduous journey he himself had been forced to undertake only to discover in the end that equations he had considered very early on were the right ones after all.[4] In the context of discovery, the physical argument had been primary and the role of the mathematical argument had been to reinforce that argument. In the context of justification, it was just the other way around. Einstein gave pride of place to the mathematical argument and used elements from his physical argument only to show that his new field equations were perfectly acceptable on physical grounds. Once the mathematical route to the field equations had been reified in his first communication to the Berlin Academy of November 1915, the physical route rapidly faded from memory. The streamlined argument of the context of justification quickly supplanted the messy reasoning of the context of discovery. Einstein succumbed to a typical case of selective amnesia. Before long he had eyes only for the mathematical strand in his reasoning and had lost sight of the physical strand altogether.[5]

In the remainder of this introduction, we give an outline, as non-technical as possible, of our new understanding of the path that took Einstein away from generally-covariant field equations and back again in the period 1912–1915. The emphasis will be on the second part of this fascinating tale. The case for our new reconstruction is strong but largely circumstantial. We shall highlight the most important pieces of evidence in the introduction, so that the reader can judge for him- or herself how well our reconstruction is supported by the documents without having to go through the detailed calculations that make up the balance of the paper.[6]

1.1 Tying the Knot: Coordinate Restrictions

The central problem frustrating Einstein's search for generally-covariant field equations in the Zurich Notebook was his peculiar use of what a modern relativist would immediately recognize as coordinate conditions. One needs to impose such conditions, which the metric tensor has to satisfy in addition to the field equations, if one wants to compare equations of general relativity, which are valid in arbitrary coordinates, to equations in Newtonian gravitational theory, which in their standard form

4 A paper written by one of us (JR) which we have cannibalized for this paper was therefore called "Progress in a Loop."

5 Jeroen van Dongen (2002, 46–47) has emphasized that Einstein's selective memory only served him all too well in his later years in his defense of relying on a purely mathematical strategy in the search for a unified field theory.

6 For a short version of our account, see (Janssen 2005, 75–82).

are valid only in inertial frames. In particular, one needs a coordinate condition to show that the relevant component of the generally-covariant field equations of general relativity reduces to the Poisson equation of Newtonian theory in the case of weak static fields (see, e.g., Wald 1984, 75–77).[7] Such conditions are not essential to the theory. One can pick whatever coordinate condition is most convenient for the problem at hand. From a modern point of view this is trivial and there would be no point in spelling it out, if it were not for the fact that Einstein's use of such additional conditions both in the Zurich Notebook and in his subsequent elaboration of the *Entwurf* theory deviated sharply from our modern use.

As Einstein was examining various generally-covariant expressions in 1912-1913 to determine whether physically acceptable field equations could be extracted from them, he assumed that he needed additional conditions not just to recover the Poisson equation for weak static fields but also to guarantee that the equations be compatible with the law of energy-momentum conservation.[8] In general relativity energy-momentum conservation is a direct consequence of the general covariance of the Einstein field equations (see Einstein 1916c and sec. 9). This result is an instantiation of one of Emmy Noether's celebrated theorems connecting symmetries and conservation laws (Noether 1918).[9] In 1912–1913, however, Einstein thought that energy-momentum conservation required that the covariance of the field equations be restricted. In the Zurich Notebook he did not make a clear distinction between conditions imposed to guarantee energy-momentum conservation and conditions imposed to recover the Poisson equation for weak static fields. On the contrary, once he had found a condition that accomplished the latter, he would investigate what further conditions, if any, were needed for the former.[10]

Einstein used these conditions to eliminate various terms from equations of broad covariance and looked upon the truncated equations of severely restricted covariance rather than upon the equations of broad covariance he started from as candidates for the fundamental field equations of his theory. Since coordinate conditions used in this manner are ubiquitous in the Zurich Notebook we introduced a special name for them. We call them *coordinate restrictions*.[11]

---

7 Einstein originally imposed the stronger requirement that this would be the only non-trivial component of the field equations in the case of weak static fields. This was because he assumed that the metric for a weak static field must be spatially flat. It was only in November 1915 that he came to realize that this is not necessary (see p. 891 below).

8 This is a lesson Einstein had learned the hard way earlier in 1912 when he had been forced to modify the field equations of his theory for static gravitational fields because the original equations violated energy-momentum conservation (Einstein 1912, sec. 4). For further discussion, see "Pathways out of Classical Physics …" (in vol. 1 of this series).

9 For careful discussion of Noether's theorems and some simple but informative applications of them, see (Brading 2002); for a discussion of how they emerged from the discussion of general relativity in Göttingen, see (Rowe 1999), (Sauer 1999), and "Hilbert's Foundation of Physics …" (in vol. 4 of this series). For a concise discussion of Einstein's ideas about energy-momentum conservation in the period 1912–1918, see sec. VIII of the introduction to CPAE 8.

10 This approach is clearly in evidence on pp. 19L–20L of the Zurich Notebook.

This notion is the key to understanding why Einstein did not publish field equations based on the Riemann tensor in 1913. He recognized that the Riemann tensor—or rather the Ricci tensor, a direct descendant of it—was the natural starting point for finding field equations, but he had great difficulty finding coordinate restrictions that would guarantee compatibility with energy-momentum conservation even in first approximation. Moreover, none of the coordinate restrictions with which he could recover the Poisson equation for weak static fields left him enough covariance to implement the equivalence principle and the generalized principle of relativity.[12]

So towards the end of the notes on gravity in the Zurich Notebook, Einstein switched from the mathematical to the physical strategy. Instead of starting from a mathematical object such as the Ricci tensor with well-defined covariance properties, he now started from the physical requirements that the Poisson equation be recovered for weak static fields and that energy-momentum conservation be satisfied. Instead of demanding broad covariance, he only demanded covariance under general linear transformations.[13] From these requirements he derived the equations that would serve as the fundamental field equations of his theory without bothering to find the generally-covariant equations of which these equations would be the truncated version or the coordinate restriction with which to do the truncating. Einstein convinced himself that this procedure led to a unique result: the *Entwurf* field equations.

---

11 See secs. 4.1 of "Commentary …" (in this volume). There is no agreement among the authors of this volume as to why Einstein used coordinate restrictions. The majority view is that Einstein at the time did not yet have the modern understanding of coordinate conditions. John Norton, however, argues that Einstein did have the modern understanding all along and offers a different explanation for why he nonetheless chose to use coordinate restrictions instead (see "What was Einstein's 'Fateful Prejudice'?" [in this volume]). The story we tell in this paper is compatible with both views. This is an indication of how difficult it is to decide between them. John Norton argues that it boils down to one's view of Einstein's *modus operandi*. For the record, we share the view presented in "What Did Einstein Know …" (in this volume).

12 John Norton (1984, 102, 111–112, 142-143) argued that the incompatibility of the harmonic coordinate condition with the spatially flat metric that Einstein thought should describe weak static fields plays a crucial role both in Einstein's rejection of field equations based on the Ricci tensor in 1912–1913 and in his choice of new field equations in the first paper of November 1915 (Einstein 1915a). We seriously doubt whether Einstein was even aware of this incompatibility either at the time of the Zurich Notebook or in November 1915. We see no evidence that this incompatibility played any role in Einstein's search for gravitational field equations (cf. sec. 5.4 of "Commentary …" [in this volume] and the conclusion of sec. 5 below). Einstein's prejudice about the form of the metric for weak static field, for which there is abundant textual evidence, did play a role—as Norton (1984, 146–148) also emphasized—in that it was incompatible with field equations containing a term with the trace of the energy-momentum tensor of matter. Aside from general covariance, this trace term is the most important feature distinguishing the Einstein field equations from their *Entwurf* counterpart (see sec. 7 and the appendix).

13 Einstein's hope was that the equations would also be invariant under what he later called "non-autonomous" transformations to accelerating frames of reference. See sec. 3.3 below for discussion of the concept of non-autonomous transformations.

1.2 Tightening the Noose: Covariance Properties of the Entwurf Field Equations

Einstein's further elaboration of the *Entwurf* theory in 1913–1914 centered on clarifying the covariance properties of the *Entwurf* field equations. In August 1913, he produced an argument purporting to show that because of energy-momentum conservation the equations' covariance group had to be limited to general linear transformations. This argument was published in an addendum to the journal version of the *Entwurf* paper (Einstein and Grossmann 1914a). Within a few months, Einstein realized that it was based on a faulty premise.[14] The idea that energy-momentum conservation circumscribes the covariance of the field equations nonetheless survived. In the same paper in which he retracted his fallacious argument (Einstein and Grossmann 1914b), he and Grossmann presented a new argument tying the covariance of the *Entwurf* field equations to energy-momentum conservation. With Noether's theorems still four years into the future, Einstein's intuition that energy-momentum conservation is closely related to the covariance of the field equations is quite remarkable. It will play a crucial role in our story.

Einstein and Grossmann (1914b) found four conditions, compactly written as $B_\mu = 0$, that in conjunction with the *Entwurf* field equations imply energy-momentum conservation. They then used a variational formalism to show that these same conditions determine the covariance properties of the *Entwurf* field equations. Einstein thought that these conditions were the coordinate restriction with which the *Entwurf* field equations could be extracted from generally-covariant equations.[15] He had no interest in finding the latter, since his infamous 'hole argument' had meanwhile convinced him that the field equations could not possibly be generally covariant.[16] In fact, Einstein and Grossmann claimed that the four conditions they had found gave the field equations the maximum covariance allowed by the hole argument. With these results, the theory appeared to have reached its definitive form.

In the spring of 1914, Einstein left Zurich and Grossmann and moved to Berlin. In October 1914, nearly three months into the Great War, he completed a lengthy review article on his new theory, no longer called "a general*ized* theory of relativity" (Einstein and Grossmann 1913) but "the *general* theory of relativity" (Einstein 1914c). In this article, he reiterated and tried to improve on the results of his second paper with Grossmann. He now used the variational formalism to deal both with the covariance properties of the field equations and with energy-momentum conservation. And he did so without specifying the Lagrangian ahead of time as he had done in the paper with Grossmann. He only assumed that the Lagrangian transforms as a scalar under

---

14 For discussion of this episode, see (Norton 1984, sec. 5), "Pathways out of Classical Physics …" (in vol. 1 of this series), and sec. 2 of "What Did Einstein Know …" (in this volume).

15 Einstein (1914b, 178) believed that there is a corresponding generally-covariant equation for any physically meaningful equation that is not. See note 57 below for the relevant passage.

16 For a discussion of (the origin of) the hole argument, see sec. 4 of "What Did Einstein Know …" (in this volume).

general linear transformations. He found that a generic version of the set of conditions he had found with Grossmann, still written as $B_{\mu} = 0$, is necessary both for the covariance of the field equations and for their compatibility with energy-momentum conservation. Energy-momentum conservation, however, called for an additional set of conditions, compactly written as $S_{\sigma}^{\nu} = 0$. Einstein believed that these extra conditions uniquely picked out the Lagrangian giving the *Entwurf* field equations. As a matter of fact they do no such thing.

1.3 At the End of His Rope: The Demise of the Entwurf Field Equations

In early 1915, the Italian mathematician Tullio Levi-Civita contested some of the results of Einstein's review article but Einstein did not give ground.[17] Curiously, Levi-Civita did not take aim at Einstein's uniqueness argument, even though his interest in Einstein's article had been triggered by a letter from Max Abraham complaining about the arbitrariness of Einstein's choice of the Lagrangian (Cattani and De Maria 1989, 185). It was not until October 1915 that Einstein himself realized that his uniqueness argument was illusory. This setback came hard on the heels of another one. He had discovered that the *Entwurf* field equations are incompatible with one of the guiding ideas of the theory—the idea that the inertial forces of rotation can be conceived of as gravitational forces. Michele Besso had already put his finger on this problem two years earlier, but Einstein had ignored his friend's warnings.[18] He finally faced up to the problem in September 1915.

In a letter to H. A. Lorentz of October 12, 1915 (CPAE 8, Doc. 129), Einstein explained where his uniqueness argument went wrong. The extra conditions $S_{\sigma}^{\nu} = 0$ that he had used to determine the Lagrangian are trivially satisfied by any Lagrangian invariant under general linear transformations. So both sets of conditions—$B_{\mu} = 0$ and $S_{\sigma}^{\nu} = 0$—needed for energy-momentum conservation also emerge from the analysis of the theory's covariance properties. From a modern point of view, this is just an instance of one of Noether's theorems. If one sets the Lagrangian in Einstein's variational formalism equal to the Riemann curvature scalar, as Einstein (1916c) himself would do the following year, the four conditions $B_{\mu} = 0$ become the contracted Bianchi identities.

1.4 Pulling a Thread: from the Entwurf Field Equations to the November Tensor and the Einstein Field Equations

Despite the problem of rotation and the evaporation of the uniqueness argument, Einstein was not ready to part with the *Entwurf* field equations just yet. He told Lorentz that they are still the only equations that allow one to recover the Poisson equation for weak static fields. Just a few weeks later, however, on November 4, 1915, he submit-

17 See the correspondence between Einstein and Levi-Civita in March–May 1915 in (CPAE 8).

18 For discussion of Einstein's struggles with the problem of rotation in 1913–1915, see sec. 3 of "What Did Einstein Know …" (in this volume) and (Janssen 2005, 68–71).

ted a short paper to the Berlin Academy in which he replaced the *Entwurf* field equations by equations based on the Riemann tensor. He had examined and rejected these exact same equations three years earlier in the Zurich Notebook. Three more short communications to the academy followed in rapid succession, two of them with further modifications of the field equations (Einstein 1915b, d) and one on the perihelion motion of Mercury (Einstein 1915c). By the end of November, Einstein had thus arrived at the generally-covariant field equations that still bear his name and he had solved an outstanding puzzle in planetary astronomy.

What happened those last few weeks of October? Einstein has left us some tantalizing clues. In the first November paper, he singled out one element and called it "a fateful prejudice."[19] In a letter written later that month, shortly after the dust had settled, he wrote that changing that one element had been "the key to [the] solution."[20] The element in question is the definition of the components of the gravitational field. In the *Entwurf* theory, they are essentially just the derivatives of components of the metric field. This is the straightforward generalization of the definition of the gravitational field in Newtonian theory as the gradient of the gravitational potential. In Einstein's theory the components of the metric field play the role of the gravitational potentials. In the final version of the theory, the gravitational field is represented by the so-called Christoffel symbols. The Christoffel symbols consist of a sum of three terms with derivatives of the metric. These objects play an important role in Riemannian geometry. They also occur in the geodesic equation, which makes them the natural candidates for representing the gravitational field. Surprisingly from a modern point of view, the first November paper is the first place where Einstein actually makes this observation. Why did he put so much emphasis all of a sudden on the definition of the gravitational field?[21]

In the *Entwurf* theory, both the field equations and the equation for energy-momentum conservation were originally formulated in terms of the metric, the quantity representing the gravitational potentials, not in terms of the quantity representing the gravitational field.[22] Einstein, however, also tried to write both equations in terms of the field. In this form, the analogy between the *Entwurf* theory and electrodynam-

19 "ein verhängnisvolles Vorurteil" (Einstein 1915a, 782). Cf. note 38 below.

20 "Den Schlüssel zu dieser Lösung …" Einstein to Arnold Sommerfeld, 28 November 1915 (CPAE 8, Doc. 153). Cf. note 39 below.

21 Without the analysis of the Zurich Notebook presented in this volume, Einstein's remarks about the definition of the gravitational field have, as John Norton (1984, 145) put it, "all the flavor of an after-the-fact rationalization." Norton was also right in that these comments do not help us understand why Einstein turned his back on equations extracted from the Riemann tensor in 1913.

22 As John Stachel points out in "The Story of Newstein …" (in vol. 4 of this series), this is in part because Einstein had to make do with the mathematics available to him. Far from providing all the tools he needed, differential geometry at the time still lacked the concept of an affine connection, which is a much more natural object than the metric to describe the inertio-gravitational field of general relativity. The absence of the notion of parallel displacement and the concept of an affine connection also tripped up H. A. Lorentz in 1916 when he tried to give a coordinate-free formulation of general relativity (Janssen 1992).

ics, which Einstein had consciously pursued in constructing the theory,[23] was brought out more clearly. This in turn made the physical meaning of the equations more perspicuous.

In August 1913, Einstein had found that the *Entwurf* field equations can be written in the form of Maxwell's equations, with the four-divergence of the field on the left-hand side and the field's sources—the sum of the energy-momentum densities of matter and gravitational field[24]—on the right-hand side (Einstein 1913, 1258, eq. 7b). He had written the equations in this form ever since.[25] In the Zurich Notebook he had already noticed that the term representing the gravitational force density in the energy-momentum balance equation can be interpreted as an inner product of the field and its sources, just like the Lorentz force that some extended charge distribution experiences from its self-field. Most importantly, in his review article of 1914, Einstein wrote the Lagrangian for the *Entwurf* field equations in terms of the components of the gravitational field (Einstein 1914c, 1076, note 1). The *Entwurf* Lagrangian is the same quadratic expression in the field as the Lagrangian for the free Maxwell field. These structural similarities to electrodynamics—in the field equations and in the expressions for the force density and the Lagrangian—carry over from the *Entwurf* theory to the theory of the November 1915 papers if the components of the gravitational field are redefined as the Christoffel symbols.[26]

In view of this continuity, Einstein's remark that the new definition of the gravitational field was "the key to the solution" suggests a natural pathway along which, sometime during the second half of October 1915, Einstein found his way back to field equations of broad covariance discarded three years earlier in the Zurich Notebook.

Not long after his letter to Lorentz of October 12, Einstein must have come to accept that the problem of rotation was the nemesis of the *Entwurf* field equations (Janssen 1999). He needed new field equations or rather a new Lagrangian from which such new equations could be derived. His variational formalism would give him the conditions to guarantee compatibility with energy-momentum conservation.

---

23 For further discussion of the role of this analogy, see "Pathways out of Classical Physics …" (in vol. 1 of this series).

24 Because of the equivalence of energy and mass (inertial and gravitational) it is clear that the gravitational field contributes to its own source. This, of course, is a major disanalogy between the gravitational field equations and Maxwell's equations. For one thing, unlike Maxwell's equations, the gravitational field equations will not be linear in the components of the field.

25 See, e.g., (Einstein 1914a, 289, eq. 5), (Einstein 1914b, 179, eq. 6), (Einstein and Grossmann 1914b, 217, eq. II), and (Einstein 1914c, 1077, eq. 81). For the original form of the *Entwurf* field equations, see (Einstein and Grossmann 1913, 15–17, eqs. 13–16, 18, and 21).

26 Both in the *Entwurf* theory and in the theory of the first November paper, the Lagrangian has the form $g^{\mu\nu}\Gamma^{\alpha}_{\beta\mu}\Gamma^{\beta}_{\alpha\nu}$, where $\Gamma^{\beta}_{\alpha\nu}$ are the components of the gravitational field. In the *Entwurf* theory, $\Gamma^{\beta}_{\alpha\nu} \equiv \frac{1}{2}g^{\beta\tau}g_{\tau\alpha,\nu}$, in the first November paper, $\Gamma^{\beta}_{\alpha\nu} \equiv -\begin{Bmatrix} \beta \\ \alpha\nu \end{Bmatrix} \equiv -\frac{1}{2}g^{\beta\tau}(g_{\tau\alpha,\nu} + g_{\tau\nu,\alpha} - g_{\alpha\nu,\tau})$. The Lagrangian is modelled on the Lagrangian for the free Maxwell field, $-\frac{1}{4}F_{\mu\nu}F^{\mu\nu}$, where $F_{\mu\nu}$ and $F^{\mu\nu}$ are the covariant and contravariant components of the electromagnetic field, respectively.

The analogy with electrodynamics could be used to narrow the range of plausible candidates for the new Lagrangian. Replacing the derivatives of the metric field by the Christoffel symbols as components of the gravitational field in the *Entwurf* Lagrangian may well have been one of the first things he tried.

The expression for the left-hand side of the field equations that one finds upon feeding this new Lagrangian into the variational formalism bears a striking resemblance to the left-hand side of field equations that Einstein had extracted from the Ricci tensor in the Zurich Notebook by imposing the (relatively weak) restriction to unimodular transformations. These are transformations with a Jacobian equal to one, or, equivalently, transformations under which the determinant g of the metric transforms as a scalar. Imposing the restriction to unimodular transformations on the general variational formalism, which allows one to omit a factor of $\sqrt{-g}$ in the action, and feeding the new Lagrangian into this version of the formalism, one finds that the left-hand side of the resulting field equations is exactly the same as the left-hand side of field equations discarded in the Zurich Notebook. Because of its reappearance in November 1915, we call this expression the *November tensor*.[27]

By adjusting the physical reasoning that had gone into the derivation of the *Entwurf* field equations, Einstein had thus found new field equations that could also be derived along the lines of the mathematical strategy. This was exactly the sort of convergence of physical and mathematical considerations that had eluded Einstein in the Zurich Notebook and in his work on the *Entwurf* theory. The best he had been able to do was to convince himself in 1914 that the *Entwurf* field equations can at least in principle be extracted from generally-covariant ones with the help of the coordinate restriction $B_\mu = 0$. Now physical and mathematical considerations both pointed to the November tensor. He set the November tensor equal to the energy-momentum tensor for matter, and confidently replaced the *Entwurf* field equations by these new equations in his first communication to the Berlin Academy of November 1915.

In the Zurich Notebook, Einstein had not been able to prove compatibility of field equations based on the November tensor with energy-momentum conservation. His variational formalism, even though it had to be used with caution because of the restriction to unimodular transformations, now provided all the guidance he needed to solve that problem. This unexpected windfall, however, brought a new puzzle. Having caught on to the connection between covariance and conservation laws, Einstein had come to expect that the covariance of the field equations was determined by the four conditions $B_\mu = 0$ in his variational formalism that at the same time guarantee energy-momentum conservation. The covariance of the November tensor, however, is much broader than these conditions would seem to allow. What did Einstein make of this apparent mismatch between covariance and conservation laws? The November 1915 papers again provide some important clues.

In the first of these papers (Einstein 1915a, 785), Einstein rewrote the four conditions $B_\mu = 0$ in such a way that they can be replaced by one stronger condition. He

27 See sec. 5.5 of "Commentary …" (in this volume).

then showed that this stronger condition can be replaced by the requirement that the determinant of the metric not be a constant. In the second and in the fourth paper, Einstein proposed ways to circumvent this requirement. These moves become readily understandable if we assume that they were made in response to the discrepancy between covariance and conservation laws mentioned above. Given that the November tensor is invariant under arbitrary unimodular transformations, Einstein expected that energy-momentum conservation would not require any further restrictions. As we mentioned above, g, the determinant of the metric, transforms as a scalar under unimodular transformations. This explains why Einstein tried to rewrite the standard four conditions $B_\mu = 0$ giving energy-momentum conservation as one condition on g. It also explains why he was not satisfied with the requirement that g not be a constant. The restriction to unimodular *transformations* only requires g to transform as a scalar, not that it be either a constant or a variable. In fact, it turns out to be advantageous to impose the stronger restriction to unimodular *coordinates*, i.e., coordinates in which $g = -1$. It is thus perfectly understandable that Einstein tried to replace the condition that g not be a constant by the condition that $g = -1$. This was the driving force behind the transition from the field equations of the first November paper to those of the fourth one. In this last paper of November 1915, Einstein showed that one arrives at the desired condition $g = -1$ if a term involving the trace of the energy-momentum tensor is added to the field equations of the first November paper. These equations can be looked upon as generally-covariant equations expressed in terms of unimodular coordinates. The generally-covariant equations are the Einstein field equations.

1.5 Untying the Knot: Coordinate Conditions

Two problems that had defeated the November tensor in the Zurich Notebook still need to be addressed. How did Einstein recover the Poisson equation for weak static fields and how did he show that his new field equations allow Minkowski spacetime in rotating coordinates? These are two separate problems and they are easily solved separately, but in the Zurich Notebook they had become entangled with one another and with the problem of energy-momentum conservation. The entanglement was the result of Einstein's use of coordinate restrictions. One and the same restriction had to reduce the relevant component of the field equations to the Poisson equation in the case of weak static fields, guarantee energy-momentum conservation, and allow the metric for Minkowski spacetime in rotating coordinates. Coordinate conditions only have to do the first of these three things. The three problems can thus be disentangled by switching from coordinate restrictions to coordinate conditions.

Einstein, we believe, made this switch when he saw that field equations based on the November tensor can be made compatible with energy-momentum conservation by imposing just one weak coordinate restriction. Recovering the Poisson equation for weak static fields still required the usual four restrictions. This discrepancy of one restriction versus four opened up the possibility to handle recovery of the Poisson equation with a coordinate condition in the modern sense and impose a coordinate restriction only for energy-momentum conservation. It is impossible to say whether Einstein had arrived at the modern understanding of coordinate conditions earlier or whether he only reached this point when faced with this unexpected discrepancy. Only the separation of the two sets of conditions, however, made it possible to put the modern understanding of coordinate conditions to good use. Not only could Einstein now decouple the problem of energy-momentum conservation from the problem of recovering the Poisson equation, he could also decouple the latter from the problem of rotation. It is this disentanglement of various conditions and requirements that we tried to capture in the title of our paper: "Untying the knot."

The first November paper contains the first unambiguous instance of Einstein applying a coordinate condition in the modern sense to show that the relevant component of the field equations reduces to the Poisson equation for weak static fields (Einstein 1915a, 786). In the Zurich Notebook Einstein had used what we call the *Hertz restriction*[28] for this purpose. One of the problems with this restriction was that it does not allow the Minkowski metric in rotating coordinates. In the first November paper, Einstein used the exact same mathematical formula, but now interpreted as a coordinate condition rather than a coordinate restriction. As Einstein clearly recognized, it then no longer is a problem that the condition is not satisfied by the Minkowski metric in rotating coordinates. Right after he applied the Hertz condition, he pointed out that the class of unimodular transformations under which the field equations are invariant allow transformations to rotating coordinates. The obvious implication is that the new theory steers clear of the problem of rotation that had defeated the old one.

Einstein had untied the knot. The definition of the components of the gravitational field had been the thread he had pulled to do so. No wonder that he called the old definition "a fateful prejudice" and the new one "the key to the solution."

1.6 Tug of War: Physics or Mathematics?

How well does the text of the November 1915 papers support our reconstruction of how Einstein found his way back to generally-covariant field equations? As a matter of fact, Einstein does *not* introduce the new field equations by pointing out that they can be obtained simply by changing the definition of the gravitational field in the expression for the Lagrangian from which he had earlier derived the *Entwurf* equations. Instead, he uses that the new field equations are closely related to the generally-covariant Riemann tensor, rehearsing the argument that had led him to the November tensor in the Zurich Notebook. At first glance, this looks like a strike against us. On closer examination, it is not such a clear call. In his paper, Einstein was presumably concerned with making the strongest possible case for his new field equations. No matter how Einstein had arrived at these new field equations, it clearly was more con-

28 The only reason for this name is that the condition is discussed in Einstein to Paul Hertz, August 22, 1915 (CPAE 8, Doc. 111). See sec. 5.5.2 of "Commentary …" (in this volume).

vincing to show that these equations can easily be extracted from the Ricci tensor than to show that they can be obtained by a natural adjustment of the formalism that Einstein had used the year before in his failed attempt to prove the uniqueness of the *Entwurf* field equations. Emphasizing the former argument and suppressing the latter would have been the obvious preemptive strike against skeptical readers who might want to remind him of that fiasco. But it need not even have been a calculated rhetorical move on Einstein's part. He himself probably saw the connection to the Riemann tensor as the most convincing evidence in favor of his new field equations. It thus makes perfect sense that this is what he emphasized in his presentation and that he only availed himself of the variational formalism to do the one thing he did not know how to do any other way, namely proving compatibility with energy-momentum conservation.

If we are right, Einstein's papers of November 1915 not only gave his contemporaries and a host of later commentators a misleading picture of how he found the field equations of general relativity, they also and most importantly fooled their own author. Einstein would soon forget that he had arrived at the new field equations pursuing the physical strategy and that the complementary mathematical strategy had served mainly to give him the confidence that he was finally on the right track. In his later years, Einstein extolled the virtues of a purely mathematical approach to theory construction. As John Norton (2000) and, in much greater detail, Jeroen van Dongen (2002, 2004) have shown, the older Einstein routinely claimed that this was the lesson he had drawn from the way in which he had found general relativity. The way Einstein remembered it, physics had led him astray; it was only after he had decided to throw in his fate with mathematics that he had found the right theory. In our reconstruction, however, Einstein found his way back to generally-covariant field equations by making one important adjustment to the *Entwurf* theory, a theory born almost entirely out of physical considerations. He saw that he could redefine the components of the gravitational field without losing any of the structural similarities to electrodynamics that made the *Entwurf* theory so attractive from a physical point of view. After a few more twists and turns, this path led him to the Einstein field equations. That mathematical considerations pointed in the same direction undoubtedly inspired confidence that this was the right direction, but guiding him along this path were physical not mathematical considerations.

1.7 The Red Thread: Einstein's Variational Formalism

In the rest of this paper, we fill in the details of our new reconstruction of the transition from the *Entwurf* theory to general relativity. For those who do not want to go through the derivations, we give short summaries at the beginning of all (sub-)sections of the results derived in them. In sec. 2, we review the one result we need from the Zurich Notebook, namely the extraction of the November tensor from the Ricci tensor. In sec. 3, we give a self-contained exposition of the variational formalism of (Einstein 1914c) that plays a pivotal role in our account. In sec. 4, we show how Einstein used this formalism to make what he considered his most compelling case for the *Entwurf* theory. In secs. 5–7, we analyze how Einstein used the formalism in his papers of November 1915 (Einstein 1915a, b, d). In secs. 8 and 9, we turn to two papers (Einstein 1916a, 1916c) in which the results of November 1915 were consolidated, again with the help of the formalism of 1914. In sec. 10, we address the discrepancy noted above between how Einstein presented and remembered his discovery of general relativity and how he actually discovered it. Finally, in the appendix, drawing on calculations scattered throughout the body of the paper, we present a concise and sanitized version of the transition from the *Entwurf* field equations to the Einstein field equations, which makes the relation between these two sets of equations more perspicuous.

2. THE NOVEMBER TENSOR IN THE ZURICH NOTEBOOK

We review how the field equations based on the November tensor in (Einstein 1915a) made their first appearance in the Zurich Notebook. Einstein extracted the November tensor from the Ricci tensor by imposing a restriction to unimodular transformations. He then showed how the Hertz restriction reduces the November tensor to the d'Alembertian acting on the metric in the case of weak fields.

On p. 22R of the Zurich Notebook—at the instigation, it seems, of his friend and collaborator Marcel Grossmann whose name appears at the top of the page—Einstein wrote down the Ricci tensor in the form

$$T_{il} = \frac{\partial}{\partial x^l}\begin{Bmatrix} k \\ ik \end{Bmatrix} - \frac{\partial}{\partial x^k}\begin{Bmatrix} k \\ il \end{Bmatrix} + \begin{Bmatrix} \lambda \\ ik \end{Bmatrix}\begin{Bmatrix} k \\ \lambda l \end{Bmatrix} - \begin{Bmatrix} \lambda \\ il \end{Bmatrix}\begin{Bmatrix} k \\ \lambda k \end{Bmatrix}, \tag{1}$$

where

$$\begin{Bmatrix} k \\ il \end{Bmatrix} \equiv \frac{1}{2}g^{k\alpha}(g_{i\alpha,l} + g_{l\alpha,i} - g_{il,\alpha}) \tag{2}$$

are the Christoffel symbols.[29] Einstein extracted an expression from the Ricci tensor that transforms as a tensor under unimodular transformations. Under such transformations, the quantity

$$T_i \equiv \begin{Bmatrix} k \\ ik \end{Bmatrix} = (\lg\sqrt{-g})_{,i} \tag{3}$$

(with g the determinant of $g_{\mu\nu}$), transforms as a vector, and its covariant derivative,

$$\frac{\partial T_i}{\partial x^l} - \begin{Bmatrix} \lambda \\ il \end{Bmatrix}T_\lambda = \frac{\partial}{\partial x^l}\begin{Bmatrix} k \\ ik \end{Bmatrix} - \begin{Bmatrix} \lambda \\ il \end{Bmatrix}\begin{Bmatrix} k \\ \lambda k \end{Bmatrix}, \tag{4}$$

transforms as a tensor. It follows that the quantity

$$T_{il}^{x} \equiv \frac{\partial}{\partial x^k}\begin{Bmatrix} k \\ il \end{Bmatrix} - \begin{Bmatrix} \lambda \\ ik \end{Bmatrix}\begin{Bmatrix} k \\ \lambda l \end{Bmatrix}, \tag{5}$$

which is (minus) the difference between the generally-covariant Ricci tensor in eq. (1) and the expression in eq. (4), also transforms as a tensor under unimodular transformations. This is the quantity we call the *November tensor.*

Setting the November tensor equal to the energy-momentum tensor, $T_{\mu\nu}$, multiplied by the gravitational constant κ, one arrives at the field equations of Einstein's first paper of November 1915 (Einstein 1915a, 783, eq. 16a):

$$\frac{\partial}{\partial x^\alpha}\begin{Bmatrix} \alpha \\ \mu\nu \end{Bmatrix} - \begin{Bmatrix} \alpha \\ \mu\beta \end{Bmatrix}\begin{Bmatrix} \beta \\ \alpha\nu \end{Bmatrix} = \kappa T_{\mu\nu}, \tag{6}$$

The first term on the left-hand side does not reduce to the d'Alembertian acting on the metric in the weak-field case. There are additional terms with unwanted second-order derivatives of the metric. At the time of the Zurich Notebook, this made the November tensor itself unacceptable as a candidate for the left-hand side of the field equations. Einstein, however, extracted a candidate for the left-hand side of the field equations from the November tensor by imposing what we call the *Hertz restriction*,

$$g_k^{k\alpha} \equiv \frac{\partial g^{k\alpha}}{\partial x^k} = 0. \tag{7}$$

Expanding the Christoffel symbols in the first term of the November tensor, one finds

$$\frac{\partial}{\partial x^k}\begin{Bmatrix} k \\ il \end{Bmatrix} = \frac{1}{2}(g^{k\alpha}(g_{i\alpha,l} + g_{l\alpha,i} - g_{il,\alpha}))_k.$$

Using the Hertz restriction and the relation

29 We have adopted a notation that lies somewhere between slavishly following the original text and translating everything into modern language. Our guiding principle has been to use a notation that makes the equations both easy to follow for those familiar with the standard notation of modern general relativity and easy to compare with the original sources for those who want to check our claims against Einstein's own writings. On this basis, we have adopted the following rules. We typically follow Einstein's choice of letters for quantities and indices in the document under discussion. E.g., the Ricci tensor is not written as $R_{\mu\nu}$, as it is in most modern texts, but as T_{il} in our discussion of the Zurich Notebook in this section and as G_{im} in our discussion of (Einstein 1915b) in sec. 7. As in Einstein's writings of this period, all indices, Greek and Latin, run from 1 through 4. However, we do not follow Einstein's idiosyncratic convention before (Einstein 1914c) of writing nearly all indices downstairs and distinguishing between covariant and contravariant components (e.g., the components $g_{\mu\nu}$ and $g^{\mu\nu}$ of the metric) by using a Latin letter for one ($g_{\mu\nu}$) and a Greek letter for the other ($\gamma_{\mu\nu}$). We use Latin letters for all quantities and write all covariant indices downstairs and all contravariant indices upstairs. Following Einstein, we use Fraktur for tensor densities (e.g., $\mathfrak{T}_\nu^\mu = \sqrt{-g}T_\nu^\mu$). Deviating from Einstein, we occasionally use commas and semi-colons for ordinary and covariant differentiation, respectively. We consistently use the summation convention (introduced in Einstein 1916a, 788).

$$g^{k\alpha}(g_{i\alpha,l} + g_{l\alpha,i}) = -g_l^{k\alpha}g_{i\alpha} - g_i^{k\alpha}g_{l\alpha}$$

(where $g_l^{\kappa\alpha} \equiv g^{\kappa\alpha}{}_{,l}$), one can rewrite this expression as:

$$\frac{\partial}{\partial x^k}\begin{Bmatrix} k \\ il \end{Bmatrix} = -\frac{1}{2}(g^{k\alpha}g_{il,k\alpha} + g_l^{k\alpha}g_{i\alpha,k} + g_i^{k\alpha}g_{l\alpha,k}). \tag{8}$$

The first term in parentheses reduces to $-\Box g_{il}$ for weak fields. The other two terms are quadratic in first-order derivatives of the metric like the contribution coming from the term in the November tensor quadratic in the Christoffel symbols. All these terms can be neglected for weak fields.

On p. 23L of the notebook, Einstein tried to bring down the number of terms quadratic in first-order derivatives of the metric in his field equations by introducing yet another coordinate restriction in addition to the Hertz restriction and the restriction to unimodular transformations. We call this new restriction the ϑ-restriction.[30] Einstein discovered that this new coordinate restriction could be used to eliminate the unwanted terms with second-order derivatives as well, so that there was no longer any need for the Hertz restriction. Einstein eventually abandoned the ϑ-restriction because the ϑ-restriction—like the Hertz restriction for that matter—ruled out transformations to rotating frames in Minkowski spacetime. After a few more twists and turns, Einstein settled on the *Entwurf* field equations. The November tensor and the Hertz restriction—used now as a coordinate condition—only reappeared in November 1915.

3. EINSTEIN'S VARIATIONAL FORMALISM: FIELD EQUATIONS, ENERGY-MOMENTUM CONSERVATION, AND COVARIANCE PROPERTIES

We cover various aspects of the variational formalism that Einstein used both in his review article on the Entwurf theory of October 1914 and in a number of papers on general relativity in 1915–1918. The Lagrangian is left unspecified, so all results hold both in the Entwurf theory and in modern general relativity. The main point of the section is to show how two very different lines of reasoning—one aimed at finding conditions to ensure energy-momentum conservation, the other aimed at finding coordinate transformations leaving the action invariant—lead to the exact same conditions on the Lagrangian, written as $S_\sigma^\nu = 0$ and $B_\mu = 0$. The convergence of these two lines of reasoning confirmed what Einstein had come to suspect in the fall of 1913, namely that energy-momentum conservation is directly related to the covariance of the gravitational field equations.

30 For discussion of the ϑ-restriction, see note 69 below and secs. 5.5.4–5.5.10 in "Commentary …" (in this volume).

In early November 1915, Einstein (1915a) replaced the *Entwurf* field equations of severely limited covariance by field equations just a few tweaks away from the generally-covariant Einstein field equations of (Einstein 1915d). The way Einstein initially described it, it was a wholesale replacement:

> After all confidence in *the result and the method* of the earlier theory had thus given way, I saw clearly that a satisfactory solution could only be found in a connection to the general theory of covariants, i.e., to Riemann's covariant (our emphasis).[31]

About six weeks later, he recognized that the old had not been all bad:

> The series of my papers on gravitation is a chain of erroneous paths, which nonetheless gradually brought me closer to my goal.[32]

It is true that Einstein discarded some of his earlier *results*, but he retained the *method* that he had used to obtain those results. This method is the variational formalism first presented in (Einstein and Grossmann 1914b) and further developed in the definitive exposition of the *Entwurf* theory (Einstein 1914c, part D). In the latter paper, he used this formalism to produce an elegant derivation of the *Entwurf* field equations, to investigate their covariance properties, and to prove their compatibility with energy-momentum conservation. We argue that he used this same formalism to find the successor to the *Entwurf* field equations, published in the first of his four communications to the Prussian Academy in November 1915 (Einstein 1915a).

What complicates the use of the formalism both in the four November papers (Einstein 1915a, b, c, d) and in the first systematic exposition of the new theory (Einstein 1916a) is a restriction to unimodular transformations in the first paper and the choice of unimodular coordinates[33] in the other four. In all these papers, Einstein nonetheless relied heavily on the formalism to guide him in his analysis of the relation between field equations and energy-momentum conservation.

---

31 "Nachdem so jedes Vertrauen im Resultate und Methode der früheren Theorie gewichen war, sah ich klar, dass nur durch einen Anschluss an die allgemeine Kovariantentheorie, d.h. an Riemanns Kovariante, eine befriedigende Lösung gefunden werden konnte." Einstein to Arnold Sommerfeld, November 28, 1915 (CPAE 8, Doc. 153; our emphasis). Unless otherwise noted, all translations are based on those in the companion volumes to the Einstein edition. This letter to Sommerfeld provides the most detailed account of the developments of November 1915 that culminated in the publication of the Einstein field equations and the explanation of the anomalous motion of Mercury's perihelion. This document, however, needs to be treated with care. It was a calculated move on Einstein's part to tell Sommerfeld the whole story rather than, say, Lorentz, with whom he had corresponded much more intensively on matters general relativistic. In the fall of 1915, Sommerfeld was kept apprised of developments not only by Einstein but also by Hilbert. Writing to Sommerfeld, Einstein probably first and foremost wanted to make sure that Sommerfeld knew that he had put his house in order without any help from Hilbert.

32 "Die Serie meiner Gravitationsarbeiten ist eine Kette von Irrwegen, die aber doch allmählich dem Ziele näher führten." Einstein to H. A. Lorentz, January 17, 1916 (CPAE 8, Doc. 183). The mixing of metaphors ("Kette von Irrwegen") is Einstein's, not ours.

33 Recall the discussion of the difference between coordinate restrictions and coordinate conditions in the introduction.

It was only in (Einstein 1916c), written in October 1916, that Einstein first presented the new theory entirely in arbitrary rather than in unimodular coordinates. This paper follows the exposition of the variational formalism in (Einstein 1914c) almost to the letter (as already emphasized in Norton 1984, 141). In early 1918, Einstein used the formalism again to defend his approach to energy-momentum conservation in general relativity against objections from Levi-Civita, Lorentz, Klein, and others (Einstein 1918d). The formalism can also be found in Einstein's lecture notes for a course on general relativity in Berlin in 1919 (CPAE 7, Doc. 19, [pp. 13–17]).

Einstein's reliance on this variational formalism thus provides an important element of continuity in the transition from the *Entwurf* theory to general relativity and puts the lie to Einstein's remark to Sommerfeld that he had lost all confidence in both "the result and the method" of the old theory.

In this section, we cover various aspects of Einstein's formalism: the derivation of the field equations (sec. 3.1), the treatment of energy-momentum conservation (sec. 3.2), and the investigation of covariance properties (sec. 3.3). In subsequent sections, we discuss the applications of the formalism in the period 1914–1916. In sec. 4, we examine the relevant portion of (Einstein 1914c) published in November 1914. In secs. 5–7, we turn to the papers of November 1915 documenting the transition from the *Entwurf* theory to general relativity (Einstein 1915a, b, d). In sec. 8, we present the streamlined version of the argument of November 1915 given in the review article completed in March 1916 (Einstein 1916a, part C). In sec. 9, we cover what is probably the most elegant application of the formalism, the demonstration that energy-momentum conservation in general relativity is a direct consequence of the general covariance of the field equations. This argument was made in (Einstein 1916c), presented to the Prussian Academy in November 1916. Our story thus covers a timespan of two years, from November 1914 to November 1916. Our main focus will be on the tumultuous developments of one month in the middle of this period, November 1915.

3.1 Field Equations

With the appropriate definition of the gravitational energy-momentum pseudo-tensor, the field equations can be written in a form resembling Maxwell's equations, with the divergence of the gravitational field on the left-hand side and the sum of the energy-momentum densities of matter and gravitational field on the right-hand side.

Consider the gravitational part of the action

$$J = \int Q d\tau, \tag{9}$$

where $Q = H\sqrt{-g}$ is the gravitational part of the Lagrangian,[34] H is some as yet completely undetermined function of $g^{\mu\nu}$ and its first derivatives $g^{\mu\nu}_{\alpha}$, and $d\tau = d^4x$ is a four-dimensional volume element. The condition that J be an extremum, $\delta J = 0$, leads to the Euler-Lagrange equations

$$\frac{\partial}{\partial x^{\alpha}}\left(\frac{\partial Q}{\partial g^{\mu\nu}_{\alpha}}\right) - \frac{\partial Q}{\partial g^{\mu\nu}} = 0. \tag{10}$$

Although his proof did not satisfy Levi-Civita, Einstein thought he could show that the expression on the left-hand side transforms as a tensor density under all transformations under which H transforms as a scalar.

He generalized the vacuum field equations (10) to

$$\frac{\partial}{\partial x^{\alpha}}\left(\frac{\partial Q}{\partial g^{\mu\nu}_{\alpha}}\right) - \frac{\partial Q}{\partial g^{\mu\nu}} = -\kappa\sqrt{-g}T_{\mu\nu} \tag{11}$$

in the presence of matter described by the energy-momentum tensor $T_{\mu\nu}$. This equation can be written in a form analogous to Maxwell's equations, with the divergence of the gravitational field on the left-hand side and the sources of the field, the energy-momentum densities of matter and gravitational field, on the right-hand side. Contraction with $g^{\nu\lambda}$ of the left-hand side of eq. (11) gives:

$$\frac{\partial}{\partial x^{\alpha}}\left(g^{\nu\lambda}\frac{\partial Q}{\partial g^{\mu\nu}_{\alpha}}\right) - g^{\nu\lambda}_{\alpha}\frac{\partial Q}{\partial g^{\mu\nu}_{\alpha}} - g^{\nu\lambda}\frac{\partial Q}{\partial g^{\mu\nu}};$$

contraction with $g^{\nu\lambda}$ of the right-hand $-\kappa\sqrt{-g}T^{\lambda}_{\mu}$. If the gravitational energy-momentum pseudo-tensor t^{λ}_{μ} is defined as

$$\kappa\sqrt{-g}t^{\lambda}_{\mu} \equiv -g^{\nu\lambda}_{\alpha}\frac{\partial Q}{\partial g^{\mu\nu}_{\alpha}} - g^{\nu\lambda}\frac{\partial Q}{\partial g^{\mu\nu}}, \tag{12}$$

then eq. (11) can be rewritten as

$$\frac{\partial}{\partial x^{\alpha}}\left(g^{\nu\lambda}\frac{\partial Q}{\partial g^{\mu\nu}_{\alpha}}\right) = -\kappa(\mathfrak{T}^{\lambda}_{\mu} + \mathfrak{t}^{\lambda}_{\mu}), \tag{13}$$

where $\mathfrak{T}^{\lambda}_{\mu} \equiv \sqrt{-g}T^{\lambda}_{\mu}$ and $\mathfrak{t}^{\lambda}_{\mu} \equiv \sqrt{-g}t^{\lambda}_{\mu}$ are mixed tensor densities. These field equations fulfill an important requirement: the energy-momentum of the gravitational field enters the source term in the same way as the energy-momentum of matter.

If the quantity in parentheses on the left-hand side of eq. (13) is identified as the gravitational field, the equations have the same structure as Maxwell's equations, $\partial_{\mu}F^{\mu\nu} = \mu_0 j^{\nu}$, where $F^{\mu\nu}$ is the electromagnetic field tensor, μ_0 is a constant, and j^{ν} is the charge-current density, the source of the electromagnetic field.

34 Strictly speaking, Q is the Lagrangian density (for detailed discussion, see Wald 1984, Appendix E).

3.2 Energy-Momentum Conservation

In addition to the field equations, it is assumed that the covariant divergence of the energy-momentum tensor of matter vanishes. This equation can be rewritten as the vanishing of the ordinary divergence of the sum of the energy-momentum tensor densities for matter and gravitational field, provided that the gravitational energy-momentum pseudo-tensor is defined appropriately. Compatibility of this definition and the definition in the preceding subsection leads to the conditions $S_\sigma^\nu = 0$ *on the Lagrangian. In 1914, Einstein (erroneously) thought that these conditions uniquely pick out the Entwurf Lagrangian. Einstein imposed four more conditions, written as* $B_\mu = 0$*. Taken together with the field equations, the conditions* $B_\mu = 0$ *imply energy-momentum conservation. In general relativity, these conditions turn into the contracted Bianchi identities.*

The energy-momentum balance for matter in a gravitational field can be written as[35]

$$\mathfrak{T}^\alpha_{\mu,\alpha} = \begin{Bmatrix}\beta \\ \mu\alpha\end{Bmatrix}\mathfrak{T}^\alpha_\beta. \tag{14}$$

This equation is equivalent to (cf. Einstein 1914c, 1056, eq. 42a):[36]
$\mathfrak{T}^\alpha_{\mu,\alpha} = \frac{1}{2}g^{\beta\rho}g_{\rho\alpha,\mu}\mathfrak{T}^\alpha_\beta$. four-momentum density of matter at any given point can only change in two ways: it can flow to or from neighboring points and it can be transferred to or from the gravitational field at that point. The left-hand sides of eqs. (14)–() describe the former process, the right-hand sides the latter. The right-hand sides give the rate at which four-momentum density is transferred from gravitational field to matter. This term thus represents the gravitational force density. The analogy with the Lorentz force density $f_\mu = F_{\mu\nu}j^\nu$ —the contraction of the electromagnetic field $F_{\mu\nu}$ and its source, the charge-current density j^ν —suggests that this quantity should be equal to minus the contraction of the gravitational field and its source, the energy-momentum tensor of matter. The minus sign reflects that the gravitational force

---

35 Eq. (14) is equivalent to $T^\alpha_{\mu;\alpha} = 0$:

$$\sqrt{-g}T^\alpha_{\mu;\alpha} = \sqrt{-g}\left(T^\alpha_{\mu,\alpha} + \begin{Bmatrix}\alpha \\ \beta\alpha\end{Bmatrix}T^\beta_\mu - \begin{Bmatrix}\beta \\ \mu\alpha\end{Bmatrix}T^\alpha_\beta\right) = 0$$

Upon substitution of $\begin{Bmatrix}\alpha \\ \beta\alpha\end{Bmatrix} = (\sqrt{-g})_{,\beta}/\sqrt{-g}$, this equation turns into

$$\sqrt{-g}T^\alpha_{\mu,\alpha} + (\sqrt{-g})_{,\alpha}T^\alpha_\mu - \begin{Bmatrix}\beta \\ \mu\alpha\end{Bmatrix}\sqrt{-g}T^\alpha_\beta = 0,$$

which can be rewritten as eq. (14).

36 Using definition (2) of the Christoffel symbols, one can rewrite the right-hand side of eq. (14) as

$$\begin{Bmatrix}\beta \\ \mu\alpha\end{Bmatrix}\mathfrak{T}^\alpha_\beta = \frac{1}{2}g^{\beta\rho}(g_{\rho\mu,\alpha} + g_{\rho\alpha,\mu} - g_{\mu\alpha,\rho})\mathfrak{T}^\alpha_\beta = \frac{1}{2}\mathfrak{T}^{\alpha\rho}(g_{\rho\mu,\alpha} - g_{\mu\alpha,\rho}) + \frac{1}{2}g^{\beta\rho}g_{\rho\alpha,\mu}\mathfrak{T}^\alpha_\beta.$$

The first term of this last expression vanishes since it is the contraction of a quantity symmetric in the indices α and ρ and a quantity anti-symmetric in those same indices.

between two masses is attractive whereas the electric force between two like charges is repulsive. In the *Entwurf* theory, Einstein read off the expression for the gravitational field from the right-hand side of eq. ().[37] In the November 1915 theory, he used the right-hand side of eq. (14) instead. Commenting on this switch in his first paper of November 1915, Einstein wrote:

> This conservation law [essentially eq. ()] has led me in the past to look upon the quantities $[(1/2)g^{\tau\mu}g_{\mu\nu,\sigma}]$ as the natural expressions of the components of the gravitational field, even though the formulas of the absolute differential calculus suggest the Christoffel symbols […] instead. *This was a fateful prejudice*.[38]

After the fourth paper of November 1915, he told Sommerfeld:

> *The key to this solution* was my realization that not $[g^{l\alpha}g_{\alpha i,m}]$ but the related Christoffel symbols […] are to be regarded as the natural expression for the "components" of the gravitational field.[39]

To appreciate the full significance of these comments, one needs to see how the November tensor drops out of Einstein's variational formalism (see sec. 5 below). For now, we shall follow the treatment in (Einstein 1914c) and work with eq. () rather than with eq. (14).

In relativistic continuum mechanics, which is carefully tailored to electrodynamics, the theory for which it was first developed, a four-force density can be written as the four-divergence of a suitably chosen energy-momentum tensor.[40] So Einstein tried to write the gravitational force density in eq. () as the four-divergence of a suitably chosen gravitational energy-momentum (pseudo-)tensor density $\mathfrak{t}^{\alpha}_{\mu}$. If this can be done, energy-momentum conservation can be written as the vanishing of an ordinary divergence:

$$(\mathfrak{T}^{\lambda}_{\mu}+\mathfrak{t}^{\lambda}_{\mu})_{,\lambda} = 0. \tag{15}$$

Eq. () can indeed be written in this form, but the resulting expression for $\mathfrak{t}^{\alpha}_{\mu}$ differs from expression (12) for $\mathfrak{t}^{\alpha}_{\mu}$ found earlier. Einstein therefore had to add an extra condition to his theory that sets these two expressions equal to one another. As he discovered in October 1915, this same condition pops up in the analysis of the covariance properties of the theory.

---

37 In doing so, Einstein omitted a minus sign. The motivation for using eq. () rather than eq. (14) to identify the components of the gravitational field is explained in (Einstein 1914c, 1060, note 1). See also Einstein to Hans Thirring, 7 December 1917 (CPAE 8, Doc. 405, note 4).

38 "Diese Erhaltungsgleichung hat mich früher dazu verleitet, die Größen […] als den natürlichen Ausdruck für die Komponenten des Gravitationsfeldes anzusehen, obwohl es im Hinblick auf die Formeln des absoluten Differentialkalküls näher liegt, die Christoffelschen Symbole statt jener Größen einzuführen. Dies war ein verhängnisvolles Vorurteil" (Einstein 1915a, 782; our emphasis).

39 "Den Schlüssel zu dieser Lösung lieferte mir die Erkenntnis, dass nicht […] sondern die damit verwandten Christoffel'schen Symbole […] als natürlichen Ausdruck für die "Komponente" des Gravitationsfeldes anzusehen ist." Einstein to Arnold Sommerfeld, 28 November 1915 (CPAE 8, Doc. 153). Our emphasis.

40 See, e.g., sec. 20 of Einstein's "Manuscript on the Special Theory of Relativity" (CPAE 4, Doc. 1).

Since $g^{\beta\rho}g_{\rho\alpha,\mu} = -g^{\beta\rho}_{\mu}g_{\rho\alpha}$, eq. () can also be written as:

$$\mathfrak{T}^{\alpha}_{\mu,\alpha} + \frac{1}{2}g^{\alpha\beta}_{\mu}\mathfrak{T}_{\alpha\beta} = 0. \tag{16}$$

The second term has to be written in the form of $t^{\lambda}_{\mu,\lambda}$. To this end, the field equations (11) are used to replace $\mathfrak{T}_{\alpha\beta}$ by an expression in terms of the metric field and its derivatives:[41]

$$\frac{1}{2}g^{\alpha\beta}_{\mu}\mathfrak{T}_{\alpha\beta} = \frac{1}{2\kappa}g^{\alpha\beta}_{\mu}\left(\frac{\partial Q}{\partial g^{\alpha\beta}} - \frac{\partial}{\partial x^{\lambda}}\left(\frac{\partial Q}{\partial g^{\alpha\beta}_{\lambda}}\right)\right). \tag{17}$$

The right-hand side of this expression can indeed be written in the form $t^{\lambda}_{\mu,\lambda}$ with:[42]

$$t^{\lambda}_{\mu} \equiv \frac{1}{2\kappa}\left(\delta^{\lambda}_{\mu}Q - g^{\alpha\beta}_{\mu}\frac{\partial Q}{\partial g^{\alpha\beta}_{\lambda}}\right). \tag{18}$$

We introduce the more explicit notations $t^{\lambda}_{\mu}(Q, \text{cons})$ for t^{λ}_{μ} as defined in eq. (18) and $t^{\lambda}_{\mu}(Q, \text{source})$ for t^{λ}_{μ} as defined in eq. (12).[43]

Compatibility between these two definitions is assured if the quantity S^{ν}_{σ}, defined as (Einstein 1914c, 1075, eq. 76a),[44]

$$S^{\nu}_{\sigma} \equiv g^{\beta\nu}_{\alpha}\frac{\partial Q}{\partial g^{\beta\sigma}_{\alpha}} + g^{\alpha\nu}\frac{\partial Q}{\partial g^{\alpha\sigma}} + \frac{1}{2}\delta^{\nu}_{\sigma}Q - \frac{1}{2}g^{\alpha\beta}_{\sigma}\frac{\partial Q}{\partial g^{\alpha\beta}_{\nu}}, \tag{19}$$

41 This method for identifying the expression for the gravitational energy-momentum pseudo-tensor can already be found at several places in the Zurich Notebook (see, e.g., p. 19R and p. 24R, discussed in secs. 5.4.2 and 5.6.1, respectively, of "Commentary …" [in this volume]). It was also used in (Einstein and Grossmann 1913, 15). Einstein had used a completely analogous method in his earlier theory for static fields (Einstein 1912, 456).

42 Using that

$$\frac{\partial Q}{\partial x^{\mu}} = \frac{\partial Q}{\partial g^{\alpha\beta}}g^{\alpha\beta}_{\mu} + \frac{\partial Q}{\partial g^{\nu\sigma}_{\lambda}}g^{\alpha\beta}_{\lambda\mu},$$

with $g^{\alpha\beta}_{\lambda\mu} \equiv g^{\alpha\beta}{}_{,\lambda\mu}$, one can rewrite the first term on the right-hand side of eq. (17) as

$$g^{\alpha\beta}_{\mu}\frac{\partial Q}{\partial g^{\alpha\beta}} = \frac{\partial Q}{\partial x^{\mu}} - \frac{\partial Q}{\partial g^{\nu\sigma}_{\lambda}}g^{\alpha\beta}_{\lambda\mu}.$$

One thus arrives at

$$\frac{\partial Q}{\partial x^{\mu}} - g^{\alpha\beta}_{\mu\lambda}\frac{\partial Q}{\partial g^{\nu\sigma}_{\lambda}} - g^{\alpha\beta}_{\mu}\frac{\partial}{\partial x^{\lambda}}\left(\frac{\partial Q}{\partial g^{\alpha\beta}_{\lambda}}\right) = \frac{\partial}{\partial x^{\lambda}}\left(\delta^{\lambda}_{\mu}Q - g^{\alpha\beta}_{\mu}\frac{\partial Q}{\partial g^{\alpha\beta}_{\lambda}}\right),$$

from which eq. (18) follows.

43 The designations "source" and "cons[ervation]" refer to the fact that these two definitions are found from considerations of the source term of the field equations and considerations of energy-momentum conservation, respectively.

44 Our derivation of eq. (19) follows Einstein to H. A. Lorentz, 12 October 1915 (CPAE 8, Doc. 129). In his 1914 review article, Einstein derived this equation by substituting the left-hand side of eq. (11) for $\mathfrak{T}_{\mu\nu}$ in both terms on the left-hand side of eq. (16).

vanishes. This quantity is equal to (κ times) the difference between the two definitions of t^{λ}_{μ}:

$$S^{\nu}_{\sigma}(Q) = \kappa t^{\nu}_{\sigma}(Q, \text{cons}) - \kappa t^{\nu}_{\sigma}(Q, \text{source}). \tag{20}$$

At the time of his review article on the *Entwurf* theory, Einstein thought that this condition uniquely determined $Q = \sqrt{-g}H$ to be the Lagrangian for the *Entwurf* field equations. As it turns out, S^{ν}_{σ} vanishes for *any* H that transforms as a scalar under general linear transformations (see eqs. (30)–(31) below).

Einstein imposed another set of conditions on the Lagrangian density Q which guarantee energy-momentum conservation. Taking the divergence of both sides of the field equations (13), one arrives at

$$-\frac{\partial^2}{\partial x^{\lambda}\partial x^{\alpha}}\left(g^{\nu\lambda}\frac{\partial Q}{\partial g^{\mu\nu}_{\alpha}}\right) = \kappa(\mathfrak{T}^{\lambda}_{\mu} + t^{\lambda}_{\mu})_{,\lambda}. \tag{21}$$

The field equations thus imply energy-momentum conservation, if Q satisfies the condition (Einstein 1914c, 1077):

$$B_{\mu} \equiv \frac{\partial^2}{\partial x^{\rho}\partial x^{\sigma}}\left(g^{\rho\tau}\frac{\partial Q}{\partial g^{\mu\tau}_{\sigma}}\right) = 0.. \tag{22}$$

Energy-momentum thus calls for two sets of conditions:

$$S^{\nu}_{\sigma} = 0,\ B_{\mu} = 0, \tag{23}$$

These same conditions, it turns out, also express the covariance properties of the field equations (see eq. (33) below).

3.3 Covariance Properties

The conditions the Lagrangian has to satisfy for the action to be invariant under a given coordinate transformation are determined. The assumption that the action is at least invariant under general linear transformations leads to the conditions $S^{\nu}_{\sigma} = 0$. Einstein did not explicitly write down these conditions in 1914, which explains why he thought that these conditions, which he did encounter in the context of energy-momentum conservation, could be used to determine the Lagrangian. The conditions for additional non-linear transformations leaving the action invariant are $B_{\mu} = 0$, which, as Einstein did recognize, were also the conditions guaranteeing energy-momentum conservation. In the Entwurf theory, these four conditions determine the class of what Einstein called "adapted coordinates." In general relativity, they turn into the generally-covariant contracted Bianchi identities.

What are the transformations that leave the action $J = \int H\sqrt{-g}d\tau$ (with H an arbitrary function of $g^{\mu\nu}$ and $g^{\mu\nu}_{\alpha}$) invariant?[45] Consider an arbitrary infinitesimal coordinate transformation $x'^{\mu} = x^{\mu} + \Delta x^{\mu}$. The changes in $g^{\mu\nu}$ and $g^{\mu\nu}_{\alpha}$ under this transformation are given by[46,47]

$$\Delta g^{\mu\nu} \equiv g'^{\mu\nu}(x') - g^{\mu\nu}(x) = g^{\mu\alpha}\frac{\partial \Delta x^{\nu}}{\partial x^{\alpha}} + g^{\nu\alpha}\frac{\partial \Delta x^{\mu}}{\partial x^{\alpha}}, \tag{24}$$

$$\Delta g^{\mu\nu}_{\alpha} \equiv g'^{\mu\nu}_{\alpha}(x') - g^{\mu\nu}_{\alpha}(x) = \frac{\partial}{\partial x^{\alpha}}(\Delta g^{\mu\nu}) - \frac{\partial \Delta x^{\beta}}{\partial x^{\alpha}} g^{\mu\nu}_{\beta}. \tag{25}$$

The change in J is given by

$$\Delta J = J' - J = \int Q' d\tau' - \int Q d\tau. \tag{26}$$

The Jacobian $|\partial x'/\partial x|$ can be written as $1 + \partial\Delta x^{\mu}/\partial x^{\mu}$.[48] ΔJ can then be rewritten as

$$\Delta J = \int Q'\left(1 + \frac{\partial \Delta x^{\mu}}{\partial x^{\mu}}\right) d\tau - \int Q d\tau = \int Q \delta^{\nu}_{\sigma}\frac{\partial \Delta x^{\sigma}}{\partial x^{\nu}} d\tau + \int \Delta Q d\tau. \tag{27}$$

Since Q is a function of $g^{\mu\nu}$ and $g^{\mu\nu}_{\alpha}$, $\Delta Q \equiv Q' - Q$ is given by

$$\Delta Q = \frac{\partial Q}{\partial g^{\mu\nu}}\Delta g^{\mu\nu} + \frac{\partial Q}{\partial g^{\mu\nu}_{\alpha}}\Delta g^{\mu\nu}_{\alpha}.$$

Inserting eqs. (24)–(25) for $\Delta g^{\mu\nu}$ and $\Delta g^{\mu\nu}_{\alpha}$, one finds[49]

$$\Delta Q = \left\{2g^{\alpha\nu}\frac{\partial Q}{\partial g^{\alpha\sigma}} + 2g^{\beta\nu}_{\alpha}\frac{\partial Q}{\partial g^{\beta\sigma}_{\alpha}} - g^{\alpha\beta}_{\sigma}\frac{\partial Q}{\partial g^{\alpha\beta}_{\nu}}\right\}\frac{\partial \Delta x^{\sigma}}{\partial x^{\nu}} + 2\frac{\partial Q}{\partial g^{\mu\tau}_{\sigma}} g^{\rho\tau}\frac{\partial^2 \Delta x^{\mu}}{\partial x^{\sigma}\partial x^{\rho}}. \tag{28}$$

Inserting this expression for ΔQ into expression (27) for ΔJ, one finds:

$$\begin{aligned}\frac{1}{2}\Delta J = \int\left\{g^{\beta\nu}_{\alpha}\frac{\partial Q}{\partial g^{\beta\sigma}_{\alpha}} + g^{\alpha\nu}\frac{\partial Q}{\partial g^{\alpha\sigma}} + \frac{1}{2}\delta^{\nu}_{\sigma}Q - \frac{1}{2}g^{\alpha\beta}_{\sigma}\frac{\partial Q}{\partial g^{\alpha\beta}_{\nu}}\right\}\frac{\partial \Delta x^{\sigma}}{\partial x^{\nu}} d\tau \\ + \int g^{\rho\tau}\frac{\partial Q}{\partial g^{\mu\tau}_{\sigma}}\frac{\partial^2 \Delta x^{\mu}}{\partial x^{\rho}\partial x^{\sigma}} d\tau\cdot\end{aligned} \tag{29}$$

The expression in curly brackets in the first integral is just the quantity S^{ν}_{σ} defined in eq. (19) in the course of the discussion of energy-momentum conservation. Eq. (29) can thus be written more compactly as

45 Ultimately, the question is under which transformations the field equations are invariant. Both in (Einstein and Grossmann 1914b) and in (Einstein 1914c, 1069–1071), Einstein argued that these are just the transformations under which the action is invariant. Levi-Civita's criticism was aimed at this part of Einstein's argument, which for our purposes is not important. Einstein and Grossmann (1914b, 219, note 2) credit Paul Bernays with the suggestion to use a variational formalism to investigate the covariance properties of the *Entwurf* field equations.

$$\frac{1}{2}\Delta J = \int S_\sigma^\nu \frac{\partial \Delta x^\sigma}{\partial x^\nu} d\tau + \int g^{\rho\tau} \frac{\partial Q}{\partial g_\sigma^{\mu\tau}} \frac{\partial^2 \Delta x^\mu}{\partial x^\rho \partial x^\sigma} d\tau. \tag{30}$$

Einstein now focused his attention on functions H that transform as scalars under arbitrary linear transformations. This implies that the action J is invariant under linear transformations ($\sqrt{-g}d\tau$ is an invariant volume element). For linear transformations the second-order derivatives of Δx^μ all vanish, so the second integral in eq. (30) does not contribute to ΔJ. This means that the first integral must vanish identically for arbitrary values of the first-order derivatives of Δx^μ, i.e., that:

$$S_\sigma^\nu = 0. \tag{31}$$

Through partial integration, the second integral in eq. (30) can be rewritten as

$$\int \frac{\partial^2}{\partial x^\rho \partial x^\sigma}\left(g^{\rho\tau} \frac{\partial Q}{\partial g_\sigma^{\mu\tau}}\right) \Delta x^\mu d\tau$$

plus surface terms that can all be assumed to vanish. The integrand is the contraction of Δx^μ and an expression which is exactly equal to the quantity B_μ defined in eq. (22) in the context of the discussion of energy-momentum conservation:

$$B_\mu \equiv \frac{\partial^2}{\partial x^\rho \partial x^\sigma}\left(g^{\rho\tau} \frac{\partial Q}{\partial g_\sigma^{\mu\tau}}\right).$$

Assuming that condition (31) holds, one can thus rewrite eq. (30) as

46 Einstein had an idiosyncratic way of computing the variations induced by coordinate transformations. Felix Klein, David Hilbert, Emmy Noether, and other mathematicians in or closely affiliated with Göttingen (such as Hermann Weyl in Zurich) used what is called the "Lie variation" of the metric tensor, defined as $\delta g_{\mu\nu} \equiv g'_{\mu\nu}(x) - g_{\mu\nu}(x)$. Commenting on (Weyl 1918), which he was reading in proof, Einstein wrote: "He [Weyl] derives the energy law for matter with the same variational trick that you used in the note that recently appeared [Klein 1917]" ("Den Energiesatz der Materie leitet er mit demselben Variations-Kunstgriff ab wie Sie in Ihrer neulich erschienenen Note," Einstein to Felix Klein, 24 March 1918 [CPAE 8, Doc. 492]). Two years earlier he had already noted that Lie variation and differentiation commute (Einstein to David Hilbert, 30 March 1916 [CPAE 8, Doc. 207]). This is not the case if one does the variation the way Einstein does (see eq. (25)). One would have expected Einstein to pick up on this quickly. After all, the crucial distinction between $g'_{\mu\nu}(x')$ and $g'_{\mu\nu}(x)$ was very familiar to him from the hole argument (see, e.g., sec. 4 in "What Did Einstein Know …". [in this volume]). In fact, Einstein still had not fully assimilated the notion of Lie variation in late 1918, as can be inferred from his comments on (Klein 1918a): "At first I had some trouble understanding your equation (6) [involving Lie variation]. The point is that with your preferred way of doing variations $\delta(\partial g^{\mu\nu}/\partial x_\sigma) = \partial/\partial x_\sigma(\delta g^{\mu\nu})$" ("Anfänglich hatte ich etwas Mühe, Ihre Gleichung (6) zu begreifen. Der Witz ist eben, dass bei der von Ihnen bevorzugten Art zu varieren […] ist," Einstein to Felix Klein, 22 October 1918 [CPAE 8, Doc. 638]). Old habits die hard. In his lectures on general relativity in Berlin in 1919, Einstein still vacillated between his own way of doing the variations and that of Weyl and Klein (CPAE 7, Doc. 19, [pp. 13–17]). For discussion of different types of variation that played a role in the early years of general relativity, see (Kichenassamy 1993), sec. 3.

$$\frac{1}{2}\Delta J = \int B_\mu \Delta x^\mu d\tau. \tag{32}$$

The upshot then is that the following two sets of conditions have to be satisfied for the action to be invariant under some coordinate transformation $x'^\mu = x^\mu + \Delta x^\mu$:

$$S_\sigma^\nu = 0, \; B_\mu = 0. \tag{33}$$

These conditions are the same as the conditions guaranteeing energy-momentum conservation that we found in eq. (23).

In the case of the four conditions $B_\mu = 0$, Einstein clearly recognized in his 1914 review article that they play this dual role (Einstein 1914c, 1076–1077). In the case of the conditions $S_\sigma^\nu = 0$, however, he did not. He only encountered these conditions in the context of energy-momentum conservation. He did not encounter them in his analysis of the covariance of the action J. Einstein started from

$$\Delta J = \int \Delta H \sqrt{-g}\, d\tau \tag{34}$$

rather than from eq. (26). He did not bother to write down the coefficients of $\partial \Delta x^\sigma / \partial x^\nu$ in ΔH as we did for ΔQ (see eq. (28)). He wrote:

> We now assume that H is invariant under linear transformations, i.e., that ΔH should vanish if the [$\partial^2 \Delta x^\mu / \partial x^\alpha \partial x^\sigma$] vanish. On this assumption we arrive at

---

47 Eq, (24) for $\Delta g^{\mu\nu}$ follows from

$$\Delta g^{\mu\nu} = g'^{\mu\nu} - g^{\mu\nu} = \frac{\partial x'^\mu}{\partial x^\alpha}\frac{\partial x'^\nu}{\partial x^\beta} g^{\alpha\beta} - g^{\mu\nu} = \left(\delta_\alpha^\mu + \frac{\partial \Delta x^\mu}{\partial x^\alpha}\right)\left(\delta_\beta^\nu + \frac{\partial \Delta x^\nu}{\partial x^\beta}\right) g^{\alpha\beta} - g^{\mu\nu};$$

eq. (25) for $\Delta g_\sigma^{\mu\nu}$ from:

$$\begin{aligned}\Delta g_\sigma^{\mu\nu} &= g'^{\mu\nu}_\sigma - g_\sigma^{\mu\nu} = \frac{\partial x^\rho}{\partial x'^\sigma}\frac{\partial}{\partial x^\rho}\left(\frac{\partial x'^\mu}{\partial x^\alpha}\frac{\partial x'^\nu}{\partial x^\beta} g^{\alpha\beta}\right) - g_\sigma^{\mu\nu} \\ &= \left(\delta_\sigma^\rho - \frac{\partial \Delta x^\rho}{\partial x^\sigma}\right)\frac{\partial}{\partial x^\rho}\left(\left(\delta_\alpha^\mu + \frac{\partial \Delta x^\mu}{\partial x^\alpha}\right)\left(\delta_\beta^\nu + \frac{\partial \Delta x^\nu}{\partial x^\beta}\right) g^{\alpha\beta}\right) - g_\sigma^{\mu\nu}.\end{aligned}$$

48 The Jacobian can be computed as follows:

$$\left|\frac{\partial x'}{\partial x}\right| = \varepsilon_{\mu..\nu}\frac{\partial x'^\mu}{\partial x^1}..\frac{\partial x'^\nu}{\partial x^4} = \varepsilon_{\mu..\nu}\delta_1^\mu..\delta_4^\nu + \varepsilon_{\mu..\nu}\frac{\partial x'^\mu}{\partial x^1}..\delta_4^\nu + \ldots + \ldots + \varepsilon_{\mu..\nu}\delta_1^\mu..\frac{\partial x'^\nu}{\partial x^4} = 1 + \frac{\partial \Delta x^\mu}{\partial x^\mu}.$$

49 Eq. (28) is found as follows:

$$\begin{aligned}\Delta Q &= \frac{\partial Q}{\partial g^{\mu\nu}}\left(g^{\mu\alpha}\frac{\partial \Delta x^\nu}{\partial x^\alpha} + g^{\nu\alpha}\frac{\partial \Delta x^\mu}{\partial x^\alpha}\right) + \frac{\partial Q}{\partial g_\alpha^{\mu\nu}}\left(\frac{\partial}{\partial x^\alpha}(\Delta g^{\mu\nu}) - \frac{\partial \Delta x^\beta}{\partial x^\alpha} g_\beta^{\mu\nu}\right) \\ &= 2\frac{\partial Q}{\partial g^{\mu\nu}} g^{\mu\alpha}\frac{\partial \Delta x^\nu}{\partial x^\alpha} + \frac{\partial Q}{\partial g_\alpha^{\mu\nu}}\left(\frac{\partial}{\partial x^\alpha}\left(2 g^{\mu\beta}\frac{\partial \Delta x^\nu}{\partial x^\beta}\right) - \frac{\partial \Delta x^\beta}{\partial x^\alpha} g_\beta^{\mu\nu}\right) \\ &= 2\frac{\partial Q}{\partial g^{\mu\nu}} g^{\mu\alpha}\frac{\partial \Delta x^\nu}{\partial x^\alpha} + 2\frac{\partial Q}{\partial g_\alpha^{\mu\nu}} g_\alpha^{\mu\beta}\frac{\partial \Delta x^\nu}{\partial x^\beta} + 2\frac{\partial Q}{\partial g_\alpha^{\mu\nu}} g^{\mu\beta}\frac{\partial^2 \Delta x^\nu}{\partial x^\alpha \partial x^\beta} - \frac{\partial Q}{\partial g_\alpha^{\mu\nu}} g_\beta^{\mu\nu}\frac{\partial \Delta x^\beta}{\partial x^\alpha}.\end{aligned}$$

Grouping terms in $\partial \Delta x^\nu / \partial x^\alpha$ and in $\partial^2 \Delta x^\nu / \partial x^\alpha \partial x^\beta$ and relabeling indices, one arrives at eq. (28).

$$\left[\frac{1}{2}\Delta H = g^{\nu\alpha}\frac{\partial H}{\partial g^{\mu\nu}_{\sigma}}\frac{\partial^2\Delta x^{\mu}}{\partial x^{\alpha}\partial x^{\sigma}}\right]$$

With the help of [this equation and $\Delta(\sqrt{-g}d\tau) = 0$] one arrives at

$$\left[\frac{1}{2}\Delta J = \int d\tau g^{\nu\alpha}\frac{\partial H\sqrt{-g}}{\partial g^{\mu\nu}_{\sigma}}\frac{\partial^2\Delta x^{\mu}}{\partial x^{\alpha}\partial x^{\sigma}}\right]$$

and from this through partial integration at

$$\left[\frac{1}{2}\Delta J = \int d\tau(\Delta x^{\mu}B_{\mu}) + \text{surface terms}\right]^{50}$$

Using eq. (28) with H instead of Q and the invariance of $\sqrt{-g}d\tau$, one arrives at

$$\frac{1}{2}\Delta J = \int\left\{g^{\alpha\nu}\frac{\partial H}{\partial g^{\alpha\sigma}} + g^{\beta\nu}_{\alpha}\frac{\partial H}{\partial g^{\beta\sigma}_{\alpha}} - \frac{1}{2}g^{\alpha\beta}_{\sigma}\frac{\partial H}{\partial g^{\alpha\beta}_{\nu}}\right\}\frac{\partial\Delta x^{\sigma}}{\partial x^{\nu}}\sqrt{-g}d\tau$$
$$+\int g^{\rho\tau}\frac{\partial H}{\partial g^{\mu\tau}_{\sigma}}\frac{\partial^2\Delta x^{\mu}}{\partial x^{\rho}\partial x^{\sigma}}\sqrt{-g}d\tau.$$

The condition that the expression in curly brackets vanish guarantees that $\Delta J = 0$ for linear transformations. Comparing this expression to S^{ν}_{σ}, the coefficients of $\partial\Delta x^{\sigma}/\partial x^{\nu}$ in eq. (29), one sees that the condition read off of the expression for ΔJ above is equivalent but not identical to the conditions $S^{\nu}_{\sigma} = 0$. So even if Einstein actually did calculate the coefficients of $\partial\Delta x^{\sigma}/\partial x^{\nu}$ in ΔH and ΔJ, he would not have arrived at the conditions $S^{\nu}_{\sigma} = 0$ found in the context of energy-momentum conservation. Consequently, he would still not have realized that $S^{\nu}_{\sigma} = 0$ for any function that transforms as a scalar under general linear transformations and that these conditions therefore cannot be used to determine the Lagrangian.[51]

Almost a year went by before Einstein discovered his error. As he wrote to Lorentz in early October 1915:

> The invariant-theoretical method actually does not tell us more than the Hamiltonian principle when it comes to the determination of your function Q $(= H\sqrt{-g})$ [see Lorentz 1915, 763]. That I did not realize this last year is because I nonchalantly introduced the assumption on p. 1069 of my paper [Einstein 1914c] that H be an invariant under *linear* transformations.[52]

The conditions on Q coming from the “Hamiltonian principle” are presumably the ones coming from the requirement that expressions (13) and (19) for t^{μ}_{ν} are equal to one another. This is how Einstein derives the conditions $S^{\nu}_{\sigma} = 0$ in his letter to

50 “Wir nehmen nun an daß H bezüglich linearer Transformationen eine Invariante sei; d. h. ΔH soll verschwinden, falls die […] verschwinden. Unter dieser Voraussetzung enthalten wir […]. Mit hilfe von […] erhält man […] und hieraus durch partielle Integration […]” (Einstein 1914c, 1069–1070).

51 In (Einstein 1916c, 1113–1115), the conditions $S^{\nu}_{\sigma} = 0$ and $B_{\mu} = 0$ are derived in yet another way (see sec. 9, eqs. (117)–(122)).

Lorentz. He then adds: “This is also the condition for $[Q d\tau]$ being an invariant under linear transformations” (ibid.).[53]

We now turn from the conditions $S_\sigma^\nu = 0$ to the conditions $B_\mu = 0$. These are the conditions for “coordinates adapted to the gravitational field.”[54] or “adapted coordinates” for short. If some metric field $g^{\mu\nu}$ expressed in coordinates x^μ satisfies these conditions, the coordinates are called adapted to that field. Transformations from one adapted coordinate system to another are called “justified.”[55] Such transformations are not mappings of the form $x^\mu \rightarrow x'^\mu$, like ordinary coordinate transformations, but mappings of the form $(x^\mu, g^{\mu\nu}(x)) \rightarrow (x'^\mu, g'^{\mu\nu}(x'))$. Because of their dependence on the metric, Einstein, at Ehrenfest’s suggestion, called them “non-autonomous” transformations at one point.[56]

Einstein looked upon the conditions $B_\mu = 0$ for adapted coordinates as the coordinate restriction with which the *Entwurf* field equations could be extracted from unknown generally-covariant equations.[57] He must have been pleased to see that these coordinate restrictions follow from energy-momentum conservation. In March 1914, Einstein wrote a letter to Besso reporting on the results that would be published a few months later in (Einstein and Grossmann 1914b). He showed how the conditions $B_\mu = 0$ follow from the field equations and energy-momentum conservation (cf. eqs. (21)–(22)). The *Entwurf* field equations, he told Besso, “*hold in every frame of reference adapted to this condition*.”[58] Einstein claimed that this class of reference frames included all sorts of accelerated frames, including the important case of a rotating frame.[59] This is not true,[60] but that does not matter for our purposes. What is interesting for our story is the following passage from the draft of Besso’s reply:

> You already had the fundamental insight that the conservation laws represent the condition for positing an admissible coordinate system; but it did not appear to be ruled out that a restriction to Lorentz transformations was thereby essentially already given, so that nothing particularly interesting epistemologically comes out of it. Now everything is fundamentally completely satisfactory.[61]

---

52 “Thatsächlich liefert die invariantentheoretische Methode nicht mehr als das Hamilton’sche Prinzip wenn es sich um die Bestimmung der Ihrer Funktion Q $(= H\sqrt{-g})$ handelt. Dass ich dies letztes Jahr nicht merkte liegt daran, dass ich auf Seite 1069 meiner Abhandlung leichtsinnig die Voraussetzung einführte, H sei eine Invariante bezüglich linearer Transformationen.” Einstein to H. A. Lorentz, October 12, 1915 (CPAE 8, Doc. 129). The function Q was introduced in (Lorentz 1915, 763).

53 “Dies ist gleichzeitig die Bedingung dafür, dass $[Q d\tau]$ eine Invariante bezüglich linearer Substitutionen ist.”

54 “dem Gravitationsfeld angepaßte Koordinatensysteme” (Einstein 1914c, 1070). In (Einstein and Grossmann 1914b, 221), such coordinates are called “‘adapted’ to the manifold” (“der Mannigfaltigkeit „*angepaßte*“”).

55 “„*berechtigte*“” (Einstein and Grossmann 1914b, 221).

56 “„unselbständige“.” Einstein to H. A. Lorentz, 14 August 1913 (CPAE 5, Doc. 467). Non-autonomous transformations play an important role in the Zurich Notebook (see sec. 4.3 of “Commentary …” [in this volume]).

In the first sentence Besso is referring to Einstein's argument of late 1913 which seemed to show that energy-momentum conservation limits the covariance of the *Entwurf* field equations to linear transformations. Einstein had touted this specious result in several places. To Paul Ehrenfest, for instance, he wrote in November 1913, referring both to the argument from energy-momentum conservation and to the 'hole argument':

> The gravitation affair has been resolved to my *full satisfaction* (namely the circumstance that the equations of the gravitational field are only invariant under linear transformations[)]. It turns out that one can prove that *generally-covariant* equations that *fully* determine the field on the basis of the matter [energy-momentum] tensor cannot exist at all. What can be more beautiful than that the necessary specialization follows from the conservation laws?[62]

The argument that energy-momentum conservation restricts the covariance of the field equations to linear transformations evaporated early in 1914. But the triumphant rhetorical question in the passage above can also be applied to the argument leading to the condition $B_\mu = 0$ that took its place: "What can be more beautiful than that the necessary specialization follows from the conservation laws?"

Neither in (Einstein and Grossmann 1914b) nor in (Einstein 1914c) do we find statements drawing attention to the close connection between covariance of the field equations and energy-momentum conservation. Einstein probably did emphasize this connection though in his Wolfskehl lectures in Göttingen in the summer of 1915.[63] Afterwards, in two letters to his friend Heinrich Zangger,[64] Einstein expressed his satisfaction that he had been able to convince the Göttingen mathematicians, and David Hilbert in particular, of his *Entwurf* theory. In November 1915, Einstein found himself in a race against Hilbert to find field equations to replace the discarded *Entwurf* equa-

57 This fits with Einstein's general attitude towards general covariance at the time. In (Einstein 1914b, 177–178), he wrote: "When one has equations relating certain quantities that only hold in certain coordinate systems, one has to distinguish between two cases: 1. There are generally-covariant equations corresponding to the equations […]; 2. There are no generally-covariant equations that can be found on the basis of the equations given for a particular choice of reference frame. In case 2, the equations do not tell us anything about the things represented by these quantities; they only restrict the choice of reference frame. If the equations tell us anything at all about the things represented by these quantities, we are always dealing with case 1" ("Wenn Gleichungen zwischen irgendwelchen Größen gegeben sind, die nur bei spezieller Wahl des Koordinatensystems gültig sind, so sind zwei Fälle zu unterscheiden: 1. Es entsprechen den Gleichungen allgemein kovariante Gleichungen […]; 2. es gibt keine allgemein kovarianten Gleichungen, die aus den für spezielle Wahl des Bezugssystems gegebenen Gleichungen gefolgert werden können. Im Falle 2 sagen die Gleichungen über die durch die Größen dargestellten Dinge gar nichts aus; sie beschränken nur die Wahl des Bezugssystems. Sagen die Gleichungen über die durch die Größen dargestellten Dinge überhaupt etwas aus, so liegt stets der Fall 1 vor […]."). Similarly, he told Ehrenfest: "Grossmann wrote to me that he has now also been able to derive the gravitational [field] equations from the theory of general covariants. That would be a neat addition to our investigation [i.e., Einstein and Grossmann 1914b]" ("Grossmann schrieb mir, dass es ihm nun auch gelinge, die Gravitationsgleichungen aus der allgemeinen Kovariantentheorie abzuleiten. Es wäre dies eine hübsche Ergänzung zu unserer Untersuchung," Einstein to Paul Ehrenfest, before 10 April 1914 [CPAE 8, Doc. 2]).

tions. Page proofs of (Hilbert 1915),[65] the final version of which would not be published until late March 1916, show that the theory originally proposed by Hilbert has a structure that is remarkably similar to that of the *Entwurf* theory as presented in (Einstein and Grossmann 1914b) and (Einstein 1914c). In these page proofs, Hilbert introduces field equations that are invariant under arbitrary transformations of the coordinates—or, as Hilbert calls them, "world parameters" ("Weltparameter"). He then rehearses what is essentially Einstein's 'hole argument' to argue that these "world parameters" need to be restricted to what he calls "spacetime coordinates" ("Raum-Zeitkoordinaten"). Such spacetime coordinates are defined as those world parameters for which a condition called the "energy theorem" ("Energiesatz") holds.[66] It is probably because of these similarities between Hilbert's original theory and the *Entwurf* theory, that Einstein accused Hilbert of "nostrification" ("Nostrifikation") in another letter to Zangger.[67] This episode is interesting for our purposes since it provides circumstantial evidence for our conjecture that Einstein did mention in his Wolfskehl lectures that it should be possible to extract the *Entwurf* field equations from (unknown) generally-covariant equations by imposing the coordinate restriction $B_\mu = 0$ given by the demands of energy-momentum conservation.

---

58 "… *für jedes Bezugssystem gelten, welches dieser Bedingung angepasst ist*." Einstein to Michele Besso, ca. 10 March 1914 (CPAE 5, Doc. 514). Levi-Civita constructed a counter-example to Einstein's claim, Tullio Levi-Civita to Einstein, 28 March 1915 (CPAE 8, Doc. 67). In our notation, Levi-Civita found a non-autonomous transformation $(x^\mu, g^{\mu\nu}(x)) \rightarrow (x'^\mu, g'^{\mu\nu}(x'))$ satisfying the condition for justified transformations between adapted coordinates (i.e., $B_\mu(g^{\mu\nu}(x)) = B_\mu(g'^{\mu\nu}(x')) = 0$), under which the *Entwurf* field equations were nonetheless *not* invariant (i.e., $g^{\mu\nu}(x)$ is a solution but $g'^{\mu\nu}(x')$ is not). In Levi-Civita's example, $g^{\mu\nu}(x) = \eta^{\mu\nu} = \mathrm{diag}(-1, -1, -1, 1)$.

59 This claim was based on a general argument given in Einstein to H.A. Lorentz, 23 January 1915 (CPAE 8, Doc. 47).

60 For the Minkowski metric in rotating coordinates, $B_\mu \neq 0$ (Janssen 1999, 150–151, note 47)

61 "Du hattest schon principiell eingesehen, dass die Erhaltungssätze die Bedingung für die Aufstellung eines zulässigen Coordinatensystems darstellen; aber es schien nicht ausgeschlossen, dass schon dadurch, im Wesentlichen, die Beschränkung auf die Lorentztransformationen gegeben sei, so dass nichts erkenntnistheoretisch besonders interessantes herauskam. Nun ist alles principiell vollkommen befriedigend." Michele Besso to Einstein, draft, 20 March 1914 (CPAE 5, Doc. 516). For further discussion of this letter, see sec. 2.2 of "What Did Einstein Know …" (in this volume).

62 "Die Gravitationsaffäre hat sich zu meiner *vollen Befriedigung* aufgeklärt (der Umstand nämlich, dass die Gleichungen des Gr. Feldes nur linearen Transformationen gegenüber kovariant sind. Es lässt sich nämlich beweisen, dass *allgemein kovariante* Gleichungen, die das Feld aus dem materiellen Tensor vollständig bestimmen, überhaupt nicht existieren können. Was kann es schöneres geben, als dies, dass jene nötige Spezialisierung aus den Erhaltungssätzen fliesst?" Einstein to Paul Ehrenfest, before 7 November 1913 (CPAE 5, Doc. 481).

63 Notes taken by an unknown auditor present at (some of) these lectures were found by Leo Corry and are published in Appendix B of CPAE 6. These notes, however, do not touch on the field equations, nor on energy-momentum conservation.

4. THE MAXWELLIAN *ENTWURF* LAGRANGIAN

We feed the gravitational part of the Lagrangian for the Entwurf field equations, modelled on the Lagrangian for the free Maxwell field, into the general formalism of (Einstein 1914c) and derive the field equations, the expression for the gravitational energy-momentum pseudo-tensor, and the condition for "adapted coordinates" that determines for each solution what other coordinate representations of the solution are allowed by the Entwurf field equations.

One arrives at the gravitational part of the Lagrangian $Q = H\sqrt{-g}$ for the *Entwurf* field equations through the following choice for the function H (Einstein 1914c, 1076, note 1):[68]

$$H = -g^{\mu\nu}\Gamma^{\alpha}_{\beta\mu}\Gamma^{\beta}_{\alpha\nu}, \tag{35}$$

where $\Gamma^{\alpha}_{\beta\mu}$ are the components of the gravitational field, defined as (Einstein 1914, p. 1077, eq. 81a):

$$\Gamma^{\alpha}_{\beta\mu} \equiv \frac{1}{2}g^{\alpha\rho}g_{\rho\beta,\mu}. \tag{36}$$

The Lagrangian is modelled on the Lagrangian $-\frac{1}{4}F_{\mu\nu}F^{\mu\nu}$ for the free Maxwell field. Since H is a scalar under linear transformations, the conditions $S^{\nu}_{\sigma} = 0$ are satisfied (see eqs. (30)–(31)).

---

64 Einstein to Heinrich Zangger, 7 July 1915 (CPAE 8, Doc. 94) and between 24 July and 7 August 1915 (CPAE 8, Doc. 101).

65 These page proofs are located at the Niedersächsische Staats- und Universitätsbibliothek in Göttingen (Cod. Ms. D. Hilbert 634), where they were discovered by Leo Corry. For discussion, see (Corry et al. 1997), (Sauer 1999), and "Hilbert's Foundation of Physics …" (in vol. 4 of this series).

66 This restriction is stated in "Axiom III (Axiom of space and time)" ("Axiom III (Axiom von Raum und Zeit")) in the page proofs. This axiom is also mentioned in David Hilbert to Einstein, 13 November 1915 (CPAE 8, Doc. 140). It no longer occurs in (Hilbert 1915).

67 Einstein to Heinrich Zangger, 26 November 1915 (CPAE 8, Doc. 152). From Einstein to David Hilbert, 18 November 1915 (CPAE 8, Doc. 148) it can be inferred that Einstein saw a manuscript with an early version of the theory that would eventually be published in (Hilbert 1915).

68 The expression for H actually differs by a factor $1/2$ from the one that leads to the *Entwurf* field equations as given in Einstein's publications prior to (Einstein 1914c). Substituting eq. (36) for $\Gamma^{\rho}_{\mu\tau}$ into eq. (35), one can rewrite the function H as:

$$H = -g^{\mu\nu}\left(\frac{1}{2}g^{\alpha\sigma}g_{\sigma\beta,\mu}\right)\left(\frac{1}{2}g^{\beta\tau}g_{\tau\alpha,\nu}\right).$$

Since $g^{\alpha\sigma}g^{\beta\tau}g_{\tau\alpha,\nu} = -g^{\sigma\beta}_{\nu}$, this expression can also be written as (Einstein 1914, 1076, eq. 78):

$$H = \frac{1}{4}g^{\mu\nu}g_{\sigma\beta,\mu}g^{\sigma\beta}_{\nu}.$$

To recover the *Entwurf* field equations, the factor $1/4$ should be replaced by $1/2$ (see Einstein and Grossmann 1914, 219, eq.Va, and note 72 below). In other words, the function H should be defined as $H = -2g^{\mu\nu}\Gamma^{\alpha}_{\beta\mu}\Gamma^{\beta}_{\alpha\nu}$.

The definition of the components of the gravitational field is suggested by the energy-momentum balance equation, written in the form of eq. (), which with the help of eq. (36) can be rewritten as

$$\mathfrak{T}^{\alpha}_{\mu,\alpha} - \Gamma^{\beta}_{\alpha\mu}\mathfrak{T}^{\alpha}_{\beta} = 0.$$

The second term represents the gravitational force density and has the same form as the Lorentz force density, $f_\mu = F_{\mu\nu}j^\nu$. It is the contraction of the field $\Gamma^{\alpha}_{\beta\mu}$ and its source $\mathfrak{T}^{\alpha}_{\beta}$.

The quantities $\Gamma^{\alpha}_{\beta\mu}$ in eq. (36) are truncated versions of the Christoffel symbols[69]

$$\left\{\begin{matrix}\alpha\\ \beta\mu\end{matrix}\right\} = \frac{1}{2}g^{\alpha\rho}(g_{\rho\beta,\mu} + g_{\rho\mu,\beta} - g_{\beta\mu,\rho}).$$

Note that, unlike the Christoffel symbols, the $\Gamma^{\alpha}_{\beta\mu}$-s are not symmetric in their lower indices.

We derive the left-hand side of the *Entwurf* field equations from the action principle $\delta J = 0$. The gravitational part of the action is (see eq. (9))

$$J = \int Q d\tau,$$

with $Q = \sqrt{-g}H$. There are two contributions to δQ:

$$\delta Q = \sqrt{-g}\delta H + (\delta\sqrt{-g})H. \tag{37}$$

---

69 On p. 23L of the Zurich Notebook, Einstein had tried to extract field equations from the November tensor by truncating the Christoffel symbols in a similar fashion. Introducing the quantities

$$\vartheta_{il\alpha} \equiv \frac{1}{2}(g_{il,\alpha} + g_{l\alpha,i} + g_{\alpha i,l})$$

Einstein could write the Christoffel symbols as

$$\left\{\begin{matrix}k\\ il\end{matrix}\right\} = g^{k\alpha}(\vartheta_{il\alpha} - g_{il,\alpha}).$$

Inserting this expression into the November tensor,

$$T^x_{il} = \frac{\partial}{\partial x^k}\left\{\begin{matrix}k\\ il\end{matrix}\right\} - \left\{\begin{matrix}\lambda\\ ik\end{matrix}\right\}\left\{\begin{matrix}k\\ \lambda l\end{matrix}\right\}$$

(see eq. (5)), and eliminating all terms involving $\vartheta_{il\alpha}$ with the help of the appropriate coordinate restriction, Einstein arrived at the following candidate for the left-hand side of the field equations

$$-\frac{\partial}{\partial x^k}(g^{k\alpha}g_{il,\alpha}) - (g^{\lambda\alpha}g_{ik,\alpha})(g^{k\beta}g_{\lambda l,\beta}).$$

As Einstein realized, this expression can be obtained in one fell swoop by setting $\vartheta_{il\alpha} = 0$ and substituting $-g^{k\alpha}g_{il,\alpha}$ for the Christoffel symbols in the November tensor. Like all other candidates extracted from the Riemann tensor in the Zurich Notebook, this candidate was rejected because the necessary coordinate restriction turned out to be too restrictive. For a more detailed analysis, see sec. 5.5.4 of "Commentary …" (in this volume).

Since $\delta g = -g g_{\alpha\beta}\delta g^{\alpha\beta}$ (see Einstein 1914c, 1051, eqs. 32–34; Einstein 1916a, 796, eq. 29),

$$\delta\sqrt{-g} = -\frac{1}{2}\sqrt{-g}\,g_{\mu\nu}\delta g^{\mu\nu}. \tag{38}$$

Inserting this expression and eq. (35) for H into the second contribution to δQ in eq. (37), one arrives at:

$$(\delta\sqrt{-g})H = \left(\frac{1}{2}\sqrt{-g}\,g_{\mu\nu}g^{\rho\sigma}\Gamma^{\alpha}_{\beta\rho}\Gamma^{\beta}_{\alpha\sigma}\right)\delta g^{\mu\nu}. \tag{39}$$

Variation of H in the first contribution to δQ in eq. (37) gives:

$$\delta H = \delta(-g^{\mu\nu}\Gamma^{\alpha}_{\beta\mu}\Gamma^{\beta}_{\alpha\nu}) = -\delta g^{\mu\nu}\Gamma^{\alpha}_{\beta\mu}\Gamma^{\beta}_{\alpha\nu} - 2g^{\mu\nu}\Gamma^{\alpha}_{\beta\mu}\delta\Gamma^{\beta}_{\alpha\nu}.$$

For $\delta\Gamma^{\beta}_{\alpha\nu}$ one finds:[70]

$$\delta\Gamma^{\beta}_{\alpha\nu} = -\frac{1}{2}g_{\alpha\lambda}\delta g^{\beta\lambda}_{\nu} - g_{\alpha\rho}\Gamma^{\beta}_{\sigma\nu}\delta g^{\rho\sigma}.$$

It follows that:[71]

$$-2g^{\mu\nu}\Gamma^{\alpha}_{\beta\mu}\delta\Gamma^{\beta}_{\alpha\nu} = g^{\alpha\beta}g_{\rho\nu}\Gamma^{\rho}_{\mu\beta}\delta g^{\mu\nu}_{\alpha} + 2g^{\rho\tau}g_{\alpha\mu}\Gamma^{\alpha}_{\beta\rho}\Gamma^{\beta}_{\nu\tau}\delta g^{\mu\nu}.$$

Substituting this into the expression for δH above and collecting terms with $\delta g^{\mu\nu}$ and $\delta g^{\mu\nu}_{\alpha}$, one finds:

$$\delta H = (2g^{\rho\tau}g_{\alpha\mu}\Gamma^{\alpha}_{\beta\rho}\Gamma^{\beta}_{\nu\tau} - \Gamma^{\alpha}_{\beta\mu}\Gamma^{\beta}_{\alpha\nu})\delta g^{\mu\nu} + (g^{\alpha\beta}g_{\rho\nu}\Gamma^{\rho}_{\mu\beta})\delta g^{\mu\nu}_{\alpha}. \tag{40}$$

Inserting eqs. (39) and (40) into eq. (37), one finds:

$$\delta Q = \sqrt{-g}\Big[\Big(2g^{\rho\tau}g_{\alpha\mu}\Gamma^{\alpha}_{\beta\rho}\Gamma^{\beta}_{\nu\tau} - \Gamma^{\alpha}_{\beta\mu}\Gamma^{\beta}_{\alpha\nu} + \frac{1}{2}g_{\mu\nu}g^{\rho\sigma}\Gamma^{\alpha}_{\beta\rho}\Gamma^{\beta}_{\alpha\sigma}\Big)\delta g^{\mu\nu} + (g^{\alpha\beta}g_{\rho\nu}\Gamma^{\rho}_{\mu\beta})\delta g^{\mu\nu}_{\alpha}\Big].$$

Comparison of this expression with $\delta Q = \dfrac{\partial Q}{\partial g^{\mu\nu}}\delta g^{\mu\nu} + \dfrac{\partial Q}{\partial g^{\mu\nu}_{\alpha}}\delta g^{\mu\nu}_{\alpha}$ gives

---

70 Using eq. (36) for $\Gamma^{\beta}_{\alpha\nu}$, one finds:

$$\delta\Gamma^{\beta}_{\alpha\nu} = \frac{1}{2}\delta(g^{\beta\lambda}g_{\lambda\alpha,\nu}) = -\frac{1}{2}\delta(g^{\beta\lambda}_{\nu}g_{\lambda\alpha}) = -\frac{1}{2}\delta g^{\beta\lambda}_{\nu}g_{\lambda\alpha} - \frac{1}{2}g^{\beta\lambda}_{\nu}\delta g_{\lambda\alpha}.$$

Since $\delta g_{\alpha\lambda} = -g_{\alpha\rho}g_{\lambda\sigma}\delta g^{\rho\sigma}$, the last term can be rewritten as:

$$-\frac{1}{2}g^{\beta\lambda}_{\nu}\delta g_{\alpha\lambda} = \frac{1}{2}g^{\beta\lambda}_{\nu}g_{\alpha\rho}g_{\lambda\sigma}\delta g^{\rho\sigma} = -\frac{1}{2}g^{\beta\lambda}g_{\lambda\sigma,\nu}g_{\alpha\rho}\delta g^{\rho\sigma} = -g_{\alpha\rho}\Gamma^{\beta}_{\sigma\nu}\delta g^{\rho\sigma}.$$

71 This follows from

$$-2g^{\mu\nu}\Gamma^{\alpha}_{\beta\mu}\delta\Gamma^{\beta}_{\alpha\nu} = g^{\mu\nu}\Gamma^{\alpha}_{\beta\mu}(g_{\alpha\lambda}\delta g^{\beta\lambda}_{\nu}) + 2g^{\mu\nu}\Gamma^{\alpha}_{\beta\mu}(g_{\alpha\rho}\Gamma^{\beta}_{\sigma\nu}\delta g^{\rho\sigma})$$

after relabeling indices.

$$\frac{\partial Q}{\partial g^{\mu\nu}} = \sqrt{-g}\Big(2g^{\rho\tau}g_{\alpha\mu}\Gamma^{\alpha}_{\beta\rho}\Gamma^{\beta}_{\nu\tau} - \Gamma^{\alpha}_{\beta\mu}\Gamma^{\beta}_{\alpha\nu} + \frac{1}{2}g_{\mu\nu}g^{\rho\sigma}\Gamma^{\alpha}_{\beta\rho}\Gamma^{\beta}_{\alpha\sigma}\Big), \tag{41}$$

$$\frac{\partial Q}{\partial g^{\mu\nu}_{\alpha}} = \sqrt{-g}g^{\alpha\beta}g_{\rho\nu}\Gamma^{\rho}_{\mu\beta}. \tag{42}$$

Inserting equations (41)–(42) into the general form of the field equations,

$$\frac{\partial}{\partial x^{\alpha}}\Big(\frac{\partial Q}{\partial g^{\mu\nu}_{\alpha}}\Big) - \frac{\partial Q}{\partial g^{\mu\nu}} = -\kappa\mathfrak{T}_{\mu\nu} \tag{43}$$

(see eq. (11)), one can write the *Entwurf* field equations as:[72]

$$\begin{aligned}\frac{\partial}{\partial x^{\alpha}}(\sqrt{-g}g^{\alpha\beta}g_{\rho\nu}\Gamma^{\rho}_{\mu\beta}) - \sqrt{-g}\Big(2g^{\rho\tau}g_{\alpha\mu}\Gamma^{\alpha}_{\beta\rho}\Gamma^{\beta}_{\nu\tau} - \Gamma^{\alpha}_{\beta\mu}\Gamma^{\beta}_{\alpha\nu}\\ + \frac{1}{2}g_{\mu\nu}g^{\rho\sigma}\Gamma^{\alpha}_{\beta\rho}\Gamma^{\beta}_{\alpha\sigma}\Big) = -\kappa\mathfrak{T}_{\mu\nu}.\end{aligned} \tag{44}$$

The *Entwurf* field equations can be written in a more compact form (Einstein 1914c, 1077). Instead of using eq. (43), one can use the field equations in the general form

72 Expressing $\Gamma^{\alpha}_{\mu\nu}$ in terms of $g_{\mu\nu}$ and multiplying the left-hand side of eq. (44) by 2 to correct for the error in eq. (35) for H (see note 68), one recovers the *Entwurf* field equation as originally given in (Einstein and Grossmann 1913). The first two terms on the left-hand side of eq. (44) can be written as:

$$\frac{\partial}{\partial x^{\alpha}}\Big(\sqrt{-g}g^{\alpha\beta}g_{\rho\nu}\Big[\frac{1}{2}g^{\rho\lambda}g_{\lambda\mu,\beta}\Big]\Big) - 2\sqrt{-g}g^{\rho\tau}g_{\alpha\mu}\Big(\frac{1}{2}g^{\alpha\sigma}g_{\sigma\beta,\rho}\Big)\Big(\frac{1}{2}g^{\beta\lambda}g_{\lambda\nu,\tau}\Big)$$
$$= \frac{1}{2}\sqrt{-g}\Big\{\frac{1}{\sqrt{-g}}\frac{\partial}{\partial x^{\alpha}}(\sqrt{-g}g^{\alpha\beta}g_{\mu\nu,\beta}) - g^{\alpha\beta}g^{\tau\rho}g_{\mu\tau,\alpha}g_{\nu\rho,\beta}\Big\}.$$

The expression in curly brackets is the quantity $D_{\mu\nu}(g)$ defined in (Einstein and Grossmann 1913, 16, eq. 16). The last two terms on the left-hand side eq. (44) can likewise be written as:

$$\sqrt{-g}\Big(\Big[\frac{1}{2}g^{\alpha\rho}g_{\rho\beta,\mu}\Big]\Big[\frac{1}{2}g^{\beta\sigma}g_{\sigma\alpha,\nu}\Big] - \frac{1}{2}g_{\mu\nu}g^{\rho\sigma}\Big[\frac{1}{2}g^{\alpha\tau}g_{\tau\beta,\rho}\Big]\Big[\frac{1}{2}g^{\beta\lambda}g_{\lambda\alpha,\sigma}\Big]\Big)$$
$$= -\frac{1}{2}\sqrt{-g}\Big\{\frac{1}{2}\Big[g^{\alpha\beta}_{\mu}g_{\alpha\beta,\nu} - \frac{1}{2}g_{\mu\nu}g^{\alpha\beta}g^{\tau\rho}_{\alpha}g_{\tau\rho,\beta}\Big]\Big\}.$$

The expression in square brackets is the quantity $-2\kappa t_{\mu\nu}$ in (Einstein and Grossmann 1913, 16, eq. 14). The left-hand side of eq. (44) can thus be rewritten as:

$$\frac{1}{2}\sqrt{-g}\{D_{\mu\nu}(g) + \kappa t_{\mu\nu}\}.$$

Omitting the erroneous factor $1/2$ and dividing by $\sqrt{-g}$, one sees that eq. (44) can be rewritten as

$$-D_{\mu\nu}(g) = \kappa(t_{\mu\nu} + T_{\mu\nu}),$$

which is just (Einstein and Grossmann 1913, 17, eq. 21).

$$\frac{\partial}{\partial x^{\alpha}}\left(g^{\nu\lambda}\frac{\partial Q}{\partial g^{\mu\nu}_{\alpha}}\right) = -\kappa(\mathfrak{T}^{\lambda}_{\mu} + t^{\lambda}_{\mu}) \tag{45}$$

(see eq. (13)), along with eq. (18) for the gravitational energy-momentum tensor[73]

$$\kappa t^{\lambda}_{\mu} = \frac{1}{2}\left(\delta^{\lambda}_{\mu}Q - g^{\alpha\beta}_{\mu}\frac{\partial Q}{\partial g^{\alpha\beta}_{\lambda}}\right). \tag{46}$$

Using expression (35) for H in $Q = H\sqrt{-g}$, one recovers the *Entwurf* field equations as given by Einstein. From eq. (42) it follows that eq. (45) in this case is:[74]

$$\frac{\partial}{\partial x^{\alpha}}(\sqrt{-g}g^{\alpha\beta}\Gamma^{\lambda}_{\mu\beta}) = -\kappa(\mathfrak{T}^{\lambda}_{\mu} + t^{\lambda}_{\mu}). \tag{47}$$

From eqs. (35) and (42) it follows that eq. (46) in this case is:

$$\kappa t^{\lambda}_{\mu} = \frac{1}{2}(\delta^{\lambda}_{\mu}\sqrt{-g}[-g^{\rho\tau}\Gamma^{\alpha}_{\beta\rho}\Gamma^{\beta}_{\alpha\tau}] - g^{\alpha\beta}_{\mu}[\sqrt{-g}g^{\lambda\rho}g_{\tau\beta}\Gamma^{\tau}_{\alpha\rho}]). \tag{48}$$

Simplifying this expression, one finds[75]

$$\kappa t^{\lambda}_{\mu} = \sqrt{-g}\left(g^{\lambda\rho}\Gamma^{\alpha}_{\tau\mu}\Gamma^{\tau}_{\alpha\rho} - \frac{1}{2}\delta^{\lambda}_{\mu}g^{\rho\tau}\Gamma^{\alpha}_{\beta\rho}\Gamma^{\beta}_{\alpha\tau}\right). \tag{49}$$

This is indeed the expression for the gravitational energy-momentum pseudo-tensor as given in (Einstein 1914c, 1077, eq. 81b). And with this expression for t^{λ}_{μ}, the field equations (47) are indeed the *Entwurf* field equations as given in (Einstein 1914c, 1077, eq. 81).[76]

From eq. (47) it follows that the conditions $B_{\mu} = 0$ —playing the dual role of restricting the coordinates (see eq. (32)) and guaranteeing the vanishing of the divergence of $t^{\lambda}_{\mu} + \mathfrak{T}^{\lambda}_{\mu}$ (see eqs. (21)–(22))—take on the specific form:

$$B_{\mu} = \frac{\partial^2}{\partial x^{\alpha}\partial x^{\lambda}}(\sqrt{-g}g^{\alpha\beta}\Gamma^{\lambda}_{\mu\beta}) = 0. \tag{50}$$

---

73 This expression is simpler than expression (12), which depends both on $\partial Q/\partial g^{\alpha\beta}_{\lambda}$ and on $\partial Q/(\partial g^{\mu\nu}.)$ Since $S^{\nu}_{\sigma} = 0$, expressions (12) and (18) are equivalent (see eq. (18)–(20)).

74 In detail: $\frac{\partial}{\partial x^{\alpha}}(g^{\nu\lambda}[\sqrt{-g}g^{\alpha\beta}g_{\rho\nu}\Gamma^{\rho}_{\mu\beta}]) = \frac{\partial}{\partial x^{\alpha}}(\sqrt{-g}g^{\alpha\beta}\delta^{\lambda}_{\rho}\Gamma^{\rho}_{\mu\beta}) = \frac{\partial}{\partial x^{\alpha}}(\sqrt{-g}g^{\alpha\beta}\Gamma^{\lambda}_{\mu\beta})$

75 The last term in eq. (48) can be rewritten as:

$$-\frac{1}{2}\sqrt{-g}g^{\alpha\beta}_{\mu}g^{\lambda\rho}g_{\tau\beta}\Gamma^{\tau}_{\alpha\rho} = \frac{1}{2}\sqrt{-g}g^{\alpha\beta}g^{\lambda\rho}g_{\tau\beta,\mu}\Gamma^{\tau}_{\alpha\rho} = \sqrt{-g}g^{\lambda\rho}\Gamma^{\alpha}_{\tau\mu}(\Gamma_{\alpha\rho},)^{\tau}$$

where in the last step eq. (36) was used.

76 To obtain the *Entwurf* field equations of Einstein's earlier publications, one needs to multiply both the left-hand side of eq. (47) and the right-hand sides of eqs. (48)–(49) by 2 (see notes 68 and 72).

Commenting on eq. (36) for the gravitational field, eq. (47), the *Entwurf* field equations, and eq. (49) for the gravitational energy-momentum pseudo-tensor, Einstein wrote

> Despite its complexity, the system of equations admits of a simple physical interpretation. The left-hand side [of eq. (47)] expresses a kind of divergence of the gravitational field [eq. (36)]. This [divergence] is—as the right-hand side shows—determined by the components of the total energy tensor. What is very important is the result that the energy tensor of the gravitational field [eq. (49)] acts as a source of the field in the same way as the energy tensor of matter.[77]

On the preceding page, Einstein boasted that his new derivation of the *Entwurf* field equations is essentially free from physical considerations. After showing that the expression for H in eq. (35) satisfies the conditions $S_\sigma^\nu = 0$ —as would any other expression transforming as a scalar under linear transformations—he wrote:

> We have now in a completely formal manner, i.e., without direct use of our physical knowledge about gravity, arrived at very definite field equations.[78]

Even if we forget for a moment that the uniqueness argument immediately preceding it is hogwash, this statement is patently false. The derivation of the *Entwurf* field equations in (Einstein 1914c, part D) may be more formal than earlier derivations, but it still relies heavily on physical considerations. The function H giving the Lagrangian is modelled on the Lagrangian for the free Maxwell field. It is assumed to depend only on first-order derivatives of the metric because the Poisson equation of Newtonian theory suggests that the gravitational field equations do not contain anything higher than second-order derivatives of the metric (Einstein and Grossmann 1913, 11). The conditions $S_\sigma^\nu = 0$ that supposedly determine H uniquely are derived from the energy-momentum balance law of matter in a gravitational field (see

77 "Das Gleichungssystem […] läßt trotz seiner Kompliziertheit eine einfache physikalische Interpretation zu. Die linke Seite drückt eine Art Divergenz des Gravitationsfeldes aus. Diese wird—wie die rechte Seite zeigt—bedingt durch die Komponente des totalen Energietensors. Sehr wichtig ist dabei das Ergebnis, daß der Energietensor des Gravitationsfeldes selbst in gleicher Weise felderregend wirksam ist wie der Energietensor der Materie" (Einstein 1914c, 1077).

78 "Wir sind nun auf rein formalem Wege, d. h. ohne direkte Heranziehung unserer physikalischen Kenntnisse von der Gravitation, zu ganz bestimmten Feldgleichungen gelangt" (Einstein 1914c, 1076). Einstein had already announced this proudly in the introduction of the paper: "In particular, it was possible to obtain the equations for the gravitational field in a purely covariant-theoretical way" ("Es gelang insbesondere, die Gleichungen des Gravitationsfeldes auf einem rein kovarianten-theoretischen Wege zu gewinnen;" ibid., 1030). Earlier in 1914, in a paper co-authored with Adriaan D. Fokker, Einstein had made a similar claim for his reformulation of the Nordström theory in terms of Riemannian geometry. In the conclusion of their paper, the authors wrote: "In the foregoing it was possible to show that, if one bases oneself on the principle of the constancy of the velocity of light, one can arrive at Nordström's theory by purely formal considerations, i.e., without recourse to additional physical hypotheses" ("Im vorstehenden konnte gezeigt werden, daß man bei Zugrundelegung des Prinzips von der Konstanz der Lichtgeschwindigkeit durch rein formale Erwägungen, d.h. ohne Zuhilfenahme weiterer physikalischen Hypothesen zur Nordströmschen Theorie gelangen kann;" Einstein and Fokker 1914, 328).

sec. 3.2). Moreover, H only gives the gravitational part of the field equations. The matter part, $\mathfrak{T}_{\mu}^{\lambda}$, is inserted on the basis of physical considerations. The same is true for the way in which the gravitational part of the field equations is split into a term with the divergence of the gravitational field and a term with the gravitational energy-momentum pseudo-tensor. All this is hard to reconcile with Einstein's claim to have derived the equations "in a completely formal manner."[79]

The claim, we suggest, should be understood against the backdrop of Einstein's obvious satisfaction that physical and mathematical considerations now seemed to point to the same field equations. Material in the Zurich Notebook shows that when Einstein began to generate field equations from their weak-field form by imposing energy-momentum conservation—the method that originally gave him the *Entwurf* field equations—he also tried to recover the resulting equations from the November tensor.[80] Such an alternative derivation of the field equations would have thrown light on their covariance properties. What Einstein presents in the review article of 1914 amounts to the same thing. Although he still had not found any connection between the *Entwurf* field equations and the Ricci tensor or the November tensor, he did supplement the physical considerations in the derivation of the *Entwurf* field equations by mathematical considerations that clarify—or so Einstein thought—their covariance properties. It therefore need not surprise us that Einstein overrated the importance of mathematical considerations in his new derivation of the *Entwurf* field equations.

5. A 'FATEFUL PREJUDICE' AND THE 'KEY TO THE SOLUTION': FROM THE *ENTWURF* LAGRANGIAN TO THE NOVEMBER LAGRANGIAN

When in the Entwurf Lagrangian the gravitational field is redefined as minus the Christoffel symbols, we find new field equations that bear a striking resemblance to field equations based on the November tensor found in the Zurich Notebook. When a factor $\sqrt{-g}$ is omitted in the action with the new definition of the gravitational field, the field equations are exactly the same as these equations in the Zurich Notebook. Einstein called the old definition of the gravitational field a "fateful prejudice" and the new definition "the key to the solution." This strongly suggests that the way in which Einstein found his way back to these discarded field equations of the Zurich Notebook shortly before he published them in (Einstein 1915a) was essentially the same as the way in which they are recovered in this section.

What are the field equations if one retains the form of (the gravitational part of) the *Entwurf* Lagrangian $Q = H\sqrt{-g}$, with

79 Einstein likewise overestimated the importance of purely mathematical considerations in deriving the field equations of the Nordström theory in (Einstein and Fokker 1914).

80 See pp. 24R–25R of the Zurich Notebook and sec. 5.6 of "Commentary …" (in this volume) for detailed analysis.

$$H = -g^{\mu\nu}\Gamma^{\alpha}_{\beta\mu}\Gamma^{\beta}_{\alpha\nu} \tag{51}$$

(see eq. (35)), but changes the definition of the components of the gravitational field from

$$\Gamma^{\alpha}_{\beta\mu} \equiv \frac{1}{2}g^{\alpha\rho}g_{\rho\beta,\mu} \tag{52}$$

(see eq. (36)) to

$$\Gamma^{\alpha}_{\beta\mu} \equiv -\begin{Bmatrix}\alpha\\ \beta\mu\end{Bmatrix} = -\frac{1}{2}g^{\alpha\rho}(g_{\rho\beta,\mu} + g_{\rho\mu,\beta} - g_{\beta\mu,\rho}), \tag{53}$$

as Einstein did in his first November 1915 paper?

As before (see eqs. (35)-(44)), the left-hand side of the field equations follows from $\delta J = 0$, where

$$J = \int Q d\tau = \int H\sqrt{-g}d\tau. \tag{54}$$

Variation of Q gives two contributions (see eq. (37)):

$$\delta Q = \sqrt{-g}\delta H + (\delta\sqrt{-g})H. \tag{55}$$

The second contribution is the same as before (see eq. (39)):

$$(\delta\sqrt{-g})H = \left(\frac{1}{2}\sqrt{-g}g_{\mu\nu}g^{\rho\sigma}\Gamma^{\alpha}_{\beta\rho}\Gamma^{\beta}_{\alpha\sigma}\right)\delta g^{\mu\nu}. \tag{56}$$

Likewise, δH can once again be written as

$$\delta H = -\delta g^{\mu\nu}\Gamma^{\alpha}_{\beta\mu}\Gamma^{\beta}_{\alpha\nu} - 2g^{\mu\nu}\Gamma^{\alpha}_{\beta\mu}\delta(\Gamma_{\alpha\nu\cdot})^{\beta}$$

However, since $\Gamma^{\alpha}_{\beta\mu}$ in eq. (53) is symmetric in its lower indices whereas $\Gamma^{\alpha}_{\beta\mu}$ in eq. (52) is not, the expression for δH ends up being much simpler than before (cf. eq. (40)). The expression for δH above can be rewritten as

$$\delta H = \Gamma^{\alpha}_{\beta\mu}\Gamma^{\beta}_{\alpha\nu}\delta g^{\mu\nu} - 2\Gamma^{\alpha}_{\beta\mu}\delta(g^{\mu\nu}\Gamma^{\beta}_{\alpha\nu}),$$

which reduces to:[81]

$$\delta H = \Gamma^{\alpha}_{\beta\mu}\Gamma^{\beta}_{\alpha\nu}\delta g^{\mu\nu} - \Gamma^{\alpha}_{\mu\nu}\delta g^{\mu\nu}_{\alpha}. \tag{57}$$

---

81 Using the definition of $\Gamma^{\beta}_{\alpha\nu}$ in eq. (53), one can rewrite the last term of the expression above as

$$-\Gamma^{\alpha}_{\beta\mu}\delta(g^{\mu\nu}g^{\beta\lambda}(g_{\lambda\alpha,\nu} + g_{\lambda\nu,\alpha} - g_{\alpha\nu,\lambda}))$$

Since $\Gamma^{\alpha}_{\beta\mu}g^{\mu\nu}g^{\beta\lambda}$ is symmetric in λ and ν and $g_{\lambda\alpha,\nu} - g_{\nu\alpha,\lambda}$ is anti-symmetric in λ and ν, their contraction vanishes and the expression above reduces to:

$$-\Gamma^{\alpha}_{\beta\mu}\delta(g^{\mu\nu}g^{\beta\lambda}g_{\lambda\nu,\alpha}).$$

Using that $g^{\mu\nu}g^{\beta\lambda}g_{\lambda\nu,\alpha} = -g^{\mu\beta}_{\alpha}$, one can rewrite this as $\Gamma^{\alpha}_{\mu\nu}\delta g^{\mu\nu}_{\alpha}$.

Inserting eqs. (56) and (57) into eq. (55), one finds:

$$\delta Q = \sqrt{-g}\Big[\Big(\Gamma^{\alpha}_{\beta\mu}\Gamma^{\beta}_{\alpha\nu} + \frac{1}{2}g_{\mu\nu}g^{\rho\sigma}\Gamma^{\alpha}_{\beta\rho}\Gamma^{\beta}_{\alpha\sigma}\Big)\delta g^{\mu\nu} - \Gamma^{\alpha}_{\mu\nu}\delta g^{\mu\nu}_{\alpha}\Big]. \tag{58}$$

It follows that

$$\frac{\partial Q}{\partial g^{\mu\nu}} = \sqrt{-g}\Gamma^{\alpha}_{\beta\mu}\Gamma^{\beta}_{\alpha\nu} + \frac{1}{2}\sqrt{-g}g_{\mu\nu}g^{\rho\sigma}\Gamma^{\alpha}_{\beta\rho}\Gamma^{\beta}_{\alpha\sigma}, \tag{59}$$

$$\frac{\partial Q}{\partial g^{\mu\nu}_{\alpha}} = -\sqrt{-g}\Gamma^{\alpha}_{\mu\nu}. \tag{60}$$

Inserting equations (59)–(60) into eq. (43), one finds the field equations:

$$\frac{\partial}{\partial x^{\alpha}}(\sqrt{-g}\Gamma^{\alpha}_{\mu\nu}) + \sqrt{-g}\Gamma^{\alpha}_{\beta\mu}\Gamma^{\beta}_{\alpha\nu} + \frac{1}{2}\sqrt{-g}g_{\mu\nu}g^{\rho\sigma}\Gamma^{\alpha}_{\beta\rho}\Gamma^{\beta}_{\alpha\sigma} = \kappa\mathfrak{T}_{\mu\nu}. \tag{61}$$

If one omits the factors $\sqrt{-g}$ in the first two terms and uses eq. (53) for the components of the gravitational field, these two terms become:

$$-\frac{\partial}{\partial x^{\alpha}}\begin{Bmatrix}\alpha\\ \mu\nu\end{Bmatrix} + \begin{Bmatrix}\alpha\\ \beta\mu\end{Bmatrix}\begin{Bmatrix}\beta\\ \nu\alpha\end{Bmatrix}.$$

This is just minus the November tensor which Einstein had extracted from the Ricci tensor in the Zurich Notebook by imposing the restriction to unimodular transformations. It is not hard to see how the calculation in eqs. (54)–(61) needs to be modified in order to recover the field equations (6) based on the November tensor without any of the additional extra terms and factors in eq. (61). First, one only requires J to transform as a scalar under unimodular transformations whenever H does. One can then start from

$$J \equiv -\int H d\tau. \tag{62}$$

Before (see eq. (54)) a factor $\sqrt{-g}$ was needed because only the combination $\sqrt{-g}d\tau$ is an invariant volume element under arbitrary transformations. Under unimodular transformations, however, $d\tau$ by itself is invariant.

It turns out (see sec. 6) that omission of a factor $\sqrt{-g}$ in the action seriously complicates the use of the formalism of (Einstein 1914c). It would have been easier to work with the field equations (61) retaining all factors of $\sqrt{-g}$. The covariance properties of these equations, however, look as intractable as those of the *Entwurf* equations. Omission of a factor $\sqrt{-g}$ was a small price to pay for field equations with a broad well-defined covariance group closely connected to the Riemann tensor.[82] This was the connection Einstein had been looking for in vain in the days of the Zurich Notebook. In (Einstein 1914c) he thought that such a connection was no longer needed, that instead he could supplement the physical considerations going into the derivation of the *Entwurf* field equations with mathematical ones establishing, or so it seemed, that their covariance was broad enough for the generalization of the relativity principle he envisioned. That approach had ultimately failed. The action (62) now promised to resurrect the old ideal of the Zurich Notebook in which physical and mathematical considerations would point to the same field equations. Complications coming from omitting a factor $\sqrt{-g}$ could be dealt with later.

Eq. (62) is indeed the action from which the field equations are derived in (Einstein 1915a, 784, eq. 17). Einstein used the notation $\mathfrak{L}$ for $-H$. $\mathfrak{L}$ is given by (see eq. (51)):

$$\mathfrak{L} = -H = g^{\mu\nu}\Gamma^{\alpha}_{\beta\mu}\Gamma^{\beta}_{\alpha\nu}, \tag{63}$$

The field equations are:

$$\frac{\partial}{\partial x^{\alpha}}\left(\frac{\partial\mathfrak{L}}{\partial g^{\mu\nu}_{\alpha}}\right) - \frac{\partial\mathfrak{L}}{\partial g^{\mu\nu}} = -\kappa T_{\mu\nu} \tag{64}$$

(Einstein 1915a, 784, eq. 18). The variation $\delta\mathfrak{L} = -\delta H$ can be read off of eq. (57). It follows that (Einstein 1915a, 784, eqs. 19–19a):[83]

$$\frac{\partial\mathfrak{L}}{\partial g^{\mu\nu}} = -\Gamma^{\alpha}_{\beta\mu}\Gamma^{\beta}_{\alpha\nu}, \tag{65}$$

$$\frac{\partial\mathfrak{L}}{\partial g^{\mu\nu}_{\alpha}} = \Gamma^{\alpha}_{\mu\nu}. \tag{66}$$

Inserting eqs. (65)–(66) into eq. (64), one finds the field equations (Einstein 1915a, 783, eq. 16a):

$$\Gamma^{\alpha}_{\mu\nu,\alpha} + \Gamma^{\alpha}_{\beta\mu}\Gamma^{\beta}_{\alpha\nu} = -\kappa T_{\mu\nu}. \tag{67}$$

When minus the Christoffel symbols are substituted for the quantities $\Gamma^{\alpha}_{\beta\mu}$, eq. (67) turns into eq. (6) based on the "November" tensor familiar from the Zurich Notebook. These field equations replace the *Entwurf* equations in (Einstein 1915a).

By changing the definition of the gravitational field in the *Entwurf* Lagrangian and by restricting the variational formalism of sec. 3 to unimodular transformations,

82 Another advantage was that unimodular transformations are autonomous. Einstein had been greatly relieved when, in August 1913, he hit upon the (fallacious) argument that energy-momentum conservation limited the covariance of the *Entwurf* equations to linear transformations. That meant that he could stop searching for non-linear non-autonomous transformation leaving the equations invariant. He had been unable to find a single one up to that point (Einstein to Lorentz, 14 and 16 August 1913 [CPAE 5, Docs. 467 and 470]; for further discussion, see sec. 3 of "What Did Einstein Know …" [in this volume]). After the argument had evaporated, Einstein had been forced to reconsider non-autonomous transformations. He eventually concluded that the condition $B_{\mu} = 0$ for the *Entwurf* theory (see eq. (50)) allows non-autonomous transformations to arbitrarily moving systems. The simple case of rotation in Minkowski spacetime proved him wrong. With field equations invariant under unimodular transformations, Einstein could avoid these problematic non-autonomous transformations altogether.

we have thus arrived at field equations with a clearly defined covariance group. Einstein's comments that the old definition (52) of the gravitational field was a "fateful prejudice" (cf. note 38) and that the new definition (53) was "the key to the solution" (cf. note 39) provide strong textual evidence that he found his way back to the field equations (67) in essentially the same way in which they were derived in this section.

This provides a remarkably simple solution to one of the central puzzles in reconstructing Einstein's path to the Einstein field equations. As John Norton (1984, 142) put it: "Why Einstein should choose [the November tensor, i.e., the left-hand side of eq. (67)] as his gravitation tensor rather than a generally-covariant tensor, such as the Ricci tensor or even the Einstein tensor itself, has hitherto been a puzzle." Norton conjectured that it was Einstein's prejudice about the form of the metric for weak static fields that prevented him from choosing the Ricci tensor. To reduce the Ricci tensor to the d'Alembertian acting on the metric in the case of weak fields, the argument went, one needs the harmonic coordinate condition, which is not satisfied by Einstein's metric for weak static fields. In the case of the November tensor, one can use the Hertz condition for this purpose (see eqs. (7)-(8)), which is satisfied by Einstein's metric for weak static fields. Aside from solving Norton's puzzle, there is no evidence that the incompatibility between the harmonic condition and Einstein's prejudice about the form of the metric for weak static fields played a role at this juncture. Our alternative solution to the puzzle removes the need for invoking this incompatibility.[84] Why did Einstein choose the November tensor rather than the Ricci tensor? Because both the mathematical and the physical strategy he employed in his search for suitable field equations pointed to the November tensor, not to the Ricci tensor.

---

83 The operations 'doing the variations' and 'setting $\sqrt{-g} = 1$' do not commute. Setting $\sqrt{-g} = 1$ in eqs. (59)–(60)—i.e., *after* doing the variations in eqs. (54)–(58)—does not reduce these equations to eqs. (65)-(66), which are obtained by setting $\sqrt{-g} = 1$ in eq. (54)—i.e., *before* doing the variations. If the condition $\sqrt{-g} = 1$ is imposed first, the variation is done under a constraint (Kichenassamy 1993, 197). This is the analogue in functional analysis of the problem in ordinary calculus of finding the extrema of a function under a constraint. Such problems can be replaced by the problem of finding the extrema of a related function(al) without constraints through the well-known technique of Lagrange multipliers. Einstein was familiar with these techniques for functionals from his work in statistical mechanics. To derive various ensembles in statistical mechanics (micro-canonical, canonical, or grand canonical), one maximizes the entropy under two constraints, one on the total energy and one on the total particle number. The difference with the case of varying the action for the metric field is that the constraint $\sqrt{-g(x)} = 1$ has to be imposed at every point x, so that there is an infinite number of constraints. The Lagrange multipliers thus become a new field. Techniques for doing this have been worked out in the context of what has come to be known as unimodular gravity, a theory first proposed in (Einstein 1919) and first cast in Lagrangian form in (Anderson and Finkelstein 1971).

84 Norton (1984, 102) invoked this same incompatibility to explain why Einstein abandoned field equations based on the Ricci tensor in the Zurich Notebook. As we mentioned earlier (see note 12), we see no evidence that it played any role there either.

6. THE FIRST NOVEMBER PAPER: THE KNOT UNTIED

In the Zurich Notebook Einstein had not been able to show that field equations based on the November tensor are compatible with energy-momentum conservation. In 1915 the variational formalism of (Einstein 1914c) showed him how to solve this problem. He essentially just had to find the conditions $B_\mu = 0$ *for this specific Lagrangian. As we saw in sec. 3, such conditions also determine the covariance properties of the field equations. Because of the way in which the November tensor can be extracted from the Ricci tensor, it is clear that the field equations based on the November tensor are invariant under unimodular transformations. One would therefore expect that in this case the four conditions* $B_\mu = 0$ *reduce to one condition expressing the restriction to unimodular transformations. The four conditions can indeed be replaced by one, but this one condition says that* $\sqrt{-g}$ *can not be a constant. This is more restrictive than the condition that* $\sqrt{-g}$ *transform as a scalar. It was nonetheless an important result that the compatibility of the field equations with energy-momentum conservation only called for one additional condition. It still takes four conditions to show that the relevant component of the field equations reduces to the Poisson equation for weak static fields. We suggest that this made it clear to Einstein that he could use coordinate conditions in the modern sense to recover the Poisson equation and that a coordinate restriction was needed only for energy-momentum conservation.*

According to the general formalism of (Einstein 1914c), the gravitational field equations are compatible with energy-momentum conservation if the conditions $S_\sigma^\nu = 0$ and $B_\mu = 0$ are satisfied (see eqs. (19)–(23)). These conditions, however, were derived for an action of the form $\int \sqrt{-g} H d\tau$. The field equations of (Einstein 1915a) were derived from an action of the form $\int \mathfrak{L} d\tau$ without the factor $\sqrt{-g}$ (see eqs. (62)–(63)). This seriously complicates matters (see notes 83 and 88) and in his papers of November 1915, as Einstein realized, he could not simply apply the formalism. He nonetheless relied heavily on the formalism to guide him in his analysis of the new theory. In (Einstein 1915a, 784–785), for instance, he went through a calculation closely analogous to the one in the general formalism (see sec. 3.2, eqs. (16)–(22)) to establish that the field equations derived from $\int \mathfrak{K} d\tau$ are compatible with energy-momentum conservation under the restriction to unimodular transformations.

Einstein (1915a, sec. 1–2) first exploited the restriction to unimodular transformations to replace the energy-momentum balance equation, $T^\lambda_{\sigma;\lambda} = 0$, by a simpler equation. The covariant divergence of $T_{\mu\nu}$, the energy-momentum tensor of matter, is given by (see note 35):

$$T^\lambda_{\sigma;\lambda} = T^\lambda_{\sigma,\lambda} + \begin{Bmatrix} \lambda \\ \mu\lambda \end{Bmatrix} T^\mu_\sigma - \begin{Bmatrix} \mu \\ \sigma\lambda \end{Bmatrix} T^\lambda_\mu. \tag{68}$$

The second term on the right-hand side can be rewritten as (see note 35):

$$\begin{Bmatrix} \lambda \\ \mu\lambda \end{Bmatrix} T^{\mu}_{\sigma} = (\log\sqrt{-g})_{,\mu} T^{\mu}_{\sigma};$$

the third term as (see note 36):

$$\begin{Bmatrix} \mu \\ \sigma\lambda \end{Bmatrix} T^{\lambda}_{\mu} = \frac{1}{2} g^{\mu\nu} g_{\nu\lambda,\sigma} T^{\lambda}_{\mu} = -\frac{1}{2} g^{\mu\nu}_{\sigma} T_{\mu\nu}.$$

Inserting both expressions into eq. (68) and regrouping terms, one finds

$$T^{\lambda}_{\sigma;\lambda} = \left[T^{\lambda}_{\sigma,\lambda} + \frac{1}{2} g^{\mu\nu}_{\sigma} T_{\mu\nu} \right] + (\log\sqrt{-g})_{,\mu} T^{\mu}_{\sigma}. \tag{69}$$

The left-hand side of eq. (69) is a generally-covariant vector. The last term on the right-hand side transforms as a vector under unimodular transformations. It follows that the expression in square brackets must also transform as a vector under unimodular transformations. In (Einstein 1915a) the vanishing of this expression is used as the energy-momentum balance law:

$$T^{\lambda}_{\sigma,\lambda} = -\frac{1}{2} g^{\mu\nu}_{\sigma} T_{\mu\nu}. \tag{70}$$

Note that eq. (70) is *not* equivalent to $T^{\lambda}_{\sigma;\lambda} = 0$, unless the last term of eq. (69) vanishes, as it does when the restriction to unimodular transformations is strengthened by setting $\sqrt{-g} = 1$.[85]

Einstein (1915a, 784–785) investigated whether any restrictions over and above the restriction to unimodular transformations would be needed to make sure that the field equations (67) are compatible with energy-momentum conservation as expressed in eq. (70). Using the field equations (64) on the right-hand side of eq. (70), one finds (cf. eq. (17)):

---

85 Einstein explicitly acknowledged that eq. (70) is not equivalent to $T^{\lambda}_{\sigma;\lambda} = 0$. The restriction to unimodular transformations, Einstein (1915a, 780) conceded, cannot be used to simplify the basic formulae for covariant differentiation given in his systematic exposition of the *Entwurf* theory (Einstein 1914c, 1050, eqs. (29) and (30)). But, he added, the "defining definition" ("Definitionsgleichung") of the covariant divergence can be simplified. He then went through the argument following eq. (69) to redefine the covariant divergence of an arbitrary symmetric tensor $A^{\mu\nu}$ (Einstein 1915a, 780, eq. 9) as: $\partial A^{\nu}_{\sigma}/\partial x^{\nu} - (1/2) g^{\tau\mu} g_{\mu\nu,\sigma} A^{\nu}_{\tau}$. He noted (ibid., 781) that this equation has the same form as the covariant divergence of the tensor density $\mathfrak{A}^{\mu\nu} \equiv \sqrt{-g} A^{\mu\nu}$ as defined in (Einstein 1914c, 1054, eq. 41b). This illustrates the general remark he made at the beginning of sec. 1 of (Einstein 1915a): "Because of the scalar character of $\sqrt{-g}$ a simplification of the basic formulae for the formation of invariant objects is possible … which in short comes down to this, that the factors $\sqrt{-g}$ and $1/\sqrt{-g}$ no longer occur in the basic formulae and that the difference between tensors and tensor densities disappears" ("Vermöge des Skalarscharakters von $\sqrt{-g}$ lassen die Grundformeln der Kovariantenbildung … eine Vereinfachung zu, die kurz gesagt darin beruht, daß in den Grundformeln die Faktoren $\sqrt{-g}$ und $1/\sqrt{-g}$ nicht mehr auftreten und der Unterschied zwischen Tensoren und V-Tensoren wegfällt.").

$$T^{\lambda}_{\sigma,\lambda} = \frac{1}{2\kappa} g^{\mu\nu}_{\sigma}\left(\frac{\partial}{\partial x^{\alpha}}\left(\frac{\partial \mathfrak{L}}{\partial g^{\mu\nu}_{\alpha}}\right) - \frac{\partial \mathfrak{L}}{\partial g^{\mu\nu}}\right). \tag{71}$$

This can be rewritten as (Einstein 1915a, eq. 20)

$$(T^{\lambda}_{\sigma} + t^{\lambda}_{\sigma})_{,\lambda} = 0,$$

by introducing a gravitational energy-momentum pseudo-tensor defined as (ibid., eq. 20a):[86]

$$t^{\lambda}_{\sigma} = \frac{1}{2\kappa}\left(\delta^{\lambda}_{\sigma}\mathfrak{L} - g^{\mu\nu}_{\sigma}\frac{\partial \mathfrak{L}}{\partial g^{\mu\nu}_{\lambda}}\right). \tag{72}$$

Inserting eq. (63) for $\mathfrak{L}$ and eq. (66) for its derivative with respect to $g^{\mu\nu}_{\lambda}$, one finds

$$\kappa t^{\lambda}_{\sigma} = \frac{1}{2}(\delta^{\lambda}_{\sigma}[g^{\mu\nu}\Gamma^{\alpha}_{\beta\mu}\Gamma^{\beta}_{\alpha\nu}] - g^{\mu\nu}_{\sigma}\Gamma^{\lambda}_{\mu\nu}),$$

which can be rewritten as (ibid, eq. 20b)[87]

$$\kappa t^{\lambda}_{\sigma} = \frac{1}{2}\delta^{\lambda}_{\sigma} g^{\mu\nu}\Gamma^{\alpha}_{\beta\mu}\Gamma^{\beta}_{\alpha\nu} - g^{\mu\nu}\Gamma^{\alpha}_{\mu\sigma}\Gamma^{\lambda}_{\alpha\nu}. \tag{73}$$

Einstein now rewrote (the mixed form of) the field equations in such a way that they have $T^{\lambda}_{\mu} + t^{\lambda}_{\mu}$ on the right-hand side. The divergence of the left-hand side then gives the quantity B_{μ} in the new theory. The vanishing of B_{μ} in conjunction with the field equations guarantees energy-momentum conservation, i.e., the vanishing of the ordinary divergence of $T^{\lambda}_{\mu} + t^{\lambda}_{\mu}$.[88]

Contraction of the field equations (67) with $g^{\nu\lambda}$ gives:

$$g^{\nu\lambda}(\Gamma^{\alpha}_{\mu\nu,\alpha} + \Gamma^{\alpha}_{\beta\mu}\Gamma^{\beta}_{\alpha\nu}) = -\kappa T^{\lambda}_{\mu}.$$

---

86 The derivation of eq. (72) is fully analogous to the derivation of eq. (18) in the general formalism (replace Q by $\mathfrak{L}$ in the calculation in note 31).

87 Since the covariant derivative of the metric vanishes,

$$0 = g^{\mu\nu}{}_{;\sigma} = g^{\mu\nu}_{\sigma} + \begin{Bmatrix}\mu\\ \alpha\sigma\end{Bmatrix} g^{\alpha\nu} + \begin{Bmatrix}\nu\\ \alpha\sigma\end{Bmatrix} g^{\mu\alpha} = g^{\mu\nu}_{\sigma} - \Gamma^{\mu}_{\alpha\sigma}g^{\alpha\nu} - \Gamma^{\nu}_{\alpha\sigma}g^{\mu\alpha},$$

it follows that $g^{\mu\nu}_{\sigma}\Gamma^{\lambda}_{\mu\nu} = (\Gamma^{\mu}_{\alpha\sigma}g^{\alpha\nu} + \Gamma^{\nu}_{\alpha\sigma}g^{\mu\alpha})\Gamma^{\lambda}_{\mu\nu} = 2g^{\alpha\nu}\Gamma^{\mu}_{\alpha\sigma}\Gamma^{\lambda}_{\mu\nu} = 2g^{\mu\nu}\Gamma^{\alpha}_{\mu\sigma}\Gamma^{\lambda}_{\alpha\nu}$.

88 Einstein could not just replace Q by $H = -\mathfrak{L}$ in the field equations in the form of eq. (13) and read off B_{μ} from the resulting equations. Recall that there are two definitions of the gravitational energy-momentum tensor, designated earlier as $t^{\lambda}_{\mu}(Q, \text{source})$ and $t^{\lambda}_{\mu}(Q, \text{cons})$ (cf. eqs. (19)–(20)). The condition $S^{\nu}_{\sigma}(Q) = 0$ guarantees that these two quantities are equal to one another. The corresponding quantities $t^{\lambda}_{\mu}(\mathfrak{L}, \text{source})$ and $t^{\lambda}_{\mu}(\mathfrak{L}, \text{cons})$, however, are not equal to one another, since $S^{\nu}_{\sigma}(\mathfrak{L}) \neq 0$ (cf. the discussion following eq. (34)).The analogue of eq. (13) in this case would be:

$$\frac{\partial}{\partial x^{\alpha}}\left(g^{\nu\lambda}\frac{\partial \mathfrak{L}}{\partial g^{\mu\nu}_{\alpha}}\right) = -\kappa(T^{\lambda}_{\mu} + t^{\lambda}_{\mu}(\mathfrak{L}, \text{source}))$$

The gravitational energy-momentum pseudo-tensor in eq. (73) corresponds to $t^{\lambda}_{\mu}(\mathfrak{L}, \text{cons})$ (as will be clear from comparing eq. (72) to eq. (18)). This quantity cannot be used interchangeably with $t^{\lambda}_{\mu}(\mathfrak{L}, \text{source})$.

This can be rewritten as:[89]

$$(g^{\nu\lambda}\Gamma^{\alpha}_{\mu\nu})_{,\alpha} - g^{\nu\rho}\Gamma^{\lambda}_{\alpha\rho}\Gamma^{\alpha}_{\mu\nu} = -\kappa T^{\lambda}_{\mu}.$$

The second term on the left-hand side is equal to the last term of eq. (73). Hence:

$$-g^{\nu\rho}\Gamma^{\lambda}_{\alpha\rho}\Gamma^{\alpha}_{\mu\nu} = \kappa t^{\lambda}_{\mu} - \frac{1}{2}\delta^{\lambda}_{\mu}g^{\rho\sigma}\Gamma^{\alpha}_{\beta\rho}\Gamma^{\beta}_{\alpha\sigma}.$$

The field equations can thus be written as:[90]

$$(g^{\nu\lambda}\Gamma^{\alpha}_{\mu\nu})_{,\alpha} - \frac{1}{2}\delta^{\lambda}_{\mu}g^{\rho\sigma}\Gamma^{\alpha}_{\beta\rho}\Gamma^{\beta}_{\alpha\sigma} = -\kappa(T^{\lambda}_{\mu} + t^{\lambda}_{\mu}). \tag{74}$$

The quantity B_{μ} in this case is thus given by (cf. eqs. (21)–(22)):

$$B_{\mu} = \frac{\partial}{\partial x^{\lambda}}\left(\frac{\partial}{\partial x^{\alpha}}(g^{\nu\lambda}\Gamma^{\alpha}_{\mu\nu}) - \frac{1}{2}\delta^{\lambda}_{\mu}g^{\rho\sigma}\Gamma^{\alpha}_{\beta\rho}\Gamma^{\beta}_{\alpha\sigma}\right). \tag{75}$$

The condition $B_{\mu} = 0$ guarantees that $(T^{\lambda}_{\mu} + t^{\lambda}_{\mu})_{,\lambda} = 0$.

Given his analysis of the covariance properties of the *Entwurf* field equations in 1914, Einstein had come to expect that this same condition $B_{\mu} = 0$ circumscribes the covariance of the field equations. In the case of field equations (67), he knew that their covariance group is that of arbitrary unimodular transformations, i.e., transformations under which the determinant g of the metric transforms as a scalar. This only gives one condition, not four as in eq. (75). In view of this mismatch, it is understandable that Einstein tried to replace these four conditions by one condition on g.[91]

His first step was to rewrite B_{μ} in the form $\partial A/\partial x_{\mu}$ and make B_{μ} vanish by imposing the stronger condition $A = 0$. Eq. (75) can be rewritten as

$$B_{\mu} = \frac{\partial^2}{\partial x^{\lambda}\partial x^{\alpha}}(g^{\nu\lambda}\Gamma^{\alpha}_{\mu\nu}) - \frac{1}{2}\frac{\partial}{\partial x^{\mu}}(g^{\rho\sigma}\Gamma^{\alpha}_{\beta\rho}\Gamma^{\beta}_{\alpha\sigma}).$$

The first term works out to be $\frac{1}{2}g^{\alpha\beta}_{\alpha\beta\mu}$,[92] so this equation becomes (Einstein 1915a, 785, eq. 22[93])

89 The left-hand side can be written as

$$(g^{\nu\lambda}\Gamma^{\alpha}_{\mu\nu})_{,\alpha} - g^{\nu\lambda}_{\alpha}\Gamma^{\alpha}_{\mu\nu} + g^{\nu\lambda}\Gamma^{\alpha}_{\beta\mu}\Gamma^{\beta}_{\alpha\nu}.$$

Substituting $\Gamma^{\nu}_{\alpha\rho}g^{\rho\lambda} + \Gamma^{\lambda}_{\alpha\rho}g^{\nu\rho}$ for $g^{\nu\lambda}_{\alpha}$ in the second term (see note 87), one finds:

$$(g^{\nu\lambda}\Gamma^{\alpha}_{\mu\nu})_{,\alpha} - g^{\rho\lambda}\Gamma^{\nu}_{\alpha\rho}\Gamma^{\alpha}_{\mu\nu} - g^{\nu\rho}\Gamma^{\lambda}_{\alpha\rho}\Gamma^{\alpha}_{\mu\nu} + g^{\nu\lambda}\Gamma^{\alpha}_{\beta\mu}\Gamma^{\beta}_{\alpha\nu}$$

The second term cancels against the fourth and the two remaining terms form the left-hand side of the equation below.

90 Einstein omitted the manipulations to get to eq. (74). He simply wrote: "after simple rearrangement" ("nach einfacher Umformung," Einstein 1915a, 785). The second term on the left-hand side of eq. (74) is, as we shall see later (see eq. (85)), equal to $-(1/2)\delta^{\lambda}_{\mu}t$, where t is the trace of t^{λ}_{μ}.

91 As we see no other plausible explanation for this move, this provides strong evidence for our claim that Einstein relied heavily on the formalism of (Einstein 1914c) to guide him in the analysis presented in the papers of November 1915.

$$B_\mu = \frac{1}{2}\frac{\partial}{\partial x^\mu}[g^{\alpha\beta}_{\alpha\beta} - g^{\rho\sigma}\Gamma^\alpha_{\beta\rho}\Gamma^\beta_{\alpha\sigma}]. \tag{76}$$

Einstein now replaced the four conditions $B_\mu = 0$ by a single stronger condition (ibid., eq. 22a):

$$g^{\alpha\beta}_{\alpha\beta} - g^{\rho\sigma}\Gamma^\alpha_{\beta\rho}\Gamma^\beta_{\alpha\sigma} = 0. \tag{77}$$

The next step was to replace this condition by a condition on g. To this end, Einstein fully contracted the field equations and compared the resulting condition to condition (77). Contraction of the field equations (67) with $g^{\mu\nu}$ gives:

$$g^{\mu\nu}(\Gamma^\alpha_{\mu\nu,\alpha} + \Gamma^\alpha_{\beta\mu}\Gamma^\beta_{\alpha\nu}) = -\kappa T. \tag{78}$$

This equation can be rewritten as

$$(g^{\mu\nu}\Gamma^\alpha_{\mu\nu})_{,\alpha} - g^{\mu\nu}_\alpha\Gamma^\alpha_{\mu\nu} + g^{\mu\nu}\Gamma^\alpha_{\beta\mu}\Gamma^\beta_{\alpha\nu} = -\kappa T. \tag{79}$$

The first term on the left-hand side is equal to:[94]

$$g^{\alpha\beta}_{\alpha\beta} + \frac{\partial}{\partial x^\alpha}\left(g^{\alpha\beta}\frac{\partial}{\partial x^\beta}\log\sqrt{-g}\right);$$

the second to minus twice the third:[95]

$$-2g^{\mu\nu}\Gamma^\alpha_{\beta\mu}\Gamma^\beta_{\alpha\nu}.$$

Inserting these last two expressions into eq. (79), one finds (Einstein 1915a, 785, eq. 21[96]):

92 Inserting eq. (53) for $\Gamma^\alpha_{\mu\nu}$, one finds

$$\frac{\partial^2}{\partial x^\lambda \partial x^\alpha}(g^{\nu\lambda}\Gamma^\alpha_{\mu\nu}) = -\frac{1}{2}\frac{\partial^2}{\partial x^\lambda\partial x^\alpha}(g^{\nu\lambda}g^{\alpha\rho}(g_{\rho\mu,\nu} + g_{\rho\nu,\mu} - g_{\mu\nu,\rho}))$$

Using that $g^{\mu\alpha}g_{\alpha\nu,\lambda} = -g^{\mu\alpha}_\lambda g_{\alpha\nu}$, one can rewrite this as:

$$-\frac{1}{2}\frac{\partial^2}{\partial x^\lambda\partial x^\alpha}(-g^{\nu\lambda}g^{\alpha\rho}_\nu g_{\rho\mu} - g^{\nu\lambda}g^{\alpha\rho}_\mu g_{\rho\nu} + g^{\nu\lambda}_\rho g^{\alpha\rho}g_{\mu\nu}).$$

The first and the third term in parentheses can be grouped together to form a quantity anti-symmetric in λ and α and thus vanish upon contraction with $\partial^2/\partial x^\lambda\partial x^\alpha$; the second term can be rewritten as $\delta^\lambda_\rho g^{\alpha\rho}_\mu = g^{\alpha\lambda}_\mu$. Finally, $(g^{\alpha\lambda}_\mu)_{,\lambda\alpha} = g^{\alpha\lambda}_{\alpha\lambda\mu}$.

93 Einstein did not use the designation B_μ for this quantity, thereby obscuring its origin in the variational formalism of (Einstein 1914c).

94 Inserting eq. (53) for $\Gamma^\alpha_{\mu\nu}$, one finds (cf. note 92):

$$(g^{\mu\nu}\Gamma^\alpha_{\mu\nu})_{,\alpha} = -\frac{1}{2}\frac{\partial}{\partial x^\alpha}(g^{\mu\nu}g^{\alpha\rho}(2g_{\rho\mu,\nu} - g_{\mu\nu,\rho})).$$

The first term on the right-hand side can be rewritten as $(g^{\mu\nu}g^{\alpha\rho}_\nu g_{\rho\mu})_{,\alpha} = g^{\alpha\nu}_{\nu\alpha}$; the second term as

$$\frac{1}{2}\frac{\partial}{\partial x^\alpha}(g^{\alpha\rho}[g^{\mu\nu}g_{\mu\nu,\rho}]) = \frac{1}{2}\frac{\partial}{\partial x^\alpha}\left(g^{\alpha\rho}\left[2\frac{\partial}{\partial x^\rho}\log\sqrt{-g}\right]\right)$$

(see, e.g., Einstein 1914c, 1051, eq. 32). The sum of these two terms gives the expression below.

95 Substituting $\Gamma^\mu_{\alpha\rho}g^{\rho\nu} + \Gamma^\nu_{\alpha\rho}g^{\mu\rho}$ for $g^{\mu\nu}_\alpha$ in $-g^{\mu\nu}_\alpha\Gamma^\alpha_{\mu\nu}$ (cf. note 87), one finds $-2g^{\rho\nu}\Gamma^\mu_{\alpha\rho}\Gamma^\alpha_{\mu\nu}$. Relabeling of the summation indices gives the expression below.

$$g^{\alpha\beta}_{\alpha\beta} - g^{\mu\nu}\Gamma^{\alpha}_{\beta\mu}\Gamma^{\beta}_{\alpha\nu} + \frac{\partial}{\partial x^{\alpha}}\left(g^{\alpha\beta}\frac{\partial}{\partial x^{\beta}}\log\sqrt{-g}\right) = -\kappa T. \tag{80}$$

The first two terms on the left-hand side of this equation are equal to the two terms on the left-hand side of eq. (77). Given that eq. (80) follows from the field equations (67), it suffices to demand that

$$\frac{\partial}{\partial x^{\alpha}}\left(g^{\alpha\beta}\frac{\partial}{\partial x^{\beta}}\log\sqrt{-g}\right) = -\kappa T \tag{81}$$

(ibid., 785, eq. 21a) to make sure that condition (77) is also satisfied. Through eq. (76) this guarantees that $B_{\mu} = 0$. This in turn guarantees that $(T^{\lambda}_{\mu} + t^{\lambda}_{\mu})_{,\lambda} = 0$ (see eqs. (74)–(75)). One condition on the determinant of the metric thus suffices to guarantee compatibility of the field equations (67) with energy-momentum conservation. The condition is an odd one though. The energy-momentum tensor phenomenologically representing ordinary matter has a non-vanishing trace. Eq. (81) thus says that g cannot be a constant. This means that there still is a residual discrepancy between the covariance of the field equations and the coordinate restriction needed to guarantee compatibility with energy-momentum conservation. The restriction to unimodular transformations only demands g to transform as a scalar, it does not say that it cannot be a constant. Within a few weeks, Einstein published two modifications of his field equations to change this condition to the more congenial condition $\sqrt{-g} = 1$ for unimodular coordinates (see sec. 7).

It was nonetheless an extremely important result that the four conditions $B_{\mu} = 0$ can be reduced to one condition in this case. Up to this point, Einstein had not made a distinction between coordinate restrictions needed to recover the Poisson equation for weak static fields and those needed to ensure compatibility with energy-momentum conservation. It now turned out that the latter demand could be satisfied by one condition whereas the former continued to call for four. We conjecture that this drove home the point that the two requirements should be dealt with separately.

In fact, immediately after eq. (81) in (Einstein 1915a), we find the very first unambiguous instance in both Einstein's published papers and extant manuscripts and correspondence of a coordinate condition used in the modern sense. Einstein (1915a, 786, eqs. 22 and 16b) showed how the conditions $g^{\alpha\beta}_{\beta} = 0$, which we have called the Hertz condition/restriction (see eq. (7)), reduce the November tensor to the d'Alembertian acting on the metric in the case of weak fields. Einstein had done the same calculation in the Zurich Notebook (see eqs. (7)–(8)). There he had used $g^{\alpha\beta}_{\beta} = 0$ as a coordinate restriction. As such it was unacceptable because it was not satisfied, for instance, by the Minkowski metric in rotating coordinates.[97] The return of $g^{\alpha\beta}_{\beta} = 0$ in (Einstein 1915a) makes it clear that Einstein did not see this as a problem anymore in 1915. Einstein had come to realize that the conditions $g^{\alpha\beta}_{\beta} = 0$ are

96 Einstein omitted the manipulations to get from eq. (78) to eq. (80), again writing simply "after simple rearrangement" (cf. note 90)

not an integral part of the theory and only serve to facilitate comparison of the field equations to the Poisson equation of Newtonian theory in the case of weak static fields. In other words, Einstein now saw that $g^{\alpha\beta}_{\beta} = 0$ is not a coordinate restriction but a coordinate condition in the modern sense.

The theory of (Einstein 1915a) was thus of broad covariance. The only restrictions were that the determinant of the metric transform as a scalar and that it not be a constant. The way Einstein saw it, this was all that was needed to solve the problem of rotation that had brought down the *Entwurf* theory.[98] In the concluding paragraphs of (Einstein 1915a), he pointed out that transformations to rotating coordinates belong to the class of unimodular coordinates under which the new field equations are invariant.[99]

Looking back on secs. 5 and 6, we can clearly see how redefining the components of the gravitational field untied the tight knot of conditions and definitions that had been the *Entwurf* theory and retied it in a slightly different way to become a theory within hailing distance of general relativity as we know it today. First and foremost, the redefinition of the gravitational field led to the replacement of the *Entwurf* field equations and their intractable covariance properties by field equations invariant under arbitrary unimodular transformations. But that was not all. In the new theory, instead of the four additional restrictions $B_{\mu} = 0$ familiar from the *Entwurf* theory, it took only a minimal strengthening of the restriction to unimodular transformations (namely that the determinant of the metric not be a constant) to ensure that these new field equations yield energy-momentum conservation. Finally, because energy-momentum conservation only called for one extra condition whereas recovery of the Poisson equation continued to call for four, it became clear that these two types of conditions have a different status. Taking advantage of this insight, Einstein used a coordinate condition in the modern sense to show that the relevant component of the new field equations reduces to the Poisson equation for weak static fields and only used a coordinate restriction to satisfy the demands of energy-momentum conservation. There was thus enough covariance left in the new theory to meet the demands of Einstein's relativity and equivalence principles.

---

97 John Norton (1984, 119 and 143) already suggested that Einstein rejected the combination of the November tensor and the conditions $g^{\alpha\beta}_{\beta} = 0$ in the Zurich Notebook because $g^{\alpha\beta}_{\beta} \neq 0$ for the Minkowski metric in rotating coordinates. We agree. Note, however, that Einstein's argument is cogent only if the conditions $g^{\alpha\beta}_{\beta} = 0$ are seen as a coordinate restriction rather than a coordinate condition. In fact, it was in an attempt to make sense of these remarks in (Norton 1984) that one of us (JR) first hit upon the distinction between coordinate conditions and what we have come to call coordinate restrictions.

98 In fact, the problem of rotation persisted even in the final version of general relativity. General covariance does not make rotation—or *any* non-geodesic motion for that matter—relative (see Janssen 2005, 68–72).

99 On the face of it, it may seem that the theory still does not allow the Minkowski metric in rotating coordinates because its determinant equals one. Since the Minkowski metric is a vacuum solution of the field equations, however, it does not matter in this case that through eq. (81) $\sqrt{-g} = 1$ implies that $T = 0$ as well.

7. FROM THE NOVEMBER TENSOR TO THE EINSTEIN FIELD EQUATIONS

Einstein soon found ways of replacing the condition of (Einstein 1915a) that $\sqrt{-g}$ cannot be a constant by the more congenial condition $\sqrt{-g} = 1$ for unimodular coordinates. In (Einstein 1915b), he achieved this through the assumption that the trace T of the energy-momentum tensor of matter vanishes. He justified this assumption by adopting the view, promoted by Gustav Mie and others, that all matter is electromagnetic. Energy-momentum conservation now followed from the field equations in unimodular coordinates without any additional coordinate restrictions. And the field equations themselves could be looked upon as generally-covariant field equations in unimodular coordinates. So Einstein had found generally-covariant field equations at last: $R_{\mu\nu} = -\kappa T_{\mu\nu}$ (with $R_{\mu\nu}$ the Ricci tensor). In his calculations, both in November 1915 and in (Einstein 1916a), however, he continued to use unimodular coordinates. And he soon had second thoughts about paying for general covariance by committing himself to the electromagnetic view of nature. In (Einstein 1915d), he changed the condition that $\sqrt{-g}$ cannot be a constant to the condition $\sqrt{-g} = 1$ without imposing any restrictions on T. This he achieved by adding a term with T to the right-hand side of the field equations based on the November tensor. He realized that this trace term was needed anyway to ensure that the energy-momentum tensor for matter enters the field equations in the exact same way as the gravitational energy-momentum pseudo-tensor. This told Einstein that these were the equations he had been looking for. As before, they could be looked upon as generally-covariant equations expressed in unimodular coordinates. Einstein had thus found the Einstein field equations: $R_{\mu\nu} = -\kappa(T_{\mu\nu} - (1/2)g_{\mu\nu}T)$.

In the second and fourth of his communications to the Berlin Academy in November 1915, Einstein (1915b, 1915d) proposed two different ways to avoid the requirement that $\sqrt{-g}$ cannot be a constant found in the first communication (Einstein 1915a). In the second November paper, a three-page "Addendum" ("Nachtrag") to the first, he assumed that all matter is electromagnetic, in which case $T = 0$ (Einstein 1915b). Condition (81) can then be satisfied by setting $\sqrt{-g} = 1$, the condition for unimodular coordinates. This has three advantages. First, by looking upon the field equations as holding only in unimodular coordinates (rather than in coordinates related to one another by unimodular transformations) he removed the residual discrepancy between the covariance of the field equations and the restriction needed to guarantee energy-momentum conservation. Second, with $\sqrt{-g} = 1$, the energy-momentum balance equation, $T^{\mu}_{\nu;\mu} = 0$, reduces to $T^{\mu}_{\nu,\mu} + (1/2)g^{\alpha\beta}_{\nu}T_{\alpha\beta} = 0$ (see eq. (70)), the equations used in the analysis of energy-momentum conservation in (Einstein 1915a). The third and most important advantage is that the field equations (67) could now be looked upon as generally-covariant field equations expressed in unimodular coordinates. Setting $\sqrt{-g} = 1$ also has one disadvantage. It rules out a metric of the form $g_{\mu\nu} = \mathrm{diag}(-1, -1, -1, f(x, y, z))$, which Einstein still assumed was the general form of the metric for weak static fields. Either Einstein did not think of this disadvantage at this point (though we shall see that he had thought of it a week later), or it was outweighed, at least for the time being, by the advantages.

Einstein (1915b, 800, eq. 16b) wrote these generally-covariant field equations as

$$G_{\mu\nu} = -\kappa T_{\mu\nu}, \tag{82}$$

where $G_{\mu\nu}$ is the Ricci tensor. As in (Einstein 1915a), he wrote the Ricci tensor as the sum of two terms,[100]

$$G_{im} = R_{im} + S_{im}.$$

The first term is defined as minus what we called the November tensor (i.e., T_{il}^{x} in eq. (5)):

$$R_{im} \equiv -\frac{\partial}{\partial x^l}\begin{Bmatrix} l \\ im \end{Bmatrix} + \begin{Bmatrix} \rho \\ il \end{Bmatrix}\begin{Bmatrix} l \\ \rho m \end{Bmatrix}.$$

The second term is defined as:

$$S_{im} \equiv \frac{\partial}{\partial x^m}\begin{Bmatrix} l \\ il \end{Bmatrix} - \begin{Bmatrix} \rho \\ im \end{Bmatrix}\begin{Bmatrix} l \\ \rho l \end{Bmatrix}.$$

Since the first and the third Christoffel symbol in this expression are equal to the gradient of $\lg\sqrt{-g}$ it follows that $S_{im} = 0$ in unimodular coordinates. In the Zurich Notebook and in his first November paper, Einstein used the decomposition of the Ricci tensor only to show that the November tensor transforms as a tensor under unimodular *transformations*. In unimodular *coordinates*, the Ricci tensor actually reduces to the November tensor and the field equations (82) reduce to the field equations (67) of (Einstein 1915a):

$$R_{\mu\nu} = -\kappa T_{\mu\nu} \tag{83}$$

(Einstein 1915b, 801, eq. 16).

Einstein had thus finally found generally-covariant field equations. His calculations in (Einstein 1915a) show that *in unimodular coordinates* these field equations guarantee energy-momentum conservation (see eqs. (70)–(85)). Although Einstein did not explicitly show this, it was reasonable to assume that the corresponding generally-covariant equations $G_{\mu\nu} = -\kappa T_{\mu\nu}$ guarantee energy-momentum conservation in arbitrary coordinates.

Most of the "Addendum" (Einstein 1915b) is taken up by a defense of the assumption $T = 0$. The results reported in the "Addendum" all depend on this assumption. The assumption holds for electromagnetic fields in Maxwell's theory and might continue to hold in the non-linear generalizations of Maxwell's theory pursued by Gustav Mie and other proponents of the electromagnetic view of nature.[101] It does not hold for the energy-momentum tensor routinely used to give a phenomeno-

100 See (Einstein 1915a, 782; 1915b, 800; 1915d, 844).

logical description of ordinary matter. To get around this problem, Einstein assumed that gravitational fields play an important role in the constitution of matter. In that case the energy-momentum tensor phenomenologically describing matter should not be identified with $T^{\mu\nu}$ but with $T^{\mu\nu} + t^{\mu\nu}$ and the non-vanishing trace might come from $t^{\mu\nu}$ rather than $T^{\mu\nu}$. Einstein's flirtation with the electromagnetic world picture was short-lived, but three and a half years later he resurrected the idea that gravity plays a role in the structure of matter in a theory that makes the cosmological constant of (Einstein 1917) responsible both for the stability of the cosmos and the stability of elementary particles (Einstein 1919).[102]

At first Einstein was very enthusiastic about the electromagnetic turn his theory had taken. In an abstract for his third paper of November 1915 (Einstein 1915c)—the one in which he used the field equations of (Einstein 1915b) in unimodular coordinates to explain the anomalous advance of the perihelion of Mercury—Einstein wrote that this result "confirms the hypothesis of the vanishing of the scalar of the energy tensor of "matter" [i.e., $T = 0$]."[103] His enthusiasm, however, waned quickly. In a footnote to the perihelion paper itself, he announced:

> In a communication that will follow shortly it will be shown that this hypothesis [i.e., $T = 0$] is dispensable. Essential is only that a choice of reference frame is possible such that the determinant $|g_{\mu\nu}|$ takes on the value -1.[104]

As we shall see, the calculation of the perihelion advance of Mercury played an important role in showing Einstein that one can set $\sqrt{-g} = 1$ without setting $T = 0$.

In (Einstein 1915d), the fourth and final paper of November 1915 and the communication announced in the footnote quoted above, Einstein changed the field equations in such a way that condition (81) changes to

$$\frac{\partial}{\partial x^{\alpha}}\left(g^{\alpha\beta}\frac{\partial}{\partial x^{\beta}}\log\sqrt{-g}\right) = 0. \tag{84}$$

This makes it possible to choose unimodular coordinates (i.e., set $\sqrt{-g} = 1$) without putting any condition on the trace T of the energy-momentum tensor.

How did Einstein arrive at this new condition (84)? Recall that the original condi-

---

101 Einstein knew about Hilbert's work along these lines. See (Sauer 1999) and "Hilbert's Foundation of Physics …" (in vol. 4 of this series).

102 This theory is enjoying renewed interest. It now goes by the name of "unimodular gravity" (see Anderson and Finkelstein 1971); for a more recent discussion and further references, see (Finkelstein et al. 2002, Earman 2003).

103 "Dadurch wird die Hypothese vom Verschwinden des Skalars des Energietensors der "Materie" bestätigt." *Königlich Preußische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte* (1915): 803. For further discussion, see "Hilbert's Foundation of Physics …" (in vol. 4 of this series) and "Pathways out of Classical Physics …" (in vol. 1 of this series).

104 "In einer bald folgenden Mitteilung wird gezeigt werden, daß jene Hypothese entbehrlich ist. Wesentlich ist nur, daß eine solche Wahl des Bezugssystems möglich ist, daß die Determinante $|g_{\mu\nu}|$ den Wert -1 annimmt" (Einstein 1915c, 831, note 1).

tion (81) followed from combining two equations. The first is eq. (77) which comes from the condition $B_\mu = 0$ for the field equations (67) of (Einstein 1915a). The second is eq. (80) which comes from fully contracting those field equations. These two equations can be written more compactly by introducing the trace of the gravitational energy-momentum pseudo-tensor (73) as Einstein first did in (Einstein 1915d, 846):[105]

$$\kappa t = g^{\mu\nu}\Gamma^{\alpha}_{\beta\mu}\Gamma^{\beta}_{\alpha\nu}. \tag{85}$$

Inserting κt in the second terms of both eq. (77) and eq. (80), one finds:

$$g^{\alpha\beta}_{\alpha\beta} - \kappa t = 0, \tag{86}$$

$$g^{\alpha\beta}_{\alpha\beta} - \kappa t + \frac{\partial}{\partial x^{\alpha}}\left(g^{\alpha\beta}\frac{\partial}{\partial x^{\beta}}\log\sqrt{-g}\right) = -\kappa T. \tag{87}$$

The combination of these two equations gives the problematic condition (81). Upon inspection of eqs. (86)–(87), one sees that condition (81) would change to condition (84) if, instead of eqs. (86)–(87), one had (cf. Einstein 1915d, eqs. 10 and 9, respectively):

$$g^{\alpha\beta}_{\alpha\beta} - \kappa(t+T) = 0, \tag{88}$$

$$g^{\alpha\beta}_{\alpha\beta} - \kappa(t+T) + \frac{\partial}{\partial x^{\alpha}}\left(g^{\alpha\beta}\frac{\partial}{\partial x^{\beta}}\log\sqrt{-g}\right) = 0. \tag{89}$$

The difference between eq. (87) and eq. (89) is the sign of the term κT. Recall that eq. (87) was obtained by fully contracting the field equations (67). One can change the sign of the term κT in eq. (87) by adding a trace term to the right-hand side of eq. (67):[106]

$$\Gamma^{\alpha}_{\mu\nu,\alpha} + \Gamma^{\alpha}_{\beta\mu}\Gamma^{\beta}_{\alpha\nu} = -\kappa\left(T_{\mu\nu} - \frac{1}{2}g_{\mu\nu}T\right). \tag{90}$$

Fully contracting these new field equations and rewriting the resulting equations, one arrives at (cf. eqs. (78)–(80))

$$g^{\alpha\beta}_{\alpha\beta} - g^{\mu\nu}\Gamma^{\alpha}_{\beta\mu}\Gamma^{\beta}_{\alpha\nu} + \frac{\partial}{\partial x^{\alpha}}\left(g^{\alpha\beta}\frac{\partial}{\partial x^{\beta}}\log\sqrt{-g}\right) = \kappa T.$$

Using eq. (85) to substitute $-\kappa t$ for the second term on the left-hand side, one arrives at eq. (89).

The new field equations (90) contracted with $g^{\nu\lambda}$ are

$$g^{\nu\lambda}(\Gamma^{\alpha}_{\mu\nu,\alpha} + \Gamma^{\alpha}_{\beta\mu}\Gamma^{\beta}_{\alpha\nu}) = -\kappa\left(T^{\lambda}_{\mu} - \frac{1}{2}\delta^{\lambda}_{\mu}T\right),$$

105 This follows from $\kappa t \equiv \kappa t^{\lambda}_{\lambda} = 2g^{\mu\nu}\Gamma^{\alpha}_{\beta\mu}\Gamma^{\beta}_{\alpha\nu} - g^{\mu\nu}\Gamma^{\alpha}_{\mu\lambda}\Gamma^{\lambda}_{\alpha\nu}$.

106 Contracting $-\kappa(T_{\mu\nu} - (1/2)g_{\mu\nu}T)$ with $g^{\mu\nu}$, one finds κT.

which can be rewritten as (cf. eq. (74)):

$$(g^{\nu\lambda}\Gamma^{\alpha}_{\mu\nu})_{,\alpha} - \frac{1}{2}\delta^{\lambda}_{\mu}g^{\rho\sigma}\Gamma^{\alpha}_{\beta\rho}\Gamma^{\beta}_{\alpha\sigma} = -\kappa(T^{\lambda}_{\mu} + t^{\lambda}_{\mu}) + \frac{1}{2}\delta^{\lambda}_{\mu}T.$$

Using eq. (85) to substitute κt for $g^{\rho\sigma}\Gamma^{\alpha}_{\beta\rho}\Gamma^{\beta}_{\alpha\sigma}$, one arrives at:

$$(g^{\nu\lambda}\Gamma^{\alpha}_{\mu\nu})_{,\alpha} - \frac{1}{2}\delta^{\lambda}_{\mu}\kappa(t+T) = -\kappa(T^{\lambda}_{\mu} + t^{\lambda}_{\mu}). \tag{91}$$

Energy-momentum is guaranteed—i.e., $(T^{\lambda}_{\mu} + t^{\lambda}_{\mu})_{,\lambda} = 0$ —if the divergence of the left-hand side vanishes:

$$\frac{\partial}{\partial x_{\lambda}}\left[(g^{\nu\lambda}\Gamma^{\alpha}_{\mu\nu})_{,\alpha} - \frac{1}{2}\delta^{\lambda}_{\mu}\kappa(t+T)\right] = 0.$$

This can be rewritten (cf. eqs. (75)–(76)) as

$$\frac{\partial}{\partial x^{\mu}}[g^{\alpha\beta}_{\alpha\beta} - \kappa(t+T)] = 0.$$

In other words, energy-momentum conservation is guaranteed if the expression in square brackets vanishes, which is just eq. (88). As we noted above, combining eqs. (88) and (89) gives condition (84) which makes it possible to set $\sqrt{-g} = 1$ without any consequences for the value of T. Eq. (89) is a direct consequence of the field equations. Eq. (88), which guarantees energy-momentum conservation, is therefore a consequence of the field equations plus the condition $\sqrt{-g} = 1$. There is no need anymore for the highly speculative assumption that all matter is electromagnetic.

As Norton (1984, 146–147) has emphasized, the addition of the trace term $(1/2)\kappa g_{\mu\nu}T$ to the field equations was not an option for Einstein before the perihelion paper (Einstein 1915c). In the Zurich Notebook, Einstein had briefly considered adding a trace term to field equations based on the Ricci tensor in a weak-field approximation. He had rejected such modified weak-field equations because they do not allow the spatially flat metric, $g_{\mu\nu} = \mathrm{diag}(-1, -1, -1, f(x, y, z))$, which Einstein expected to describe weak static fields.[107] The derivation of the perihelion motion of Mercury freed Einstein from this prejudice. It showed that weak static

107 These considerations can be found on pp. 20L–21R of the Zurich Notebook. See “Commentary …” (in this volume), secs. 5.4.3–5.4.4 for a detailed analysis and sec. 5.4.6 for a concise summary. The problem is this. For weak static fields, the field equations with trace term reduce to:

$$\Delta g_{\mu\nu} = \kappa\left(T_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}T\right)$$

(with $\eta_{\mu\nu} \equiv \mathrm{diag}(-1, -1, -1, 1)$). For a static mass distribution described by $T_{\mu\nu} = \mathrm{diag}(0, 0, 0, \rho)$ ($T = \rho$), the non-trivial components of these equations are:

$$\Delta g_{11} = \Delta g_{22} = \Delta g_{33} = \Delta g_{44} = \frac{1}{2}\kappa\rho.$$

The spatially flat metric $\mathrm{diag}(-1, -1, -1, f)$ is not a solution of these equations.

fields do not have to be spatially flat. This insight was directly related to his use of unimodular coordinates in this calculation (Earman and Janssen 1993, 144–145). If $\sqrt{-g} = 1$ and g_{44} is non-constant, then at least some of the spatial components g_{ij} ($i, j = 1, 2, 3$) must be non-constant as well ($g_{4i} = g_{i4} = 0$ for a static field). This removes the objection against adding a trace term. Einstein could now set $\sqrt{-g} = 1$ without committing himself to the electromagnetic worldview.

The field equations (90) with the trace term have another feature that strongly recommends them. Compare the field equations (90) in the form of eq. (91) to the field equations (67) of (Einstein 1915a) in the form of eq. (74). Using eq. (85) to substitute κt in the second term on the left-hand side, one can write eq. (74) as:

$$(g^{\nu\lambda}\Gamma^{\alpha}_{\mu\nu})_{,\alpha} - \frac{1}{2}\delta^{\lambda}_{\mu}\kappa t = -\kappa(T^{\lambda}_{\mu} + t^{\lambda}_{\mu}). \tag{92}$$

The crucial difference between eq. (91) and eq. (92) is that in eq. (91) the energy-momentum tensor of matter enters the field equations in the exact same way as the energy-momentum pseudo-tensor for the gravitational field, whereas in eq. (92) it does not. In eq. (92) there is a term $-(1/2)\delta^{\lambda}_{\mu}\kappa T$ missing on the left-hand side. Einstein made this same observation comparing eq. (89) (after setting $\sqrt{-g} = 1$) to its counterpart eq. (87) for the field equations (67) without the trace term:

> Note that our additional [trace] term brings with it that in [eq. 9 of (Einstein 1915d), $g^{\alpha\beta}_{\alpha\beta} - \kappa(t + T) = 0$] the energy tensor of the gravitational field occurs alongside the one for matter in the same way, which is not the case in [the corresponding eq. 21 of (Einstein 1915a), $g^{\alpha\beta}_{\alpha\beta} - \kappa t + \ldots = -\kappa T$].[108]

In the general formalism of (Einstein 1914c), the conditions $B_{\mu} = 0$ guarantee the vanishing of $(T^{\lambda}_{\mu} + t^{\lambda}_{\mu})_{,\lambda}$ and the conditions $S^{\nu}_{\sigma} = 0$ guarantee that $T_{\mu\nu}$ enters the field equations in the same way as $t_{\mu\nu}$ (see secs. 3.1 and 3.2). The conditions $S^{\nu}_{\sigma} = 0$, however, do not hold if the restriction to unimodular transformations or unimodular coordinates is made (see note 88). In his first November paper, Einstein made sure that $(T^{\lambda}_{\mu} + t^{\lambda}_{\mu})_{,\lambda} = 0$ holds, but he did not check whether $T_{\mu\nu}$ and $t_{\mu\nu}$ enter the field equations in the same way. If he had, he would have recognized the need for the trace term right away.[109]

---

108 “Man beachte, daß es unser Zusatzglied mit sich bringt daß in (9) der Energietensor des Gravitationsfeldes neben dem der Materie in gleicher Weise auftritt, was in Gleichung (21) a. a. O. nicht der Fall ist” (Einstein 1915d, 846).

109 Einstein concisely summarized this part of his struggle to come up with satisfactory field equations in a letter to Besso a little over a month later: “The first paper [Einstein 1915a] along with the addendum [Einstein 1915b] still suffers from the problem that the term $(1/2)\kappa g_{\mu\nu}T$ is missing on the right-hand side; hence the postulate $T = 0$. Obviously, things have to be done as in the last paper [Einstein 1915d], in which case there is no condition anymore on the structure of matter” (“Die erste Abhandlung samt dem Nachtrag krankt noch daran dass auf der rechten Seite das Glied $(1/2)\kappa g_{\mu\nu}T$ fehlt; daher das Postulat $T = 0$. Natürlich muss die Sache gemäss der letzten Arbeit gemacht werden, wobei sich über die Struktur der Materie keine Bedingung mehr ergibt.”) Einstein to Michele Besso, 3 January 1916 (CPAE 8, Doc. 178).

Einstein needed to check one more thing in his fourth November paper to make sure that the new field equations (90) with trace term do indeed give energy-momentum conservation in unimodular coordinates. He had to show that $(T_{\mu}^{\lambda} + t_{\mu}^{\lambda})_{,\lambda} = 0$ is equivalent to the energy-momentum balance equation in unimodular coordinates,

$$T_{\sigma,\lambda}^{\lambda} + \frac{1}{2}g_{\sigma}^{\mu\nu}T_{\mu\nu} = 0. \tag{93}$$

(cf. eqs. (68)–(70)). The question is whether the second term on the left-hand side of this equation can be rewritten as $t_{\sigma,\lambda}^{\lambda}$? The standard procedure for doing this is to replace $T_{\mu\nu}$ in this term by the left-hand side of the field equations (see note 41, eqs. (16)–(18) and eqs. (71)–(72)). In this case, however, the field equations have $T_{\mu\nu} - (1/2)g_{\mu\nu}T$ on the right-hand side rather than simply $T_{\mu\nu}$. It turns out that this does not lead to any complications. In unimodular coordinates, as Einstein (1915d, 846) noted,

$$g_{\sigma}^{\mu\nu}g_{\mu\nu} = -(\log\sqrt{-g})_{,\sigma} = 0.$$

It follows that the second term of eq. (93) can be written as:

$$\frac{1}{2}g_{\sigma}^{\mu\nu}\left(T_{\mu\nu} - \frac{1}{2}g_{\mu\nu}T\right).$$

With the help of the field equations (90) this turns into

$$-\frac{1}{2\kappa}g_{\sigma}^{\mu\nu}(\Gamma_{\mu\nu,\alpha}^{\alpha} + \Gamma_{\beta\mu}^{\alpha}\Gamma_{\alpha\nu}^{\beta}).$$

As Einstein had already shown in his first November paper (see eqs. (71)–(73)), this expression is equal to the divergence of the gravitational energy-momentum pseudotensor in unimodular coordinates (Einstein 1915d, 846, eq. 8a; 1915a, 785, eq. 20b)

$$\kappa t_{\sigma}^{\lambda} \equiv \frac{1}{2}\delta_{\sigma}^{\lambda}g^{\mu\nu}\Gamma_{\beta\mu}^{\alpha}\Gamma_{\alpha\nu}^{\beta} - g^{\mu\nu}\Gamma_{\mu\sigma}^{\alpha}\Gamma_{\alpha\nu}^{\lambda},$$

defined in eq. (73) and used throughout this section and sec. 6. In unimodular coordinates, the field equations (90) of (Einstein 1915d) thus satisfy all requirements needed for energy-momentum conservation:

(1) $T_{\mu\nu}$ and $t_{\mu\nu}$ enter the field equations in the exact same way;

(2) The field equations guarantee the vanishing of the divergence of $T_{\mu\nu} + t_{\mu\nu}$;

(3) The divergence of $t_{\mu\nu}$ is equal to the gravitational force density.

As with the field equations (83) of the "Addendum" (Einstein 1915b), Einstein looked upon the field equations (90),

$$R_{im} = -\kappa\left(T_{im} - \frac{1}{2}g_{im}T\right),$$

where R_{im} is the November tensor (Einstein 1915d, 845, eq. 6), as generally-covariant field equations expressed in unimodular coordinates. The corresponding generally-covariant equations are the Einstein field equations,

$$G_{im} = -\kappa\left(T_{im} - \frac{1}{2}g_{im}T\right), \tag{94}$$

where G_{im} is the full Ricci tensor (ibid., eq. 2a). As in (Einstein 1915b), he tacitly assumed that these equations would guarantee energy-momentum conservation in arbitrary coordinates.

To conclude our analysis of Einstein's four papers of November 1915, we summarize what we see as the four key steps in the transition from the *Entwurf* field equations to the Einstein field equations.[110] The first step was the redefinition of the components of the gravitational field which led Einstein back to field equations invariant under unimodular transformations that he had considered but rejected three years earlier in the Zurich Notebook. The second step was to rewrite the four conditions that in conjunction with the field equations guarantee energy-momentum conservation as one condition on g, the determinant of the metric, to reflect the connection between energy-momentum conservation and the covariance of the field equations. This made it possible for Einstein to start using coordinate conditions in the modern sense. The third step was to recognize that the theory could be tweaked to turn the one condition on g into the condition $\sqrt{-g} = 1$ for unimodular coordinates. This made it possible to look upon the new field equations as generally-covariant equations expressed in unimodular coordinates. This is what Einstein had to show for his brief dalliance with the electromagnetic worldview. The fourth and final step was to recognize that energy-momentum conservation dictates that such tweaking be done in a specific way, namely through adding a term with the trace of the energy-momentum tensor for matter to the field equations. The perihelion paper (Einstein 1915c) was important in this context in that it freed Einstein from his prejudice about the form of the metric for weak static fields which he had found to be incompatible with such a trace term in the Zurich Notebook. We reiterate that this whole chain of reasoning was set in motion by replacing definition (52) of the components of the gravitational field, the "fateful prejudice," by definition (53), "the key to the solution." In all of this Einstein relied heavily on the general variational formalism of (Einstein 1914c). The exact expressions, relations, and conditions given by this formalism could not be used because of the restriction to unimodular transformations and unimodular coordinates in the papers of November 1915, but the insights encoded in the formalism were Einstein's main guide in taking steps one, two, and four.

110 The first two steps were made in (Einstein 1915a) and are discussed in secs. 5 and 6, respectively; the last two were made in (Einstein 1915b) and (Einstein 1915d), respectively, and are both discussed in this section, sec. 7.

8. THE 1916 REVIEW ARTICLE: THE ARGUMENT OF NOVEMBER 1915 STREAMLINED

Einstein's argument in (Einstein 1915d) for adding a trace term on the right-side of the field equations proved difficult to follow even for those most supportive of his efforts, such as his Leyden colleagues Paul Ehrenfest and H. A. Lorentz. Although Einstein claimed in the introduction of (Einstein 1915d) that the paper was self-contained, it in fact relied heavily on (Einstein 1915a) in its justification of the trace term. The relevant part of (Einstein 1915a) in turn relied heavily on the exposition of the Entwurf theory in (Einstein 1914c). In early 1916, in a letter to Ehrenfest, Einstein produced a self-contained version of the argument leading to the trace term and the Einstein field equations of the fourth November communication without the detour through the various discarded field equations preceding them. This letter became the blueprint for the part on field equations and energy-momentum conservation in (Einstein 1916a), the first systematic self-contained exposition of the new theory.

It was clear to Einstein that the field equations of his last communication of November 1915 met all requirements that he had imposed on such equations and that no further changes would be needed. Given the rapid succession of different field equations during that one month, however, it is understandable that this was not so clear to his readers. Even those most supportive of Einstein's efforts, such as the Leyden physicists Paul Ehrenfest and H. A. Lorentz, had difficulties following the argument.

Einstein himself best described the impression that the flurry of papers of November 1915 must have made on his colleagues. Knowing that the final result was correct and fully aware of the monumental character of his achievement, Einstein could afford to poke fun at the chaotic way in which victory had at long last been achieved. "It's convenient with that fellow Einstein," he wrote to Ehrenfest, "every year he retracts what he wrote the year before."[111] With similar self-deprecation, he told Sommerfeld: "Unfortunately I have immortalized my final errors in this battle in the academy-papers [Einstein 1915a, b] that I can soon send you."[112] When he did send the papers a week and a half later, he urged Sommerfeld to study them carefully despite the fact "that, as you are reading, the final part of the battle for the field equations unfolds right in front of your eyes."[113]

111 "Es ist bequem mit dem Einstein. Jedes Jahr widerruft er, was er das vorige Jahr geschrieben hat." Einstein to Paul Ehrenfest, 26 December 1915 (CPAE 8, Doc. 173). With this comment, Einstein prefaced his retraction of the hole argument (see sec. 4 of "What Did Einstein Know …" [in this volume]).

112 "Die letzten Irrtümer in diesem Kampfe habe ich leider in den Akademie-Arbeiten, die ich Ihnen bald senden kann, verevigt [sic]." Einstein to Arnold Sommerfeld, 28 November 1915 (CPAE 8, Doc. 153).

113 "… dass sich beim Lesen der letzte Teil des Kampfes um die Feldgleichungen vor Ihren Augen abspielt." Einstein to Arnold Sommerfeld, 9 December 1915 (CPAE 8, Doc. 161).

As was his habit, Ehrenfest pestered his friend Einstein with questions about the new theory.[114] Lorentz, who had already filled an uncounted number of pages with calculations on the *Entwurf* theory and had cast the theory in Lagrangian form (Lorentz 1915), immediately went to work on the new theory and sent Einstein three letters with comments and queries.[115] One topic of discussion was the hole argument, which Einstein had silently and unceremoniously dropped upon his return to general covariance in November 1915.[116] For our purposes in this paper, the interesting part of the discussion concerns the relation between the field equations and energy-momentum conservation and the necessity of the trace term. At one point in his correspondence with Ehrenfest, Einstein refers to "the warrant demanded by you for the inevitability of the additional term $-(1/2)g_{im}T$."[117]

Ehrenfest's obstinacy paid off. Einstein finally broke down and sent him a lengthy self-contained version of the argument that before had to be pieced together from the papers of November 1915. As Einstein promised at the beginning of the letter: "I shall not rely on the [November 1915] papers at all but show you all the calculations."[118] After delivering on this promise, Einstein closed the letter saying:

> I assume you will have no further difficulty. Show the thing to Lorentz too, who also has not yet appreciated the necessity of the structure of the right-hand side of the field equations. Could you do me a favor and send these sheets back to me as I do not have these things so neatly in one place anywhere else.[119]

Ehrenfest presumably obliged. The letter reads like the blueprint for the sections on the field equations and energy-momentum conservation in (Einstein 1916a), the first systematic exposition of general relativity, sent to Willy Wien, the editor of *Annalen der Physik*, in March 1916[120] and published in May of that year.

As in his papers of November 1915 and in the letter to Ehrenfest, Einstein used unimodular coordinates in this paper. He started with the November Lagrangian

114 For discussion of the relationship between Einstein and Ehrenfest, see (Klein 1970, chap. 12).

115 This can be inferred from Einstein to H. A. Lorentz, 17 January 1916 (CPAE 8, Doc. 183). Unfortunately, none of the letters from Ehrenfest and Lorentz to Einstein of this period (late 1915–early 1916) seem to have been preserved. For a discussion of the three-way correspondence between Einstein, Lorentz and Ehrenfest in this period, see (Kox 1988).

116 For discussion of the hole argument and its replacement, the point-coincidence argument, and references to the extensive literature on these topics, see sec. 4 of "What Did Einstein Know …" (in this volume) and (Janssen 2005, 73–74).

117 "die von Dir verlangte Gewähr der "Zwangläufigkeit" für das Zusatzglied $-(1/2)g_{im}T$." Einstein to Paul Ehrenfest, 17 January 1916 (CPAE 8, Doc. 182).

118 "Ich stütze mich gar nicht auf die Arbeiten, sondern rechne Dir alles vor." Einstein to Paul Ehrenfest, 24 January 1916 or later (CPAE 8, Doc. 185).

119 "Du wirst nun wohl keine Schwierigkeit mehr finden. Zeige die Sache auch Lorentz, der die Notwendigkeit der Struktur der rechten Seite der Feldgleichungen auch noch nicht empfindet. Es wäre mir lieb, wenn Du mir diese Blätter dann wieder zurückgäbest, weil ich die Sachen sonst nirgends so hübsch beisammen habe." Ibid.

120 Einstein to Wilhelm Wien, 18 March 1916 (CPAE 8, Doc. 203). Einstein had told Wien in February that he was in the process of writing this paper, Einstein to Wilhelm Wien, 28 February 1916 (CPAE 8, Doc. 196).

$$H = g^{\mu\nu}\Gamma^{\alpha}_{\mu\beta}\Gamma^{\beta}_{\nu\alpha} \tag{95}$$

(Einstein 1916a, 804, eq. 47a; see eq. (63) above with $\mathfrak{L} = H$ rather than $-H$). The Euler-Lagrange equations,

$$\frac{\partial}{\partial x^{\alpha}}\left(\frac{\partial H}{\partial g^{\mu\nu}_{\alpha}}\right) - \frac{\partial H}{\partial g^{\mu\nu}} = 0 \tag{96}$$

(ibid., 805, eq. 47b), for this Lagrangian are:

$$\Gamma^{\alpha}_{\mu\nu,\alpha} + \Gamma^{\alpha}_{\mu\beta}\Gamma^{\beta}_{\nu\alpha} = 0 \tag{97}$$

(ibid., 803, eq. 47; see eqs. (63)–(67) with $T_{\mu\nu} = 0$). These are the vacuum field equations. The question is how to generalize these equations in the presence of matter. To this end, Einstein rewrote the vacuum field equations in terms of the gravitational energy-momentum pseudo-tensor $t_{\mu\nu}$. He then added the energy-momentum tensor for matter $T_{\mu\nu}$ in such a way that it enters the field equations in the exact same way as $t_{\mu\nu}$. This strategy originated in the letter to Ehrenfest. After writing down the field equations in the form of eq. (91) (eq. 8 in the letter), Einstein wrote: "This equation is interesting because it shows that the origin of the gravitational [field] lines is determined solely by the sum $T^{\nu}_{\sigma} + t^{\nu}_{\sigma}$, as one has to expect."[121]

To find $t_{\mu\nu}$, Einstein contracted the left-hand side of the eq. (96) with $g^{\mu\nu}_{\sigma}$

$$g^{\mu\nu}_{\sigma}\left(\frac{\partial}{\partial x^{\alpha}}\left(\frac{\partial H}{\partial g^{\mu\nu}_{\alpha}}\right) - \frac{\partial H}{\partial g^{\mu\nu}}\right) = 0. \tag{98}$$

Since he did not have the field equations in the presence of matter yet, Einstein could not give the usual rationale for this move. Using eq. (90), the Einstein field equations in the presence of matter in unimodular coordinates, one can rewrite the left-hand side of eq. (98) as $\kappa g^{\mu\nu}_{\sigma}(T_{\mu\nu} - (1/2)g_{\mu\nu}T)$. As we saw in sec. 7, in unimodular coordinates $g^{\mu\nu}_{\sigma}g_{\mu\nu} = 0$, so this is equal to $\kappa g^{\mu\nu}_{\sigma}T_{\mu\nu}$, which is 2κ times the gravitational force density. By writing the left hand side of eq. (98) as a divergence, Einstein could thus express the gravitational force density as the divergence of gravitational energy-momentum density. As we have seen, this was Einstein's standard procedure for introducing $t_{\mu\nu}$ (see note 41, eqs. (16)–(18), eqs. (71)–(72), and the derivation following eq. (93)).

Eq. (98) can be rewritten as $t^{\alpha}_{\sigma,\alpha} = 0$, if t^{α}_{σ} is defined as[122]

---

121 "Diese Gleichung ist interessant, weil sie zeigt, dass das Entspringen der Gravitationslinien allein durch die Summe $T^{\nu}_{\sigma} + t^{\nu}_{\sigma}$ bestimmt ist, wie man ja auch erwarten muss." Einstein to Paul Ehrenfest, 24 January 1916 or later (CPAE 8, Doc. 185).

122 Einstein (1916a, 805) added a footnote saying: "The reason for the introduction of the factor -2κ will become clear later" ("Der Grund der Einführung des Faktors -2κ wird später deutlich werden."). He is referring to the generalization of the vacuum field equations (102) to the field equations (103) in the presence of matter in sec. 16 and to the discussion of energy-momentum conservation in secs. 17–18 of his paper (ibid., 807–810).

$$-2\kappa t_{\sigma}^{\alpha} = g_{\sigma}^{\mu\nu}\frac{\partial H}{\partial g_{\alpha}^{\mu\nu}} - \delta_{\sigma}^{\alpha}H \tag{99}$$

(Einstein 1916a, 805, eq. 49). Substituting eq. (66) (with $\mathfrak{L} = H$) and eq. (95) into eq. (99), one finds (cf. eq. (73))

$$\kappa t_{\sigma}^{\alpha} = \frac{1}{2}\delta_{\sigma}^{\alpha}g^{\mu\nu}\Gamma_{\beta\mu}^{\rho}\Gamma_{\rho\nu}^{\beta} - g^{\mu\nu}\Gamma_{\mu\beta}^{\alpha}\Gamma_{\nu\sigma}^{\beta} \tag{100}$$

(ibid., 806, eq. 50).

Einstein now used t_{ν}^{μ} to rewrite the field equations (97). The trace of the pseudo-tensor is (see eq. (85)):

$$\kappa t = g^{\mu\nu}\Gamma_{\mu\beta}^{\alpha}\Gamma_{\nu\alpha}^{\beta}.$$

Eq. (100) can thus be rewritten as

$$\kappa\Big(t_{\sigma}^{\alpha} - \frac{1}{2}\delta_{\sigma}^{\alpha}t\Big) = -g^{\mu\nu}\Gamma_{\mu\beta}^{\alpha}\Gamma_{\nu\sigma}^{\beta}. \tag{101}$$

The contraction of the vacuum field equations (97) with $g^{\nu\sigma}$,

$$g^{\nu\sigma}(\Gamma_{\mu\nu,\alpha}^{\alpha} + \Gamma_{\mu\beta}^{\alpha}\Gamma_{\nu\alpha}^{\beta}) = 0,$$

can be rewritten as (see footnote 89):

$$(g^{\nu\sigma}\Gamma_{\mu\nu}^{\alpha})_{,\alpha} - g^{\rho\tau}\Gamma_{\rho\beta}^{\sigma}\Gamma_{\tau\mu}^{\beta} = 0.$$

Using eq. (101) for the second term, one can thus write the vacuum field equations in unimodular coordinates as

$$(g^{\nu\sigma}\Gamma_{\mu\nu}^{\alpha})_{,\alpha} = -\kappa\Big(t_{\mu}^{\sigma} - \frac{1}{2}\delta_{\mu}^{\sigma}t\Big) \tag{102}$$

(ibid., 806, eq. 51).

On the argument that $T_{\mu\nu}$ should enter the field equations in the exact same way as $t_{\mu\nu}$, Einstein generalized the vacuum equations to

$$(g^{\nu\sigma}\Gamma_{\mu\nu}^{\alpha})_{,\alpha} = -\kappa\Big([t_{\mu}^{\sigma} + T_{\mu}^{\sigma}] - \frac{1}{2}\delta_{\mu}^{\sigma}[t + T]\Big) \tag{103}$$

in the presence of matter (ibid., 807, eq. 52). Since eq. (102) is just an alternative way of writing the vacuum field equations contracted with $g^{\nu\sigma}$, eq. (103) is equivalent to

$$\Gamma_{\mu\nu,\alpha}^{\alpha} + \Gamma_{\mu\beta}^{\alpha}\Gamma_{\nu\alpha}^{\beta} = -\kappa\Big(T_{\mu\nu} - \frac{1}{2}g_{\mu\nu}T\Big)$$

(ibid., 808, eq. 53; cf. eqs. (90)). These equations are more easily recognized as the generally-covariant Einstein field equations (94) in unimodular coordinates.

In the next section of his paper, Einstein (1916a, sec. 17) showed that energy-momentum conservation in the form $(T_\sigma^\alpha + t_\sigma^\alpha)_{,\alpha} = 0$ is a direct consequence of the field equations (103). Fully contracting eq. (103), one finds

$$(g^{\nu\sigma}\Gamma^\alpha_{\sigma\nu})_{,\alpha} = \kappa(t+T), \tag{104}$$

with the help of which eq. (103) itself can be rewritten as

$$\frac{\partial}{\partial x^\alpha}\Big(g^{\nu\sigma}\Gamma^\alpha_{\mu\nu} - \frac{1}{2}\delta^\sigma_\mu[g^{\lambda\beta}\Gamma^\alpha_{\lambda\beta}]\Big) = -\kappa(t^\sigma_\mu + T^\sigma_\mu).$$

The field equations thus guarantee energy-momentum conservation if

$$B_\mu \equiv \frac{\partial^2}{\partial x^\sigma \partial x^\alpha}\Big(g^{\nu\sigma}\Gamma^\alpha_{\mu\nu} - \frac{1}{2}\delta^\sigma_\mu[g^{\lambda\beta}\Gamma^\alpha_{\lambda\beta}]\Big) = 0. \tag{105}$$

This equation, it turns out, is an identity. The first term can be rewritten as[123]

$$(g^{\nu\sigma}\Gamma^\alpha_{\mu\nu})_{,\alpha\sigma} = \frac{1}{2}g^{\alpha\beta}_{\alpha\beta\mu};$$

the second as minus this same expression.[124] Eq. (106) gives the contracted Bianchi identities $(R^{\mu\nu} - (1/2)g^{\mu\nu}R)_{;\nu} = 0$ in unimodular coordinates.

Einstein (1916a, 809) finally showed that energy-momentum conservation in the form $(T^\alpha_\sigma + t^\alpha_\sigma)_{,\alpha} = 0$ is equivalent to the energy-momentum balance equation $T^\alpha_{\sigma,\alpha} + (1/2)g^{\mu\nu}_\sigma T_{\mu\nu} = 0$ (see the derivation following eq. (93)).

Three short sections of the review paper (Einstein 1916a, part C, secs. 15–18, pp. 804–810) thus provided a streamlined version of an argument that had been hard to piece together from the four papers of November 1915 even for the likes of Lorentz and Ehrenfest.

---

123 Using the definition (53) of $\Gamma^\alpha_{\mu\nu}$ as minus the Christoffel symbols, one can write

$$(g^{\nu\sigma}\Gamma^\alpha_{\mu\nu})_{,\alpha\sigma} = -\frac{1}{2}(g^{\sigma\beta}g^{\alpha\lambda}(g_{\lambda\mu,\beta} + g_{\lambda\beta,\mu} - g_{\mu\beta,\lambda}))_{,\alpha\sigma}$$

The combination of the first and the third term in the innermost parentheses are anti-symmetric in λ and β. They are contracted with a quantity symmetric in these same indices (cf. footnote 92 above).

$$(g^{\sigma\beta}g^{\alpha\lambda}(\ldots))_{,\alpha\sigma} = (g^{\alpha\beta}g^{\sigma\lambda}(\ldots))_{,\sigma\alpha} = (g^{\sigma\lambda}g^{\alpha\beta}(\ldots))_{,\alpha\sigma}$$

The expression above thus reduces to

$$(g^{\nu\sigma}\Gamma^\alpha_{\mu\nu})_{,\alpha\sigma} = -\frac{1}{2}(g^{\sigma\beta}g^{\alpha\lambda}g_{\lambda\beta,\mu})_{,\alpha\sigma} = \frac{1}{2}g^{\alpha\sigma}_{\mu\alpha\sigma}.$$

124 Using eq. (53), one can write

$$-\frac{1}{2}(\delta^\sigma_\mu g^{\lambda\beta}\Gamma^\alpha_{\lambda\beta})_{,\alpha\sigma} = -\frac{1}{2}(g^{\lambda\beta}\Gamma^\alpha_{\lambda\beta})_{,\alpha\mu} = \frac{1}{4}(g^{\lambda\beta}g^{\alpha\delta}(g_{\delta\lambda,\beta} + g_{\delta\beta,\lambda} - g_{\lambda\beta,\delta}))_{,\alpha\mu}.$$

Since $g^{\lambda\beta}g_{\lambda\beta,\delta} = (\log\sqrt{-g})_{,\delta}$ (Einstein 1916a, 796, eq. 29) vanishes for unimodular coordinates, the expression above reduces to

$$-\frac{1}{2}(\delta^\sigma_\mu g^{\lambda\beta}\Gamma^\alpha_{\lambda\beta})_{,\alpha\sigma} = \frac{1}{2}(g^{\lambda\beta}g^{\alpha\delta}g_{\delta\lambda,\beta})_{,\alpha\mu} = -\frac{1}{2}g^{\alpha\beta}_{\beta\alpha\mu}.$$

9. FROM THE NOVEMBER LAGRANGIAN TO THE RIEMANN SCALAR: GENERAL COVARIANCE AND ENERGY-MOMENTUM CONSERVATION

Both in the papers of November 1915 and in the review article that following March (Einstein 1916a), Einstein used the November Lagrangian—i.e., the Entwurf Lagrangian with the components of the gravitational field redefined as minus the Christoffel symbols—to derive the gravitational part of the Einstein field equations in unimodular coordinates. The use of unimodular coordinates clearly brings out the relation between the old and the new field equations (see the appendix), but complicates the use of the general formalism of (Einstein 1914c). Most seriously affected is the discussion of energy-momentum conservation. It was only in unimodular coordinates that Einstein was able to show that the field equations guarantee energy-momentum conservation in unimodular coordinates. He also did not make the connection between energy-momentum conservation and covariance of the field equations. In (Einstein 1916c) the November Lagrangian is replaced by a Lagrangian based on the Riemann scalar. Applying the formalism of (Einstein 1914c), Einstein now showed that the general covariance of the Einstein field equations guarantees that energy-momentum conservation holds in arbitrary coordinates. This is expressed in the conditions $S_\sigma^\nu = 0$ *and* $B_\mu = 0$ *. The latter are just the contracted Bianchi identities.*

Shortly after the triumphs of November 1915, Einstein acknowledged the desirability of deriving (the gravitational part of) the generally-covariant form of the field equations from a variational principle. In the November 1915 papers, as we have seen, he had only done so in unimodular coordinates (Einstein 1915a, 784). He realized that the generally-covariant Lagrangian would have to come from the Riemann curvature scalar. He also realized that terms with second-order derivatives of the metric in the Riemann scalar, which would lead to terms with third-order derivatives in the field equations, could be eliminated from the action through partial integration. He concluded that the effective Lagrangian had to be

$$L = \sqrt{-g}[g^{\sigma\tau}\Gamma^\beta_{\sigma\alpha}\Gamma^\alpha_{\tau\beta} - g^{\alpha\beta}\Gamma^\sigma_{\alpha\beta}\Gamma^\rho_{\sigma\rho}].$$

All this can be found in a letter to Lorentz of January 17, 1916. Einstein wrote that he had only gone through the calculation once, but the expression above is actually correct. He also told Lorentz that he had not attempted to derive the corresponding Euler-Lagrange equations: "The calculation of $\partial L/\partial g^{\mu\nu}$ and $(\partial L/\partial g_\sigma^{\mu\nu})$, however, is rather cumbersome, at least with my limited proficiency in calculating."[125]

When he wrote the review article (Einstein 1916a) less than two months later, he apparently still did not have the stomach for this cumbersome though straightforward

125 "Die Berechnung von $\partial L/\partial g^{\mu\nu}$ und $\partial L/\partial g_\sigma^{\mu\nu}$ ist aber ziemlich beschwerlich, wenigstens bei meiner geringen Sicherheit im Rechnen." Einstein to H. A. Lorentz, 17 January 1916 (CPAE 8, Doc. 183). He explicitly said he had not done the calculation in another letter to Lorentz two days later (Einstein to H. A. Lorentz, 19 January 1916 [CPAE 8, Doc. 184]).

calculation. As we have seen in sec. 8, both the presentation of the field equations and the discussion of energy-momentum conservation in secs. 14–18 of (Einstein 1916a) are in terms of unimodular coordinates.

Einstein may originally have planned to cover this material in arbitrary coordinates. This is suggested by a manuscript for an ultimately discarded five-page appendix to the review article (CPAE 6, Doc. 31). At the top of the first page of the manuscript, we find "§14" which was subsequently deleted and replaced by "Appendix: Formulation of the theory based on a variational principle."[126] Sec. 14 is the first of the five sections in (Einstein 1916a) on the field equations and energy-momentum conservation. In the manuscript under consideration here, Einstein gave a variational derivation of the field equations in arbitrary coordinates along the lines sketched in the letter to Lorentz discussed above. He still did not explicitly evaluate the Euler-Lagrange equations. But he did write down the quantities S_σ^ν and B_σ of the general formalism of (Einstein 1914c) for the effective Lagrangian of the new theory (now denoted by $\mathfrak{G}$). He pointed out that the general covariance of the Riemann scalar guarantees that these quantities vanish identically. He did not mention that this automatically implies energy-momentum conservation (see sec. 3.2 and 3.3). This might simply be because Einstein did not bother to finish this manuscript once he had decided to rewrite sec. 14 and the remainder of part C of his review article in unimodular coordinates. The original generally-covariant treatment was relegated to an appendix, which ultimately did not make it into the published paper. Einstein returned to it a few months later, revised and completed the manuscript, and submitted it to the Prussian Academy on October 26, 1916. This paper, (Einstein 1916c), will be the main focus of this section. But first we discuss some of Einstein's pronouncements on the topic in the intervening months.

Ehrenfest must have taken Einstein to task for using unimodular coordinates in the crucial sections of the review article (Einstein 1916a). In May 1916, shortly after the article was published, Einstein wrote to his friend in Leyden defensively: "My specialization of the coordinate system is not *just* based on laziness."[127] Did Einstein have some reason to believe that the choice of unimodular coordinates was not just convenient but physically meaningful?[128] At the time of this letter to Ehrenfest, it may not have been more than an inkling, but a month later Einstein actually published an argument purporting to show that unimodular coordinates are indeed physically privileged (Einstein 1916b).

Given that general relativity had been developed in analogy with electrodynamics (see sec. 3.1 and 3.2), it was only natural for Einstein to explore the possibility of

126 "Anhang: Darstellung der Theorie ausgehend von einem Variationsprinzip." (CPAE 6, Doc. 31, [p. 1]).

127 "Meine Spezialisierung des Bezugssystems beruht nicht *nur* auf Faulheit." Einstein to Paul Ehrenfest, 24 May 1916 (CPAE 8, Doc. 220).

128 In December 1915, Einstein had called the choice of unimodular coordinates "epistemologically meaningless" ("erkenntnistheoretisch ohne Bedeutung," Einstein to Moritz Schlick, 14 December 1915 [CPAE 8, Doc. 165]).

gravitational waves in his theory. This is what he did in (Einstein 1916b). He found three types of waves, two of which curiously do not transport energy (Einstein 1916b, 693). In an addendum to the paper, Einstein noted that these spurious waves can be eliminated by choosing unimodular coordinates. This, he concluded, shows that the choice of unimodular coordinates has "a deep physical justification."[129] He also rehearsed this argument in a letter to De Sitter, another member of the Leyden group around Lorentz and Ehrenfest working on relativity.[130]

Two letters from Leyden to Berlin the following year suggest that, in late 1917, Einstein still believed that unimodular coordinates have a special status. The author of these letters was Gunnar Nordström, who was in Leyden on a three-year fellowship (CPAE 8, Doc. 112, note 3). Nordström had a hard time convincing Einstein that in unimodular coordinates the gravitational field of the sun carries no energy.[131] Nordström also caught an error in (Einstein 1916b), which prompted Einstein to publish a corrected version of his 1916 paper on gravitational waves (Einstein 1918a). He now used a different argument to eliminate the spurious gravitational waves, one that makes no mention of unimodular coordinates (ibid., 160–161). By the time Gustav Mie, in his efforts to convince Einstein of the need for special coordinates, reminded him of the original argument, Einstein had abandoned the notion of privileged coordinates, unimodular or otherwise, altogether.[132]

Immediately following the defensive passage from the letter to Ehrenfest of May 1916 quoted above (see footnote 127), Einstein promised: "At some point I may present the matter without such specialization [of the coordinates], along the lines of [Lorentz 1915]."[133] Given Einstein's views at the time that unimodular coordinates were special, such a fully generally-covariant presentation of the theory was probably not a matter of great urgency to him. He was nonetheless forced to keep thinking about the issue, not by Ehrenfest this time but by a new correspondent.

In June 1916, Théophile de Donder, professor of mathematical physics in Brussels, respectfully informed Einstein that the latter's expression for the gravitational energy-momentum pseudo-tensor was wrong (Théophile de Donder to Einstein, 27 June 1916 [CPAE 8, Doc. 228]). An exchange of letters across enemy lines ensued, mercifully cut short a little over a month later by exhaustion on the Belgian side. De Donder began what would turn out to be the last letter of this testy correspondence with the announcement of a truce of sorts: "the extensive research and innumerable calculations that I have devoted to your theory have forced me to take some rest for a few weeks."[134] Einstein probably read this with a sigh of relief. Even though De

---

129 "eine tiefe physikalische Berechtigung" (Einstein 1916b, 696).

130 Einstein to Willem de Sitter, 22 June 1916 (CPAE 6, Doc. 32). The work of the Leyden group is described in (Kox 1992).

131 Gunnar Nordström to Einstein, 22–28 September 1917, 23 October 1917 (CPAE 8, Docs. 382, 393).

132 Gustav Mie to Einstein, 6 May 1918 (CPAE 8, Doc. 532). For brief discussions of the episode described in the last two paragraphs, see (CPAE 8, li-lii, and CPAE 7, xxv).

133 "Vielleicht werde ich die Sache auch einmal ohne die Spezialisierung darstellen, so wie Lorentz in seiner Arbeit."

Donder's missives were ostensibly about clarifying the relation of his own work to Einstein's, it is hard not to get the impression that De Donder's ulterior motive was to have Einstein concede priority for at least part of general relativity to his Belgian colleague.[135] If this was indeed De Donder's hidden agenda, he must have been bitterly disappointed by the letters from Berlin. The way Einstein saw it, De Donder had simply overlooked that the expressions whose correctness he was contesting only held in unimodular coordinates. In Einstein's last contribution to the debate, he showed how one would obtain the expression for the gravitational energy-momentum pseudo-tensor without choosing unimodular coordinates. The formula that Einstein gives is

$$\mathrm{t}_\sigma^\nu = -\frac{1}{2\kappa}\left(\frac{\partial L^*}{\partial g_\nu^{\alpha\beta}} g_\sigma^{\alpha\beta} - \delta_\sigma^\nu L^*\right).$$

L^* is Einstein's notation in this letter for the effective Lagrangian extracted from the Riemann scalar.[136] The letter shows that Einstein had no trouble with the continuation of the argument of the appendix to (Einstein 1916a) discussed above.

In the fall of 1916, Einstein finally finished what he had begun in this appendix and published a generally-covariant discussion of the field equations and energy-momentum conservation. The paper, (Einstein 1916c), brings together in a systematic fashion the various elements of this discussion that we encountered piecemeal in the letter to Lorentz, the discarded appendix, and the letter to De Donder. As is acknowledged in the introduction of (Einstein 1916c), both Hilbert (1915) and Lorentz (1916a) had already shown how to derive the Einstein field equations from a variational principle without choosing special coordinates.[137] Einstein's own paper owes little or nothing to this earlier work.[138] It follows the relevant sections of (Einstein 1914c) virtually step by step.

Einstein starts from the action

$$\int \mathfrak{H} d\tau = \int (\mathfrak{G} + \mathfrak{M}) d\tau,$$

134 "Les longues recherches et les innombrables calculs que j'ai consacré à vos théories m'obligent à prendre quelques semaines de repos." Théophile de Donder to Einstein, 8 August 1916 (CPAE 8, Doc. 249)

135 The following year, De Donder (1917) claimed priority for the field equations with cosmological constant of (Einstein 1917b). This prompted Einstein to write to Lorentz who had communicated De Donder's paper to the Amsterdam Academy. Clearly embarrassed to bother Lorentz with this matter, Einstein emphasized that it was because of a serious error in De Donder's paper not because of the priority claim that he urged his Dutch colleague to have De Donder publish a correction. Einstein to H. A. Lorentz, 18 December 1917 (CPAE 8, Doc. 413). We do not know whether Lorentz took up this matter with De Donder. We do know that no correction ever appeared.

136 In terms of the more explicit notation introduced in our discussion of the general formalism of (Einstein 1914c) in sec. 3, t_σ^ν would be written as $\mathrm{t}_\sigma^\nu(L^*, \text{cons})$ (cf. eq. (18)).

137 See (Sauer 1999) and "Hilbert's Foundation of Physics …" (in vol. 4 of this series) for discussion of Hilbert's work; and (Janssen 1992) for discussion of Lorentz's work.

138 See footnote 46 for discussion of how Einstein's variational techniques deviated from the standard techniques of the Göttingen crowd.

where $\mathfrak{G}$, the gravitational part of the Lagrangian, is the Riemann scalar and $\mathfrak{M}$, the Lagrangian for the material part of the system, is left unspecified (Einstein 1916c, 1111–1112).[139] Through partial integration, all terms involving second-order derivatives of the metric can be removed from the integral over $\mathfrak{G}$. The gravitational part of the field equations thus follows from the variational principle

$$\delta\int\mathfrak{G}^* d\tau = 0, \tag{106}$$

where $\mathfrak{G}^*$ is the effective Lagrangian we encountered at the beginning of this section

$$\mathfrak{G}^* = \sqrt{-g}g^{\mu\nu}\left[\begin{Bmatrix}\beta\\\mu\alpha\end{Bmatrix}\begin{Bmatrix}\alpha\\\nu\beta\end{Bmatrix}-\begin{Bmatrix}\alpha\\\mu\nu\end{Bmatrix}\begin{Bmatrix}\beta\\\alpha\beta\end{Bmatrix}\right] \tag{107}$$

(Ibid., 1113, note 2). For $\sqrt{-g} = 1$, this expression for $\mathfrak{G}^*$ reduces to expression (95) for the Lagrangian L in unimodular coordinates used in (Einstein 1916a).[140]

The field equations are the Euler-Lagrange equations

$$\frac{\partial}{\partial x^\alpha}\left(\frac{\partial\mathfrak{G}^*}{\partial g^{\mu\nu}_\alpha}\right) - \frac{\partial\mathfrak{G}^*}{\partial g^{\mu\nu}} = \frac{\partial\mathfrak{M}}{\partial g^{\mu\nu}} \tag{108}$$

(Einstein 1916c, 1113, eq. 7). Einstein still did not bother to evaluate the functional derivatives $\partial\mathfrak{G}^*/\partial g^{\mu\nu}_\alpha$ and $\partial\mathfrak{G}^*/\partial g^{\mu\nu}$ to show that the left-hand side reproduces the Einstein tensor (or rather, the corresponding tensor density). The right-hand side gives minus the energy-momentum tensor density for matter:[141]

---

139 As Einstein writes in the introduction: "In particular, specific assumptions about the constitution of matter should be kept to a minimum, in contrast especially to Hilbert's presentation" ("Insbesondere sollen über die Konstitution der Materie möglichst wenig spezialisierende Annahmen gemacht werden, im Gegensatz besonders zur Hilbertschen Darstellung." Einstein 1916c, 1111). Following Mie, Hilbert (1915) had endorsed the electromagnetic worldview, according to which the matter Lagrangian is a function only of $g^{\mu\nu}$ and of the components A_μ of the electromagnetic four-vector potential and their first-order derivatives. Einstein had tired of this electromagnetic program almost as fast as he had become enamored of it in November 1915. In a footnote to the discarded appendix to (Einstein 1916a), he characterized Hilbert's approach as "not very promising" ("wenig aussichtsvoll"). This phrase was meant for public consumption. He was much more dismissive of Hilbert's work in the letter to Ehrenfest from which we already quoted in footnotes 127 and 133 (see also footnote 152 below). And reporting on (Einstein 1916c) to Weyl, Einstein bluntly wrote: "Hilbert's assumption about matter seems infantile to me, in the sense of a child innocent of the deceit of the outside world" ("Der Hilbertsche Ansatz für die Materie erscheint mir kindlich, im Sinne des Kindes, das keine Tükken der Aussenwelt kennt," Einstein to Hermann Weyl, 23 November 1916 [CPAE 8, Doc. 278]).

140 The second term in square brackets in eq. (107) vanishes since $\begin{Bmatrix}\alpha\\\alpha\beta\end{Bmatrix} = (\lg\sqrt{-g})_{,\beta}$.

141 Factors of κ, the gravitational constant, are not to be found in (Einstein 1916c). Presumably, they are absorbed into the Lagrangians $\mathfrak{G}^*$ and $\mathfrak{M}$ for matter and gravitational field.

$$\mathfrak{T}_{\mu\nu} \equiv -\frac{\partial \mathfrak{M}}{\partial g^{\mu\nu}} \tag{109}$$

(Einstein 1916c, 1115, eq. 19).

Substituting this definition into eq. (108) and contracting the resulting equations with $g^{\mu\sigma}$, one arrives at:

$$g^{\mu\sigma}\left(\frac{\partial}{\partial x^{\alpha}}\left(\frac{\partial \mathfrak{G}^*}{\partial g^{\mu\nu}_{\alpha}}\right) - \frac{\partial \mathfrak{G}^*}{\partial g^{\mu\nu}}\right) = -\mathfrak{T}^{\sigma}_{\nu}.$$

As we have seen in sec. 3.1 (eqs. (11)–(13)), this equation can be rewritten as

$$\frac{\partial}{\partial x^{\alpha}}\left(g^{\mu\sigma}\frac{\partial \mathfrak{G}^*}{\partial g^{\mu\nu}_{\alpha}}\right) = -(\mathfrak{T}^{\sigma}_{\nu} + \mathfrak{t}^{\sigma}_{\nu}) \tag{110}$$

(ibid., eq. 18), if $\mathfrak{t}^{\sigma}_{\nu}$, the gravitational energy-momentum pseudo-tensor, is defined as:

$$\mathfrak{t}^{\sigma}_{\nu} \equiv - g^{\mu\sigma}_{\alpha}\frac{\partial \mathfrak{G}^*}{\partial g^{\mu\nu}_{\alpha}} - g^{\mu\sigma}\frac{\partial \mathfrak{G}^*}{\partial g^{\mu\nu}} \tag{111}$$

(ibid., eq. 20, first part).

Energy-momentum conservation in the form

$$(\mathfrak{T}^{\sigma}_{\nu} + \mathfrak{t}^{\sigma}_{\nu})_{,\sigma} = 0 \tag{112}$$

(ibid., eq. 21) is guaranteed through eq. (110) if

$$B_{\nu} \equiv \frac{\partial^2}{\partial x^{\sigma}\partial x^{\alpha}}\left(g^{\mu\sigma}\frac{\partial \mathfrak{G}^*}{\partial g^{\mu\nu}_{\alpha}}\right) = 0 \tag{113}$$

(ibid., eq. 17; Einstein does not use the notation B_{μ} in this paper).

Eq. (112) is equivalent to energy-momentum conservation in the form $T^{\mu\nu}{}_{;\nu} = 0$, or, equivalently (see eqs. (14)–(16) and footnote 35),

$$\mathfrak{T}^{\sigma}_{\nu,\sigma} + \frac{1}{2}g^{\alpha\beta}_{\nu}\mathfrak{T}_{\alpha\beta} = 0$$

(ibid., 1116, eq. 22), if

$$\mathfrak{t}^{\sigma}_{\nu,\sigma} = \frac{1}{2}g^{\alpha\beta}_{\nu}\mathfrak{T}_{\alpha\beta}.$$

Using definition (109) of $\mathfrak{T}_{\alpha\beta}$ and the field eqs. (108), one can rewrite this as

$$\mathfrak{t}^{\sigma}_{\nu,\sigma} = -\frac{1}{2}g^{\alpha\beta}_{\nu}\left(\frac{\partial}{\partial x^{\lambda}}\left(\frac{\partial \mathfrak{G}^*}{\partial g^{\alpha\beta}_{\lambda}}\right) - \frac{\partial \mathfrak{G}^*}{\partial g^{\alpha\beta}}\right).$$

As we have seen in sec. 3.2 (eqs. (17)–(18)), this equation holds, if $\mathfrak{t}^{\sigma}_{\nu}$ is defined as

$$\mathfrak{t}_\nu^\sigma \equiv \frac{1}{2}\left(\delta_\nu^\sigma \mathfrak{G}^* - g_\nu^{\alpha\beta}\frac{\partial \mathfrak{G}^*}{\partial g_\sigma^{\alpha\beta}}\right) \tag{114}$$

(ibid., 1115, eq. 20, second part).

Compatibility of the definitions (111) and (114) of $\mathfrak{t}_\nu^\sigma$ requires that[142]

$$S_\sigma^\nu \equiv 2g_\alpha^{\beta\nu}\frac{\partial \mathfrak{G}^*}{\partial g_\alpha^{\beta\sigma}} + 2g^{\alpha\nu}\frac{\partial \mathfrak{G}^*}{\partial g^{\alpha\sigma}} + \delta_\sigma^\nu \mathfrak{G}^* - g_\sigma^{\alpha\beta}\frac{\partial \mathfrak{G}^*}{\partial g_\nu^{\alpha\beta}} = 0. \tag{115}$$

Energy-momentum conservation thus requires both eq. (113) and eq. (115):

$$B_\sigma = 0, \quad S_\sigma^\nu = 0 \tag{116}$$

(see sec. 3.2, eq. (23)).

Before these considerations of energy-momentum conservation Einstein (1916c, 1114–1115) has already shown that both equations are satisfied identically for $\mathfrak{G}^*$ as a consequence of the general covariance of the action in eq. (106).[143] Consider a coordinate transformation $x'^\mu = x^\mu + \Delta x^\mu$, where the Δx^μ are chosen such that they vanish outside of some arbitrarily chosen region of spacetime. Since the integral over $\mathfrak{G}^*$ only differs by surface terms from the integral over the Riemann scalar $\mathfrak{G}$, the invariance of the latter under coordinate transformations $x'^\mu = x^\mu + \Delta x^\mu$ implies that the former is invariant under such transformations as well. Hence,

$$0 = \Delta\int \mathfrak{G}^* d\tau = \int \Delta\left(\frac{\mathfrak{G}^*}{\sqrt{-g}}\right)\sqrt{-g}\,d\tau \tag{117}$$

(the second step is justified because $\sqrt{-g}\,d\tau$ is an invariant volume element). The integrand is the sum of two terms:

$$\sqrt{-g}\,\Delta\left(\frac{\mathfrak{G}^*}{\sqrt{-g}}\right) = \Delta\mathfrak{G}^* + \sqrt{-g}\,\Delta\left(\frac{1}{\sqrt{-g}}\right)\mathfrak{G}^*. \tag{118}$$

The first term can be written as (see sec. 3.3, eq. (28))

$$\Delta\mathfrak{G}^* = \left\{2g^{\alpha\nu}\frac{\partial \mathfrak{G}^*}{\partial g^{\alpha\sigma}} + 2g_\alpha^{\beta\nu}\frac{\partial \mathfrak{G}^*}{\partial g_\alpha^{\beta\sigma}} - g_\sigma^{\alpha\beta}\frac{\partial \mathfrak{G}^*}{\partial g_\nu^{\alpha\beta}}\right\}\frac{\partial \Delta x^\sigma}{\partial x^\nu} + 2\frac{\partial \mathfrak{G}^*}{\partial g_\alpha^{\mu\sigma}}g^{\mu\nu}\frac{\partial^2 \Delta x^\sigma}{\partial x^\nu \partial x^\alpha};$$

the second term as[144]

142 In the more explicit notation of sec. 3, eq. (111) defines $\mathfrak{t}_\nu^\sigma(\mathfrak{G}^*, \text{source})$ and eq. (114) defines $\mathfrak{t}_\nu^\sigma(\mathfrak{G}^*, \text{cons})$. Compatibility requires that $S_\nu^\sigma(\mathfrak{G}^*) = 2\mathfrak{t}_\nu^\sigma(\mathfrak{G}^*, \text{cons}) - 2\mathfrak{t}_\nu^\sigma(\mathfrak{G}^*, \text{source}) = 0$. Note that definition (115) of S_ν^σ, which is the one given in (Einstein 1916c, 1114, eq. 14), differs by a factor 2 from definition (19), which is the one given in (Einstein 1914c, 1075, eq. 76a).

143 In the discarded appendix to (Einstein 1916a), the covariance properties are also discussed first. Einstein never gets to energy-momentum conservation in that document (CPAE 6, Doc. 31).

$$\sqrt{-g}\Delta\left(\frac{1}{\sqrt{-g}}\right)\mathfrak{G}^* = \mathfrak{G}^*\delta^\nu_\sigma\frac{\partial\Delta x^\sigma}{\partial x^\nu}.$$

Inserting these expressions into eq. (118), one finds

$$\sqrt{-g}\Delta\left(\frac{\mathfrak{G}^*}{\sqrt{-g}}\right) = \left\{2g^{\beta\nu}_\alpha\frac{\partial\mathfrak{G}^*}{\partial g^{\beta\sigma}_\alpha} + 2g^{\alpha\nu}\frac{\partial\mathfrak{G}^*}{\partial g^{\alpha\sigma}} + \delta^\nu_\sigma\mathfrak{G}^* - g^{\alpha\beta}_\sigma\frac{\partial\mathfrak{G}^*}{\partial g^{\alpha\beta}_\nu}\right\}\frac{\partial\Delta x^\sigma}{\partial x^\nu} + 2\frac{\partial\mathfrak{G}^*}{\partial g^{\mu\sigma}_\alpha}g^{\mu\nu}\frac{\partial^2\Delta x^\sigma}{\partial x^\nu\partial x^\alpha}. \tag{119}$$

The expression in curly brackets is just S^ν_σ as defined in eq. (115). Eq. (119) can thus be written more compactly as

$$\sqrt{-g}\Delta\left(\frac{\mathfrak{G}^*}{\sqrt{-g}}\right) = S^\nu_\sigma\frac{\partial\Delta x^\sigma}{\partial x^\nu} + 2\frac{\partial\mathfrak{G}^*}{\partial g^{\mu\sigma}_\alpha}g^{\mu\nu}\frac{\partial^2\Delta x^\sigma}{\partial x^\nu\partial x^\alpha} \tag{120}$$

(Einstein 1916c, 1114, eq. 13). Since $\mathfrak{G}^*$ transforms as a scalar under arbitrary linear transformations,

$$S^\nu_\sigma = 0 \tag{121}$$

(ibid., eq. 15). The only contribution to the action comes from the second term on the right-hand side of eq. (120). Through partial integration this contribution can be rewritten as

$$\int 2\frac{\partial^2}{\partial x^\nu\partial x^\alpha}\left(\frac{\partial\mathfrak{G}^*}{\partial g^{\mu\sigma}_\alpha}g^{\mu\nu}\right)\Delta x^\sigma d\tau$$

plus surface terms that will vanish. The invariance of $\int\mathfrak{G}^* d\tau$ thus implies that

$$B_\sigma = \frac{\partial^2}{\partial x^\nu\partial x^\alpha}\left(\frac{\partial\mathfrak{G}^*}{\partial g^{\mu\sigma}_\alpha}g^{\mu\nu}\right) = 0. \tag{122}$$

These are the contracted Bianchi identities. The general covariance of (the Lagrangian for) the Einstein field equations thus results in two identities—$S^\nu_\sigma = 0$ and $B_\sigma = 0$—that guarantee energy-momentum conservation (see eq. (116)).

144 Using that $\Delta\sqrt{-g} = -\frac{1}{2}\sqrt{-g}g_{\mu\nu}\Delta g^{\mu\nu}$ (eq. (38)), one can write

$$\Delta\left(\frac{1}{\sqrt{-g}}\right) = \frac{1}{g}\Delta\sqrt{-g} = \frac{1}{2\sqrt{-g}}g_{\mu\nu}\Delta g^{\mu\nu}.$$

Using that $\Delta g^{\mu\nu} = g^{\mu\alpha}\frac{\partial\Delta x^\nu}{\partial x^\alpha} + g^{\nu\alpha}\frac{\partial\Delta x^\mu}{\partial x^\alpha}$ (eq. (24) and Einstein 1916c, 1114, eq. 11), one finds:

$$\sqrt{-g}\Delta\left(\frac{1}{\sqrt{-g}}\right) = \frac{1}{2}\left(2g_{\mu\nu}g^{\mu\alpha}\frac{\partial\Delta x^\nu}{\partial x^\alpha}\right) = \delta^\alpha_\nu\frac{\partial\Delta x^\nu}{\partial x^\alpha},$$

from which the equation below follows.

For Einstein this was the central point of the paper. Writing to Ehrenfest, he summed up the paper as follows: "I have now given a Hamiltonian [read: variational] treatment of the essential points of general relativity as well, in order *to bring out the connection between relativity and the energy principle*" (our emphasis).[145] He told four other correspondents the same thing. Shortly after he had submitted the paper, he wrote to Besso: "You will soon receive a short paper of mine about the foundations of general relativity, in which it is shown how the requirement of relativity is connected with the energy principle. It is very amusing."[146] Similarly, he wrote to De Sitter a few days later: "Take a look at the page proofs [of (Einstein 1916c)] that I sent to Ehrenfest. There the connection between relativity postulate and energy law is brought out very clearly."[147] A little over a week later, he sent Lorentz an offprint of the paper describing it as "a short paper, in which I explained how in my opinion the relation of the conservation laws to the relativity postulate is to be understood."[148] He emphasized that the conservation laws are satisfied for any choice of the Lagrangian $\mathfrak{M}$ for matter, adding: "So the choice [of $\mathfrak{M}$] made by Hilbert appears to have no justification."[149] He made the same point in a letter to Weyl in which he once more reiterated the key point of (Einstein 1916c), namely that "[t]he connection between the requirement of general covariance and the conservation laws is also made clearer."[150]

Two years earlier Einstein had already made the connection between covariance and energy-momentum conservation in the context of the *Entwurf* theory (Einstein 1914c). He had shown that the conditions $B_\mu = 0$ and $S_\sigma^\nu = 0$ that in conjunction with the *Entwurf* field equations guarantee energy-momentum conservation also determine the class of "justified transformations" between "adapted coordinates" (see sec. 3.3). In the review article (Einstein 1916a), he had not connected the conditions guaranteeing energy-momentum conservation in unimodular coordinates to the corresponding covariance of the field equations. Instead, he had shown by direct calculation that these conditions are identically satisfied as long as unimodular coordinates

145 "Ich habe nun das Prinzipielle an der allgemeinen Relativitätstheorie auch hamiltonisch dargestellt, um den Zusammenhang zwischen Relativität und Energieprinzip zu zeigen." Einstein to Paul Ehrenfest, 24 October 1916 (CPAE 8, Doc. 269).

146 "Du erhältst bald eine kleine Arbeit von mir über die Basis der allgemeinen Relativitätstheorie, in der gezeigt wird, wie die Rel-Forderung mit dem Energieprinzip zusammenhängt. Es ist sehr amusant." Einstein to Michele Besso, 31 October 1916 (CPAE 8, Doc. 270).

147 "Sehen Sie sich die Druckbogen an, die ich Ehrenfest geschickt habe. Es kommt dort der Zusammenhang zwischen Relat. Postulat und Energiesatz besonders klar heraus." Einstein to Willem de Sitter, 4 November 1916 (CPAE 8, Doc. 273).

148 "eine kleine Arbeit, in der ich dargestellt habe, wie nach meiner Ansicht die Beziehung der Erhaltungssätze zum Relativitätspostulat aufgefasst werden soll." Einstein to H. A. Lorentz, 13 November 1916 (CPAE 8, Doc. 276).

149 "Die von Hilbert getroffene Wahl erscheint daher durch nichts gerechtfertigt." Ibid.

150 "Auch wird der Zusammenhang zwischen allgemeiner Kovarianz-Forderung und Erhaltungssätzen deutlicher." Einstein to Hermann Weyl, 23 November 1916 (CPAE 8, Doc. 278). This is the same letter in which Einstein sharply criticizes Hilbert's adherence to the electromagnetic program (see footnote 139).

are used (Einstein 1916a, sec. 17, eq. 55; cf. eq. (105) and footnotes 123 and 124). As he wrote to Ehrenfest: "In my earlier presentation [in Einstein 1916a] with $\sqrt{-g} = 1$, direct calculation establishes the identity that is here [in Einstein 1916c] presented as a consequence of the invariance [of the action]."[151] The variational treatment in arbitrary coordinates in (Einstein 1916c) thus fills two important gaps. The paper explicitly shows that energy-momentum conservation holds in arbitrary and not just in unimodular coordinates. More importantly, it establishes for the new theory what Einstein had already found for the old one, namely that there is an intimate connection between covariance and conservation laws.

It is no coincidence that the generalization of Einstein's insight—the celebrated Noether theorems—was formulated only two years later. Energy-momentum conservation in general relativity was hotly debated in Göttingen following the abstruse treatment of the topic in (Hilbert 1915).[152] In the course of his first attempt to make sense of this part of Hilbert's paper, Felix Klein (1917) claimed that energy-momentum conservation is an identity in Einstein's theory. Klein claimed—or, to be more charitable to Klein, Einstein took him to claim—that eq. (112) holds as a direct consequence of the invariance of the action (106), *independently of the field equations*. In fact, only eq. (113) is an identity (see eq. (122)). And it is only in conjunction with the field equations (110), that this identity implies energy-momentum conservation as expressed in eq. (112). Einstein immediately set Klein straight on this score.[153] This was the start of a correspondence between the two men about energy-momentum conservation in general relativity.[154]

The debate quickly shifted from the status of the identities flowing from the general covariance of the action to the (related) issue of whether or not it was acceptable in a generally-covariant theory to have a non-generally-covariant gravitational energy-momentum tensor. Unaware that Lorentz (1916b, c) and Levi-Civita (1917) had already made the same proposal, Klein suggested to define the left-hand side of the gravitational field equations as the generally-covariant gravitational energy-momentum tensor. Like Lorentz and Levi-Civita before him, Klein even wrote a paper on this proposal, which Einstein convinced him not to publish.[155]

---

151 "In meiner früheren Darstellung mit $\sqrt{-g} = 1$ wird die Identität direkt durch Ausrechnen konstatiert, welche hier als Folge der Invarianz dargestellt wird." Einstein to Paul Ehrenfest, 7 November 1916 (CPAE 8, Doc. 275).

152 In the letter to Ehrenfest from which we already quoted in footnotes 127 and 133, Einstein vented his irritation with (Hilbert 1915): "I do not care for Hilbert's presentation. It is [...] unnecessarily complicated, not honest (= Gaussian) in its structure (creating the impression of being an *übermensch* by obfuscating one's methods)" ("Hilberts Darstellung gefällt mir nicht. Sie ist [...] unnötig kompliziert, nicht ehrlich (= Gaussisch) im Aufbau (Vorspiegelung des Übermenschen durch Verschleierung der Methoden)."). Our assessment of Hilbert's paper follows "Hilbert's Foundation of Physics ..." (in vol. 4 of this series). For a more positive assessment, see (Sauer 1999).

153 See the first paragraph of Einstein to Felix Klein, 13 March 1918 (CPAE 8, Docs. 480).

154 The most interesting letters are the first three, all written in March 1918: (1) the letter cited in footnote 153; (2) Felix Klein to Einstein, 20 March 1918 (CPAE 8, Doc. 487); (3) Einstein to Felix Klein, 24 March 1918 (CPAE 8, Doc. 492).

Einstein was virtually alone at this point in his defense of the pseudo-tensor. As he stated with his usual flair for high drama in the first sentence of (Einstein 1918d): "While the general theory of relativity has met with the approval of most physicists and mathematicians, almost all my colleagues object to my formulation of the energy-momentum law."[156] Einstein was unfazed by the opposition. Drawing heavily on (Einstein 1916c), (Einstein 1918d) provides a sustained defense of the views on energy-momentum conservation that had guided Einstein in finding and consolidating the *Entwurf* theory in 1913–1914 and that had guided him again in finding and consolidating its successor theory in 1915–1916. Subsequent developments would prove Einstein right. We now know that gravitational energy-momentum is represented by a pseudo-tensor and not by a tensor because gravitational energy-momentum cannot be localized.

Klein meanwhile continued to discuss the problem of energy-momentum conservation with other Göttingen mathematicians, notably with Carl Runge and Emmy Noether—one of the wrong sex, the other too old to be sent to the front. With Runge he undertook a systematic survey of the relevant literature. These efforts resulted in two important papers on the topic (Klein 1918a, 1918b). Noether's work on the problem resulted in her seminal paper on symmetries and conservation laws (Noether 1918).[157]

The importance of (Einstein 1916c) for our story is not so much its role in the run-up to Noether's theorems, but the evidence it provides for the continuity of Einstein's reliance on the variational formalism of (Einstein 1914c) in the transition from the *Entwurf* theory to general relativity. There was no abrupt break, no sudden switch from physical to mathematical strategy. Instead, the transition was brought about by changing one key element in the formalism encoding much of the physical knowledge that went into the *Entwurf* theory and then modifying other parts of the formalism (if necessary) to accommodate the new version of this one element. Einstein himself pinpointed this one element for us. It was the definition of the components of the gravitational field. Not all modifications necessitated by changing this definition were in place by the time he published the first of his four communications of November 1915 to the Prussian Academy (Einstein 1915a). Most of them were in place by the time he published the fourth (Einstein 1915d). This Einstein made clear in his systematic exposition of the theory in (Einstein 1916a). Even this paper, how-

155 Two drafts of this paper can be found in the Klein *Nachlass* in the Niedersächsische Staats- und Universitätsbibliothek in Göttingen (see CPAE 7, Doc. 9, note 5, for more details).

156 "Während die allgemeine Relativitätstheorie bei den meisten theoretischen Physikern und Mathematikern Zustimmung gefunden hat, erheben doch fast alle Fachgenossen gegen meine Formulierung des Impuls-Energiesatzes Einspruch" (Einstein 1918d, 448). This paper pulls together and amplifies earlier comments in (Einstein 1918a, sec. 6), written in response to Levi-Civita, and (Einstein 1918b), written in response to one of two short, little-known, and inconsequential excursions into general relativity by Erwin Schrödinger (1918). For discussion of the debate over energy-momentum conservation between Einstein and Levi-Civita, see (Cattani and De Maria 1993).

157 In broad outline this story can be found in (Rowe 1999). For a particularly illuminating analysis of Noether's theorems and their applications in physics, see (Brading 2002).

ever, left at least one important question unanswered (viz., do the field equations guarantee energy-momentum conservation in arbitrary coordinates) and failed to transfer at least one important insight from the *Entwurf* theory to the new theory (viz., the relation between covariance and energy-momentum conservation). These issues were settled only with (Einstein 1916c), a paper that can be seen as the end of the consolidation phase of the theory, although one can argue that this phase was not brought to a conclusion until the publication of (Einstein 1918a, c, d).

10. HOW EINSTEIN REMEMBERED HE FOUND HIS FIELD EQUATIONS

In his papers of November 1915, Einstein introduced his new field equations by arguing that they were the natural choice given the central role of the Riemann tensor in differential geometry. The field equations are thus presented as a product of what we have called the mathematical strategy. The continuity with the Entwurf field equations, a product of the physical strategy, is lost in Einstein's presentation and the reader is left with the impression that Einstein abruptly switched from the physical to the mathematical strategy in the fall of 1915. This is exactly how Einstein himself came to remember the breakthrough of November 1915. The physics, he felt, had been nothing but a hindrance; he had been saved at the eleventh hour by the mathematics. In his later years Einstein routinely used this version of events to justify the purely mathematical approach in his work in unified field theory.

The way Einstein presented his new field equations in the first of his four papers of November 1915 (Einstein 1915a) is very different from the way we claim he found them. The paper opens with the retraction of the uniqueness argument of (Einstein 1914c) for the *Entwurf* field equations. After explaining what is wrong with this argument, Einstein writes in the third paragraph:

> For these reasons I completely lost confidence in the field equations I had constructed and looked for a way that would constrain the possibilities in a natural manner. I was thus led back to the demand of a more general covariance of the field equations, which I had abandoned with a heavy heart three years ago when I was collaborating with my friend Grossmann. In fact, back then we already came very close to the solution of the problem given below.[158]

The fourth paragraph announces a new theory in which all equations, including the field equations, are covariant under arbitrary unimodular transformations. Einstein does not explain, neither in this paragraph nor anywhere else in the paper, what made him forgo *general* covariance at this point. Our explanation is that the physical strategy pointed not to the generally-covariant Ricci tensor but to the November tensor, which only transforms as a tensor under unimodular transformations. We showed how changing the definition of the gravitational field set in motion a chain of reasoning that led from the *Entwurf* field equations to field equations based on the November tensor. Reading the passage quoted above, one would not have suspected such continuity. In fact, Einstein's revelations that he has "completely lost confidence" in the *Entwurf* field equations and that he had already come "very close to the [new] solution" three years earlier suggest a dramatic about-face. The fifth and final paragraph of Einstein's introduction confirms this impression and suggests an abrupt switch from the physical to the mathematical strategy:

> Hardly anybody who has truly understood the theory will be able to avoid coming under its spell. It is a real triumph of the method of the general differential calculus developed by Gauss, Riemann, Christoffel, Ricci, and Levi-Civita.[159]

This impression is further reinforced by the way in which the field equations are introduced in the paper. In sec. 2, on the construction of quantities transforming as tensors under unimodular transformations, Einstein shows how to extract the November tensor R_{ij} from the Ricci tensor $G_{ij} = R_{ij} + S_{ij}$ (Einstein 1915a, 782, eqs. (13), (13a), (13b)) just as he had done in the Zurich Notebook.[160] At the beginning of sec. 3, he then writes:

> After what has been said so far, it is natural to posit field equations of the form $R_{\mu\nu} = -\kappa T_{\mu\nu}$, because we already know that these equations are covariant under arbitrary transformations of determinant 1.[161]

It is only at this point that Einstein gives the Lagrangian formulation of these field equations (Einstein 1915a, 784, eq. (17)) and goes through the argument demonstrating that they are compatible with energy-momentum conservation (see sec. 6).[162]

Not surprisingly in light of the above, some of the best modern commentators on (Einstein 1915a) have concluded that its author had abruptly switched strategies in the fall of 1915. The clearest, most concise and most explicit version of this account

---

158 "Aus diesen Gründen verlor ich das Vertrauen zu den von mir aufgestellten Feldgleichungen vollständig und suchte nach einem Wege, der die Möglichkeiten in einer natürlichen Weise einschränkte. So gelangte ich zu der Forderung einer allgemeineren Kovarianz der Feldgleichungen zurück, von der ich vor drei Jahren, als ich zusammen mit meinem Freunde Grossmann arbeitete, nur mit schwerem Herzen abgegangen war. In der Tat waren wir damals der im nachfolgenden gegebenen Lösung des Problems bereits ganz nahe gekommen" (Einstein 1915a, p. 778). The Zurich Notebook shows that Einstein and Grossmann did indeed consider field equations based on the November tensor three years earlier (see sec. 2). Norton (2000, 150) inaccurately translates "gelangte … zurück" as "went back" rather than as "was led back." The difference is not unimportant. Norton's "went back" conveys discontinuity: Einstein abandoned one approach and adopted another, characterized by "the demand of a more general covariance." On this reading "a more general covariance" sounds odd. One would have expected "general covariance." Our "was led back" conveys continuity: staying the course Einstein ended up with "the demand of a more general covariance." On our reading "demand" sounds odd. One would have expected "property" or "feature" instead. Our reading, however, does fit with Einstein's remark quoted at the beginning of sec. 3 (see footnote 32): "The series of my papers on gravitation is a chain of erroneous paths, which nonetheless gradually brought me closer to my goal." Einstein to H.A. Lorentz, 17 January 1916 (CPAE 8, Doc. 183)

159 "Dem Zauber dieser Theorie wird sich kaum jemand entziehen können, der sie wirklich erfaßt hat; sie bedeutet einen wahren Triumph der durch Gauss, Riemann, Christoffel, Ricci, und Levi-Civita begründeten Methode des allgemeinen Differentialkalküls" (Einstein 1915a, p. 779).

160 See eqs. (1)–(5) in sec. 2; G_{ij}, R_{ij}, and S_{ij} are defined in the equations following eq. (82).

161 "Nach dem bisher Gesagten liegt es nahe, die Feldgleichungen in der Form $R_{\mu\nu} = -\kappa T_{\mu\nu}$ anzusetzen, da wir bereits wissen, daß diese Gleichungen gegenüber beliebigen Transformationen von der Determinante 1 kovariant sind" (Einstein 1915a, p. 783).

can be found in (Norton 2000).[163] At the beginning of sec. 5, "Reversal at the Eleventh Hour," Norton gives the following summary of the developments of fall 1915:

> *… aware of the flaws in his Entwurf theory, Einstein decided he could only find the correct theory through the expressions naturally suggested by the mathematics. He proceeded rapidly to the completion of the theory and the greatest triumph of his life […]. Einstein now saw the magic in mathematics*. (Norton 2000, 148)

He elaborates:

> In effect, [Einstein's] new tactic [in the fall of 1915] was to reverse his decision of 1913. When the physical requirements appeared to contradict the formal mathematical requirements, he had then chosen in favour of the former. He now chose the latter and, writing down the mathematically natural equations, found himself rapidly propelled towards a theory that satisfied all the requirements and fulfilled his 'wildest dreams'[164] […] Einstein's reversal was his Moses that parted the waters and led him from bondage into the promised land of his general theory of relativity—and not a moment too soon. Had he delayed, the promised land might well have been Hilbert's.[165] (Einstein 1933b, 289) recalled how he 'ruefully returned to the Riemann curvature'. He now saw just how directly the mathematical route had delivered the correct equations in 1913 and, by contrast, how treacherous was his passage if he used physical requirements as his principal compass (Norton 2000, 151–152)

There is an amusing pair of quotations from letters to Besso that can be used as evidence for 'Einstein's reversal' (cf. Norton 2000, 152). In March 1914, reporting results that seemed to solidify the *Entwurf* theory (see sec. 3), Einstein told Besso:

> The general theory of invariants only proved to be an obstacle. The direct route proved to be the only feasible one. The only thing that is incomprehensible is that I had to feel my way around for so long before I found *the obvious* [our emphasis].[166]

In December 1915, Einstein used the same term to tell Besso the exact opposite: "the obvious" ("das Nächstliegende") now refers to the mathematically rather than the physically obvious:

> This time *the obvious* was correct; however Grossmann and I believed that the conservation laws would not be satisfied and that Newton's law would not come out in first approximation [our emphasis].[167]

---

162 The same pattern can be found in the review article. Einstein (1916a, 803–804) introduces the field equations by connecting them to the Riemann tensor and then proceeds to discuss them using the variational formalism. In the letter to Ehrenfest, however, that (as we argued in sec. 8) formed the blueprint for the discussion of the gravitational field equations in the review article, Einstein writes down the Lagrangian right a way and does not say a word about the connection between the field equations and the Riemann tensor. Einstein to Paul Ehrenfest, 24 January 1916 or later (CPAE 8, Doc. 185). Of course, he could simply have omitted that part because that was not what Ehrenfest had trouble with.

163 In a recent paper, Jeroen van Dongen (2004, sec. 2) fully endorses Norton's account. He is more careful in his dissertation (Van Dongen 2002).

164 See Einstein to Michele Besso, 10 December 1915 (CPAE 8, Doc. 162).

165 See (Corry et al. 1997) for conclusive evidence of Einstein's priority (cf. footnote 65 above)

166 "Die allgemeine Invariantentheorie wirkte nur als Hemmnis. Der direkte Weg erwies sich als der einzig gangbare. Unbegreiflich ist nur, dass ich so lange tasten musste, bevor ich das Nächstliegende fand." Einstein to Michele Besso, ca. 10 March 1914 (CPAE 5, Doc. 514).

These twin quotations seem to provide strong, if anecdotal, evidence for Einstein changing horses in the fall of 1915.

Even in late 1915, however, as the last quotation illustrates, Einstein mentioned physical as well as mathematical considerations. The same thing he told Besso he told Sommerfeld and Hilbert too:

> It is easy, of course, to write down these generally-covariant field equations but difficult to see that they are a generalization of the Poisson equation and not easy to see that they satisfy the conservation laws.[168]
>
> The difficulty did not lie in finding generally-covariant equations for the $g_{\mu\nu}$; this is easily done with the help of the Riemannian tensor. Rather it was difficult to recognize that these equations formed a generalization of Newton's laws and indeed a simple and natural generalization.[169]

To be sure, these references to physical considerations fit with the alleged 'reversal' from physics to mathematics. First, we have to keep in mind that Einstein had ulterior motives in emphasizing that the mathematics was easy and that getting the physics straight was the hard part. He wanted to downplay the importance of Hilbert's work. Einstein felt that Hilbert had only worked on the theory's mathematical formalism and had not wrestled with the physical interpretation of the formalism the way he had. Second, Einstein's comments on the difficulty of the physical interpretation of the field equations still suggest that the decisive breakthrough occurred in the mathematics and that the physics then fell into place. We have argued that it was just the other way around.

Why did Einstein nonetheless choose to present the new theory as a product of his mathematical strategy? Undoubtedly, part of the answer is that the key mathematical consideration pointing to the new field equations—the November tensor's pedigree in the Riemann tensor—is much simpler than the physical reasoning that had led Einstein to these equations in the first place. But even had the physical considerations been less forbidding, they would in all likelihood not have made for an effective and

167 "Diesmal ist das Nächstliegende das Richtige gewesen; Grossmann und ich glaubten, dass die Erhaltungssätze nicht erfüllt seien, und das Newton'sche Gesetz in erster Näherung nicht herauskomme." Einstein to Michele Besso, 10 December 1915 (CPAE 8, Doc. 162).

168 "Es ist natürlich leicht, diese allgemein kovariante Gleichungen hinzusetzen, schwer aber, einzusehen, dass sie Verallgemeinerungen von Poissons Gleichungen sind, und nicht leicht, einzusehen, dass sie den Erhaltungssätzen Genüge leisten." Einstein to Arnold Sommerfeld, 28 November 1915 (CPAE 8, Doc. 153; see footnote 31 for a discussion of the context in which this letter was written). It is very telling that recovering the Poisson equation is presented as a problem that is harder than proving energy-momentum conservation. This is indeed the case—it requires overcoming the notion of coordinate restrictions and the prejudice about the form of the static metric—but one would never have guessed this from the November 1915 paper where the Poisson equation is recovered simply by applying the Hertz condition.

169 "Die Schwierigkeit bestand nicht darin allgemein kovariante Gleichungen für die $g_{\mu\nu}$ zu finden; denn dies gelingt leicht mit Hilfe des Riemann'schen Tensors. Sondern schwer war es, zu erkennen, dass diese Gleichungen eine Verallgemeinerung, und zwar eine einfache und natürliche Verallgemeinerung des Newton'schen Gesetzes bilden." Einstein to David Hilbert, 18 November 1915 (CPAE 8, Doc. 148).

convincing argument for the new field equations. After all, Einstein had essentially drawn on these same considerations a year earlier for his fallacious argument for the uniqueness of the *Entwurf* field equations. That debacle was bound to come back and haunt a new argument along similar lines.

The math envy Einstein developed in the course of his work on general relativity may also have been a factor. As he confessed to Sommerfeld early on in his collaboration with Grossmann: "One thing for sure though is that I have never before in my life exerted myself even remotely as much and that I have been infused with great respect for mathematics, the subtler parts of which I until now, in my innocence, considered pure luxury. Compared to this problem, the original theory of relativity is child's play."[170] In 1917 he told Levi-Civita that "[i]t must be a pleasure to ride through these fields on the steed of real mathematics, while the likes of us must trudge through on foot."[171]

We suspect, however, that Einstein's main reason for going with the mathematical argument was simply that he felt that this was by far the most persuasive argument in favor of the new field equations. Recall Einstein's satisfaction in October 1914 with the physical and mathematical lines of reasoning apparently converging on the *Entwurf* field equations, thereby finally rendering their covariance properties tractable (see our discussion at the end of sec. 4). If anything, the convergence of mathematical and physical lines of reasoning in late 1915 was more striking than it had been the year before. The concomitant clarification of the equations' covariance properties was accordingly more complete and more perspicuous. In the case of the *Entwurf* field equations, the clarification had taken the form of a complicated condition for non-autonomous transformations. In the case of the November tensor, the connection to the Riemann tensor immediately told Einstein that his new field equations were invariant under arbitrary unimodular transformations. Given how pleased Einstein had been with his much more modest result in 1914, this new result cannot have failed to impress him. As we saw at the end of sec. 4, Einstein got carried away by the earlier result, claiming that he had found definite field equations "in a completely formal manner, i.e., without direct use of our physical knowledge about gravity" (Einstein 1914c, 1076; cf. footnotes 78 and 79). The same happened in November 1915. In arguing for his new field equations, Einstein emphasized the covariance considerations to the exclusion of (at least) equally important considerations concerning energy-momentum conservation and the relation to Newtonian gravitational theory.

---

170 "Aber das eine ist sicher, dass ich mich im Leben noch nicht annäherend so geplag[t] habe, und dass ich grosse Hochachtung für die Mathematik eingeflösst bekommen habe, die ich bis jetzt in ihren subtileren Teilen in meiner Einfalt für puren Luxus ansah! Gegen dies Problem ist die ursprüngliche Relativitätstheorie eine Kinderei." Einstein to Arnold Sommerfeld, 29 October 1912 (CPAE 5, Doc. 421).

171 "Es muss hübsch sein, auf dem Gaul der eigentlichen Mathematik durch diese Gefilde zu reiten, während unsereiner sich zu Fuss durchhelfen muss." Einstein to Tullio Levi-Civita, 2 August 1917 (CPAE 8, Doc. 368).

These physical considerations rapidly faded from memory. The way Einstein came to remember it, the general theory of relativity—the crowning achievement of his scientific career—was the result of a purely mathematical approach to physics. Sometimes he knew better than that. In 1918, for instance, he wrote to Besso:

> Re-reading your last letter I find something that almost makes me angry: that speculation has proved itself to be superior to empiricism. You are thinking here about the development of relativity theory. I find that this development teaches something else, which is almost the opposite, namely that a theory, to deserve our trust, must be built upon generalizable *facts* [… In the case of g]eneral relativity: *equality of inertial and gravitational mass* […] Never has a truly useful and profound theory really been found purely speculatively.[172]

Statements like this, however, are the exception. Much more typical is what he wrote to Cornelius Lanczos in 1938: "the problem of gravitation has made me into a believing rationalist, i.e., one who looks for the only reliable source of truth in mathematical simplicity."[173] This distorted memory of how he had found general relativity served an important purpose in his subsequent career. Whenever the need arose to justify the speculative mathematical approach that never got him anywhere in his work on unified field theory, Einstein reminded his audience that he could boast of at least one impressive successful application of his preferred methodology.

The emblematic text documenting the later Einstein's extreme rationalist stance on scientific methodology is his Herbert Spencer lecture, held in Oxford on June 10, 1933.[174] This is where Einstein famously enthused that "[o]ur experience hitherto justifies us in believing that nature is the realization of the simplest conceivable mathematical ideas" (Einstein 1933a, 274[175]). Einstein routinely claimed that this was the lesson he had drawn from the way in which he had found general relativity. The passage in the letter to Lanczos that we just quoted is a good example. A few more exam-

---

172 "In Deinem letzten Brief finde ich beim nochmaligen Lesen etwas, das mich geradezu erbost: die Spekulation habe sich als der Empirie überlegen gezeigt. Du denkst dabei an die Entwicklung der Relativitätstheorie. Aber ich finde, dass diese Entwicklung etwas anderes lehrt, das fast das Gegenteil davon ist, nämlich, dass eine Theorie, um Vertrauen zu verdienen, auf verallgemeinerungsfähige *Thatsachen* aufgebaut sein muss […] Algemeine Relativität: *Gleichheit der trägen und schweren Masse* […] Niemals ist eine wirklich brauchbare und tiefgehende Theorie wirklich rein spekulativ gefunden worden." Einstein to Michele Besso, 28 August 1918 (CPAE 8, Doc. 607). Quoted and discussed in (Holton 1968, 246–247).

173 "… hat das Gravitationsproblem mich zu einem gläubigen Rationalisten gemacht, d.h. zu einem, der die einzige zuverlässige Quelle der Wahrheit in der mathematischen Einfachheit sucht." Einstein to Cornel Lanczos, 24 January 1938 (EA 15 267). Quoted and discussed in (Holton 1968, 259) and (Van Dongen 2002, 48).

174 For discussion, see, e.g., (Holton 1968, 251–252), (Norton 2000), and (Van Dongen 2002).

175 The German manuscript (EA 1 114) has: "Nach unserer bisherigen Erfahrung sind wir nämlich zu dem Vertrauen darin berechtigt, dass die Natur die Realisierung des mathematisch denkbar Einfachsten ist."

ples must suffice here.[176] In a letter to Louis de Broglie the year before he died Einstein wrote that he arrived at the position expounded in his Spencer lecture

> through the experiences with the gravitational theory. The gravitational [field] equations could only be found on the basis of a purely formal principle (general covariance), i.e., on the basis of trust in the largest imaginable simplicity of the laws of nature.[177]

In his autobiographical notes of 1949, he similarly wrote that

> I have learned something else from the theory of gravitation: no collection of empirical facts however comprehensive can ever lead to the formulation of such complicated equations […] Equations of such complexity as are the equations of the gravitational field can be found only through the discovery of a logically simple mathematical condition that determines the equations completely or [at least] almost completely. Once one has those sufficiently strong formal conditions, one requires only little knowledge of facts for the setting up of a theory; in the case of the equations of gravitation it is the four-dimensionality and the symmetric tensor as expression for the structure of space which together with the invariance concerning the continuous transformation group, determine the equations almost completely.[178]

Discussing this passage, Jeroen van Dongen (2002, 30) notes that this is hardly a historically balanced account and that it reads more "like a unified field theory manifesto." As we have shown in this paper, the Einstein field equations were found not, as the later Einstein would have it, by extracting the mathematically simplest equations from the Riemann tensor, but by pursuing the analogy with Maxwell's equations for the electromagnetic field, making sure that they be compatible with Newtonian gravitational theory and energy-momentum conservation. Considerations of mathematical elegance played a role at various junctures but were always subordinate to physical considerations.

---

176 For more examples, see, e.g., (Holton 1968, 259–260), (Norton 2000), and (Van Dongen 2002).

177 "durch die Erfahrungen bei der Gravitationstheorie. Die Gravitations-gleichungen waren nur auffindbar auf Grund eines rein formalen Prinzipes (allgemeine Kovarianz), d.h. auf Grund des Vertrauens auf die denkbar grösste logische Einfachheit der Naturgesetze" Einstein to Louis de Broglie, 8 February 1954. This letter is quoted and discussed in (Van Dongen 2002, 8)

178 "Noch etwas anderes habe ich aus der Gravitationstheorie gelernt: Eine noch so umfangreiche Sammlung empirischer Fakten kann nicht zur Aufstellung so verwickelter Gleichungen führen […] Gleichungen von solcher Kompliziertheit wie die Gleichungen des Gravitationsfeldes können nur dadurch gefunden werden, dass eine logisch einfache mathematische Bedingung gefunden wird, welche die Gleichungen völlig oder nahezu determiniert. Hat man aber jene hinreichend starken formalen Bedingungen, so braucht man nur wenig Tatsachen-Wissen für die Aufstellung der Theorie; bei den Gravitationsgleichungen ist es die Vierdimensionalität und der symmetrische Tensor als Ausdruck für die Raumstruktur, welche zusammen mit der Invarianz der kontinuierlichen Transformationsgruppe die Gleichungen praktisch vollkommen determinieren" (Einstein 1949, 88–89).

APPENDIX: THE TRANSITION FROM THE *ENTWURF* FIELD EQUATIONS TO THE EINSTEIN FIELD EQUATIONS SANITIZED

Drawing on calculations scattered throughout this paper and with the benefit of hindsight, we present a sanitized version of the path that took Einstein from the *Entwurf* field equations to the Einstein field equations. We start from the vacuum field equations in unimodular coordinates. In the form in which they were originally presented (Einstein and Grossmann 1913, 16-17, eqs. 15 and 18) the *Entwurf* equations look nothing like the Einstein field equations. In unimodular coordinates they can be written in a form that clearly brings out the relation to their successor.

Comparing Vacuum Field Equations. Both in the *Entwurf* theory and in general relativity, the vacuum field equations in unimodular coordinates can be derived from the action principle $\delta\int\sqrt{-g}L d\tau$. In both cases the Lagrangian is given by:

$$L = g^{\mu\nu}\Gamma^{\alpha}_{\beta\mu}\Gamma^{\beta}_{\alpha\nu} \tag{A.1}$$

(see eq. (63) with $L \equiv \mathfrak{L}$ for general relativity and eq. (35) with $L \equiv -H$ for the *Entwurf* theory). In general relativity the gravitational field is defined as (see eq. (53))

$$\Gamma^{\alpha}_{\beta\mu} \equiv -\begin{Bmatrix}\alpha\\ \beta\mu\end{Bmatrix} = -\frac{1}{2}g^{\alpha\rho}(g_{\rho\beta,\mu} + g_{\rho\mu,\beta} - g_{\beta\mu,\rho}); \tag{A.2}$$

in the *Entwurf* theory as (see eq. (52) with a minus sign)

$$\tilde{\Gamma}^{\alpha}_{\beta\mu} \equiv -\frac{1}{2}g^{\alpha\rho}g_{\rho\beta,\mu}. \tag{A.3}$$

To distinguish between corresponding quantities in the two theories, we shall write the *Entwurf* quantities with a tilde, as in eq. (A.3). Note that $\tilde{\Gamma}^{\alpha}_{\beta\mu}$ in eq. (A.3) is nothing but a truncated version of $\Gamma^{\alpha}_{\beta\mu}$ in eq. (A.2). Also note that the Lagrangian in eq. (A.1) is modelled on the Lagrangian for the free Maxwell field, $-\frac{1}{4}F^{\mu\nu}F_{\mu\nu}$.

The structural identity of the Lagrangians in the two theories does not carry over to the Euler-Lagrange equations. This is because of two complications. (1) The operations 'setting $\sqrt{-g} = 1$' and 'doing the variations' do not commute (see footnote 83). In the *Entwurf* theory we do the variations first. In general relativity we set $\sqrt{-g} = 1$ first. (2) The quantities $\Gamma^{\alpha}_{\beta\mu}$ are symmetric in their lower indices, whereas their counterparts, $\tilde{\Gamma}^{\alpha}_{\beta\mu}$, in the *Entwurf* theory are not.

In unimodular coordinates, the vacuum Einstein field equations can be written as (see footnote 89)[179]

$$(g^{\nu\lambda}\Gamma^{\alpha}_{\mu\nu})_{,\alpha} - g^{\nu\rho}\Gamma^{\lambda}_{\alpha\rho}\Gamma^{\alpha}_{\mu\nu} = 0, \tag{A.4}$$

and the vacuum *Entwurf* field equations as (see eqs. (47) and (49)):

$$(g^{\alpha\beta}\tilde{\Gamma}^{\lambda}_{\mu\beta})_{,\alpha} - g^{\lambda\rho}\tilde{\Gamma}^{\alpha}_{\tau\mu}\tilde{\Gamma}^{\tau}_{\alpha\rho} + \frac{1}{2}\delta^{\lambda}_{\mu}g^{\rho\tau}\tilde{\Gamma}^{\alpha}_{\beta\rho}\tilde{\Gamma}^{\beta}_{\alpha\tau} = 0. \tag{A.5}$$

We can get these equations to resemble each other even more closely by defining

$$\Gamma_{\alpha}^{\mu\nu} \equiv g^{\mu\beta}\Gamma_{\alpha\beta}^{\nu} \qquad \text{(A.6)}$$

in general relativity and the corresponding quantities

$$\tilde{\Gamma}_{\alpha}^{\mu\nu} \equiv g^{\nu\beta}\tilde{\Gamma}_{\alpha\beta}^{\mu} \qquad \text{(A.7)}$$

in the *Entwurf* theory. Note that the order of the indices μ and ν is different in eqs. (A.6) and (A.7).[180] Inserting these quantities into eqs. (A.4) and (A.5), we find:

$$(\Gamma_{\mu}^{\lambda\alpha})_{,\alpha} - \Gamma_{\alpha}^{\nu\lambda}\Gamma_{\nu\mu}^{\alpha} = 0, \qquad \text{(A.8)}$$

$$(\tilde{\Gamma}_{\mu}^{\lambda\alpha})_{,\alpha} - \tilde{\Gamma}_{\alpha}^{\tau\lambda}\tilde{\Gamma}_{\tau\mu}^{\alpha} + \frac{1}{2}\delta_{\mu}^{\lambda}\tilde{\Gamma}_{\beta}^{\alpha\tau}\tilde{\Gamma}_{\alpha\tau}^{\beta} = 0. \qquad \text{(A.9)}$$

The first two terms in these two equation have the exact same form.

Expressed in unimodular coordinates and in terms of $\{\Gamma_{\alpha\beta}^{\mu}, \Gamma_{\mu}^{\alpha\beta}\}$ and $\{\tilde{\Gamma}_{\alpha\beta}^{\mu}, \tilde{\Gamma}_{\mu}^{\alpha\beta}\}$, respectively, the gravitational energy-momentum pseudo-tensors of the two theories also take on the exact same form. With the help of eq. (A.6) the pseudo-tensor of general relativity in unimodular coordinates (see eq. (73)) can be written as

$$\kappa t_{\sigma}^{\lambda} = \frac{1}{2}\delta_{\sigma}^{\lambda}\Gamma_{\mu\beta}^{\alpha}\Gamma_{\alpha}^{\mu\beta} - \Gamma_{\mu\sigma}^{\alpha}\Gamma_{\alpha}^{\mu\lambda}. \qquad \text{(A.10)}$$

With the help of eq. (A.7) the pseudo-tensor of the *Entwurf* theory in unimodular coordinates, with an overall minus sign because of the switch from H to $L = -H$ (see minus eq. (49) for $\sqrt{-g} = 1$) can be written as

$$\kappa\tilde{t}_{\mu}^{\lambda} = \frac{1}{2}\delta_{\mu}^{\lambda}\tilde{\Gamma}_{\beta\rho}^{\alpha}\tilde{\Gamma}_{\alpha}^{\beta\rho} - \tilde{\Gamma}_{\tau\mu}^{\alpha}\tilde{\Gamma}_{\alpha}^{\tau\lambda}. \qquad \text{(A.11)}$$

Eqs. (A.10) and (A.11) have the exact same form. The first term in both expressions contains the trace of the pseudo-tensor:

$$\kappa t = \Gamma_{\mu\beta}^{\alpha}\Gamma_{\alpha}^{\mu\beta}, \qquad \kappa\tilde{t} = \tilde{\Gamma}_{\beta\rho}^{\alpha}\tilde{\Gamma}_{\alpha}^{\beta\rho}. \qquad \text{(A.12)}$$

Using the expressions for the gravitational energy-momentum pseudo-tensors and their trace, we can write the field equations (A.8) and (A.9) in a form that brings out

---

179 When the variation $\delta\int\sqrt{-g}Ld\tau$ for L in eq. (A.1) is done *before* setting $\sqrt{-g} = 1$, one finds

$$\Gamma_{\mu\nu,\alpha}^{\alpha} + \Gamma_{\beta\mu}^{\alpha}\Gamma_{\alpha\nu}^{\beta} + \frac{1}{2}g_{\mu\nu}g^{\rho\sigma}\Gamma_{\beta\rho}^{\alpha}\Gamma_{\alpha\sigma}^{\beta} = 0$$

(see eq. (61) with $\mathfrak{T}_{\mu\nu} = 0$). Contracting this equation with $g^{\nu\lambda}$, one finds

$$(g^{\nu\lambda}\Gamma_{\mu\nu}^{\alpha})_{,\alpha} - g^{\nu\rho}\Gamma_{\alpha\rho}^{\lambda}\Gamma_{\mu\nu}^{\alpha} + \frac{1}{2}\delta_{\nu}^{\lambda}g^{\rho\sigma}\Gamma_{\beta\rho}^{\alpha}\Gamma_{\alpha\sigma}^{\beta} = 0$$

Not surprisingly, this last equation resembles eq. (A.5) in the *Entwurf* theory more closely than eq. (A.4), obtained when the variation is done *after* setting $\sqrt{-g} = 1$.

180 Expressed in terms of the new quantities $\Gamma_{\alpha}^{\mu\nu}$ and $\tilde{\Gamma}_{\alpha}^{\mu\nu}$, the Lagrangians for the two theories retain their structural identity: $L = g^{\mu\nu}\Gamma_{\beta\mu}^{\alpha}\Gamma_{\alpha\nu}^{\beta} = \Gamma_{\mu\beta}^{\alpha}\Gamma_{\alpha}^{\mu\beta}$ and $\tilde{L} = g^{\mu\nu}\tilde{\Gamma}_{\beta\mu}^{\alpha}\tilde{\Gamma}_{\alpha\nu}^{\beta} = \tilde{\Gamma}_{\beta\mu}^{\alpha}\tilde{\Gamma}_{\alpha}^{\beta\mu}$.

the physical interpretation of the various terms more clearly. Using eq. (A.12), we can rewrite eq. (A.10) as:

$$-\Gamma^{\nu\lambda}_{\alpha}\Gamma^{\alpha}_{\nu\mu} = \kappa t^{\lambda}_{\mu} - \frac{1}{2}\delta^{\lambda}_{\mu}\kappa t.$$

Substituting this equation into the field equations (A.8), we find[181]

$$(\Gamma^{\lambda\alpha}_{\mu})_{,\alpha} + \kappa t^{\lambda}_{\mu} - \frac{1}{2}\delta^{\lambda}_{\mu}\kappa t = 0. \tag{A.13}$$

Substituting eq. (A.11) into the field equations (A.9), we find

$$(\tilde{\Gamma}^{\lambda\alpha}_{\mu})_{,\alpha} + \kappa \tilde{t}^{\lambda}_{\mu} = 0. \tag{A.14}$$

The crucial difference between these last two equations is the trace term on the left-hand side of eq. (A.13).

From the Entwurf Field Equations to the Einstein Field Equations. Eqs. (A.13)–(A.14) suggest a short-cut for getting from the *Entwurf* field equations in unimodular coordinates to the Einstein field equations, first in unimodular coordinates and then in their generally-covariant form. Comparison of eq. (A.14) to eq. (A.13) shows that changing the definition of the gravitational field from $\tilde{\Gamma}^{\alpha}_{\beta\mu}$ in eq. (A.3) to $\Gamma^{\alpha}_{\beta\mu}$ in eq. (A.2) changes the way in which the gravitational energy-momentum pseudo-tensor occurs in the gravitational part of the field equations in unimodular coordinates. Since the energy-momentum of matter should enter the field equations in the same way as the energy-momentum of the gravitational field itself, this also affects the matter part of the field equations. In the presence of matter described by an energy-momentum tensor $T_{\mu\nu}$, the vacuum equations (A.14)—based on definition (A.3) of the gravitational field, Einstein's "fateful prejudice"—should be generalized to

$$(\tilde{\Gamma}^{\lambda\alpha}_{\mu})_{,\alpha} = -\kappa(\tilde{t}^{\lambda}_{\mu} + T^{\lambda}_{\mu}). \tag{A.15}$$

These are the *Entwurf* field equations in unimodular coordinates. By the same token, eq. (A.13)—based on definition (A.2) of the gravitational field, Einstein's "key to the solution"—should be generalized to:[182]

181 If eq. (A.4) obtained by doing the variations *after* setting $\sqrt{-g} = 1$ are replaced by the equations

$$(g^{\nu\lambda}\Gamma^{\alpha}_{\mu\nu})_{,\alpha} - g^{\nu\rho}\Gamma^{\lambda}_{\alpha\rho}\Gamma^{\alpha}_{\mu\nu} + \frac{1}{2}\delta^{\lambda}_{\nu}g^{\rho\sigma}\Gamma^{\alpha}_{\beta\rho}\Gamma^{\beta}_{\alpha\sigma} = 0,$$

obtained by doing the variations *before* setting $\sqrt{-g} = 1$ (see footnote 179), then eq. (A.13) gets replaced by $(\Gamma^{\lambda\alpha}_{\mu})_{,\alpha} + \kappa t^{\lambda}_{\mu} = 0$ which has the exact same form as eq. (A.14) in the *Entwurf* theory.

182 In his first November paper, Einstein (1915a) chose field equations in the presence of matter that set the left-hand side of eq. (A.13) equal to $-\kappa T^{\lambda}_{\mu}$. In the fourth November paper, Einstein (1915d) replaced the right-hand side by $-\kappa(T^{\lambda}_{\mu} - (1/2)\delta^{\lambda}_{\mu}T)$.

$$(\Gamma^{\lambda\alpha}_{\mu})_{,\alpha} = -\kappa\Big([t^{\lambda}_{\mu} + T^{\lambda}_{\mu}] - \frac{1}{2}\delta^{\lambda}_{\mu}[t + T]\Big). \tag{A.16}$$

These are the proper field equations for the successor theory to the *Entwurf* theory.

Eq. (A.15) guarantees energy-momentum conservation, $(\tilde{t}^{\lambda}_{\mu} + T^{\lambda}_{\mu})_{,\lambda} = 0$, in the *Entwurf* theory, if—in addition to $\sqrt{-g} = 1$ —the condition

$$\tilde{B}_{\mu} \equiv (\tilde{\Gamma}^{\lambda\alpha}_{\mu})_{,\alpha\lambda} = 0 \tag{A.17}$$

holds (cf. eq. (50)).

Eq. (A.16) guarantees energy-momentum conservation, $(t^{\lambda}_{\mu} + T^{\lambda}_{\mu})_{,\lambda} = 0$, in the new theory if—in addition to $\sqrt{-g} = 1$ —the condition

$$\Big[(\Gamma^{\lambda\alpha}_{\mu})_{,\alpha} - \frac{1}{2}\delta^{\lambda}_{\mu}\kappa(t + T)\Big]_{,\lambda} = 0 \tag{A.18}$$

holds. Contracting eq. (A.16), one finds that

$$(\Gamma^{\rho\alpha}_{\rho})_{,\alpha} = \kappa(t + T)$$

(see eq. (104)). Using this equation to eliminate T from eq. (A.18), one arrives at the condition $B_{\mu} = 0$ in the new theory

$$B_{\mu} \equiv \Big[\Gamma^{\lambda\alpha}_{\mu} - \frac{1}{2}\delta^{\lambda}_{\mu}\Gamma^{\rho\alpha}_{\rho}\Big]_{,\alpha\lambda} = 0 \tag{A.19}$$

(see eq. (105)). Eq. (A.19), it turns out, is an identity (see eq. (105) and footnotes 123–124). Eq. (A.17) in the *Entwurf* theory imposes a coordinate restriction over and above unimodularity. Eq. (A.19), its counterpart in the new theory, imposes no additional restriction.

The gravitational part of the field equations (A.16), i.e., the left-hand side of eq. (A.13), is nothing but an alternative expression for the November tensor $\Gamma^{\alpha}_{\mu\nu,\alpha} + \Gamma^{\alpha}_{\beta\mu}\Gamma^{\beta}_{\alpha\nu}$ (see eq. (67)), which itself is nothing but the Ricci tensor in unimodular coordinates (see eqs. (82)–(83)). It follows that the field equations (A.16) are (the mixed form of) the generally-covariant Einstein field equations,

$$R_{\mu\nu} = -\kappa\Big(T_{\mu\nu} - \frac{1}{2}g_{\mu\nu}T\Big), \tag{A.20}$$

in unimodular coordinates (where $R_{\mu\nu}$ is the Ricci tensor). Eq. (A.19) gives the contracted Bianchi identities in unimodular coordinates.

The generally-covariant form of the Einstein field equations can be derived from the action principle, $\delta\int\sqrt{-g}R\,d\tau = 0$, where R is the Riemann curvature scalar. All terms involving second-order derivatives of the metric can be eliminated from the action through partial integration. One then arrives at an action of the form

$$\int \mathfrak{G}^{*} d\tau, \tag{A.21}$$

where

$$\mathfrak{G}^* \equiv \sqrt{-g}\,g^{\mu\nu}\left[\begin{Bmatrix}\beta\\ \mu\alpha\end{Bmatrix}\begin{Bmatrix}\alpha\\ \nu\beta\end{Bmatrix} - \begin{Bmatrix}\alpha\\ \mu\nu\end{Bmatrix}\begin{Bmatrix}\beta\\ \alpha\beta\end{Bmatrix}\right]. \tag{A.22}$$

In unimodular coordinates $\mathfrak{G}^*$ reduces to

$$\mathfrak{G}^*(\sqrt{-g} = 1) = g^{\mu\nu}\begin{Bmatrix}\beta\\ \mu\alpha\end{Bmatrix}\begin{Bmatrix}\alpha\\ \nu\beta\end{Bmatrix}. \tag{A.23}$$

This is just the Lagrangian given in eq. (A.1) with the gravitational field defined as minus the Christoffel symbols (see eq. (A.2)).[183]

ACKNOWLEDGMENTS

This paper grew out of joint work with John Norton, Tilman Sauer, and John Stachel. All along, we have had John Norton's seminal paper, "How did Einstein find his field equations?" (Norton 1984), in the cross hairs. Our deepest gratitude is therefore to John Norton for providing us with such a worthy target. We are also indebted to John Stachel for alerting us to the non-commutativity of (partial) gauge fixing and (unconstrained) variation of the action and to Serge Rudaz for further discussion of this issue. We thank Katherine Brading, David Rowe, and Tilman Sauer for discussion of (the history of) Noether's theorems and Dan Kennefick for discussion of the gravitational energy-momentum pseudo-tensor. For additional comments and stimulating discussion, we thank Paul Brinkman, Peter Damerow, Jeroen van Dongen, Tony Duncan, Joska Illy, Martin J. Klein, A. J. Kox, Christoph Lehner, Matthias Schemmel, Robert Schulmann, Bernard Schutz, and Bill Unruh. Part of this paper was written at the *Max-Planck-Institut für Gravitationsphysik* (*Albert Einstein Institut*) in Golm outside Berlin; another part at the Outing Lodge near Stillwater, Minnesota. We thank Bernard Schutz (*Albert Einstein Institut*) and Lee Gohlike (Outing Lodge) for their hospitality and support. Swetlana Freifrau von dem Bottlenberg-Minkow could have hosted us at *Gut Schloß Golm* but chose not to. Finally, we want to thank our families—Fiorenza, Suzy, Leonardo, Eleonora, and Luc—for putting up with the vicissitudes of our transatlantic cooperation.

183 This once again illustrates that 'doing the variations' does not commute with 'choosing unimodular coordinates' (cf. footnotes 83 and 179). The Lagrangian (A.23)—i.e., setting $\sqrt{-g} = 1$ first—leads to Euler-Lagrange equations that set the Ricci tensor, $R_{\mu\nu}$, in unimodular coordinates equal to zero. The Lagrangian (A.22)—i.e., doing the variations first—leads to Euler-Lagrange equations that set the Einstein tensor, $R_{\mu\nu} - (1/2)g_{\mu\nu}R$, equal to zero.

REFERENCES

Anderson, James L., and David Finkelstein. 1971. "Cosmological Constant and Fundamental Length." *American Journal of Physics* 39: 901–904.

Brading, Katherine A. 2002. "Which Symmetry? Noether, Weyl, and the Conservation of Electric Charge." *Studies in History and Philosophy of Modern Physics* 33: 3-22.

Cattani, Carlo, and Michelangelo De Maria. 1989. "The 1915 Epistolary Controversy Between Einstein and Tullio Levi-Civita." In (Howard and Stachel 1989, 175–200).

———. 1993. "Conservation Laws and Gravitational Waves in General Relativity (1915–1918)." In (Earman et al., 1993, 63–87).

Corry, Leo, Jürgen Renn, and John Stachel. 1997. "Belated Decision in the Hilbert-Einstein Priority Dispute." *Science* 278: 1270–1273.

CPAE 4: Martin J. Klein, A. J. Kox, Jürgen Renn, and Robert Schulmann (eds.), *The Collected Papers of Albert Einstein*. Vol. 4. *The Swiss Years: Writings, 1912–1914*. Princeton: Princeton University Press, 1995.

CPAE 5: Martin J. Klein, A. J. Kox, and Robert Schulmann (eds.), *The Collected Papers of Albert Einstein*. Vol. 5. *The Swiss Years: Correspondence, 1902–1914*. Princeton: Princeton University Press, 1993.

CPAE 6: A. J. Kox, Martin J. Klein, and Robert Schulmann (eds.), *The Collected Papers of Albert Einstein*. Vol. 6. *The Berlin Years: Writings, 1914–1917*. Princeton: Princeton University Press, 1996.

CPAE 7: Michel Janssen, Robert Schulmann, József Illy, Christoph Lehner, and Diana Kormos Barkan (eds.), *The Collected Papers of Albert Einstein*. Vol. 7. *The Berlin Years: Writings, 1918–1921*. Princeton: Princeton University Press, 2002.

CPAE 8: Robert Schulmann, A. J. Kox, Michel Janssen, and József Illy (eds.), *The Collected Papers of Albert Einstein*. Vol. 8. *The Berlin Years: Correspondence, 1914–1918*. Princeton: Princeton University Press, 1998.

De Donder, Théophile. 1917. "Sur les équations différentielles du champ gravifique." *Koninklijke Akademie van Wetenschappen te Amsterdam. Wis- en Natuurkundige Afdeeling. Verslagen van de Gewone Vergaderingen* 26 (1917–18): 101–104. Reprinted in *Koninklijke Akademie van Wetenschappen te Amsterdam. Section of Sciences. Proceedings* 20 (1917–18): 97–100.

Earman, John. 2003. "The Cosmological Constant, the Fate of the Universe, Unimodular Gravity, and all that." *Studies in History and Philosophy of Modern Physics* 34: 559–577.

Earman, John, and Michel Janssen. 1993. "Einstein's Explanation of the Motion of Mercury's Perihelion." In (Earman et al., 1993, 129–172).

Earman, John, Michel Janssen, and John D. Norton (eds.). 1993. *The Attraction of Gravitation: New Studies in the History of General Relativity* (*Einstein Studies,* Vol. 5). Boston: Birkhäuser.

Einstein, Albert. 1912. "Zur Theorie des statischen Gravitationsfeldes." *Annalen der Physik* 38: 443–458, (CPAE 4, Doc. 4).

———. 1913. "Zum gegenwärtigen Stande des Gravitationsproblems." *Physikalische Zeitschrift* 14: 1249–1262, (CPAE 4, Doc. 17). (English translation in vol. 3 of this series.)

———. 1914a. "Physikalische Grundlagen einer Gravitationstheorie." *Naturforschende Gesellschaft in Zürich. Vierteljahrsschrift* 58: 284–290, (CPAE 4, Doc. 16).

———. 1914b. "Prinzipielles zur verallgemeinerten Relativitätstheorie und Gravitationstheorie." *Physikalische Zeitschrift* 15: 176–180, (CPAE 4, Doc. 25).

———. 1914c. "Die formale Grundlage der allgemeinen Relativitätstheorie." *Königlich Preußische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte*: 1030–1085, (CPAE 6, Doc. 9).

———. 1915a. "Zur allgemeinen Relativitätstheorie." *Königlich Preußische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte*: 778–786, (CPAE 6, Doc. 21).

———. 1915b. "Zur allgemeinen Relativitätstheorie. (Nachtrag)." *Königlich Preußische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte*: 799–801, (CPAE 6, Doc. 22).

———. 1915c. "Erklärung der Perihelbewegung des Merkur aus der allgemeinen Relativitätstheorie." *Königlich Preußische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte*: 831–839, (CPAE 6, Doc. 24).

———. 1915d. "Die Feldgleichungen der Gravitation." *Königlich Preußische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte*: 844–847, (CPAE 6, Doc. 25).

———. 1916a. "Die Grundlage der allgemeinen Relativitätstheorie." *Annalen der Physik* 49: 769–822, (CPAE 6, Doc. 30).

———. 1916b. "Näherungsweise Integration der Feldgleichungen der Gravitation." *Königlich Preußische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte*: 688–696, (CPAE 6, Doc. 32).

———. 1916c. "Hamiltonsches Prinzip und allgemeine Relativitätstheorie." *Königlich Preußische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte*: 1111–1116, (CPAE 6, Doc. 41).

———. 1917. "Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie." *Königlich Preußische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte*: 142–152, (CPAE 6, Doc. 43).
———. 1918a. "Über Gravitationswellen." *Königlich Preußische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte*: 154–167, (CPAE 7, Doc. 1).
———. 1918b. "Notiz zu E. Schrödingers Arbeit 'Die Energiekomponenten des Gravitationsfeldes.'" *Physikalische Zeitschrift* 19: 115–116, (CPAE 7, Doc. 2).
———. 1918c. "Prinzipielles zur allgemeinen Relativitätstheorie." *Annalen der Physik* 55: 241–244, (CPAE 7, Doc. 4).
———. 1918d. "Der Energiesatz der allgemeinen Relativitätstheorie." *Königlich Preußische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte*: 448–459, (CPAE 7, Doc. 9).
———. 1919. "Spielen Gravitationsfelder im Aufbau der materiellen Elementarteilchen eine wesentliche Rolle?" *Königlich Preußische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte*: 349–356, (CPAE 7, Doc. 17).
———. 1933a. *On the Method of Theoretical Physics*. (The Herbert Spencer lecture. Oxford, June 10, 1933.) Oxford: Clarendon Press. Reprinted in more colloquial translation in (Einstein 1954, 270–276). Page reference to this reprint. A German manuscript is available (EA 1 114) under the title: "Zur Methode der theoretischen Physik."
———. 1933b. *Origins of the General Theory of Relativity.* (George A. Gibson foundation lecture at the University of Glasgow, June 20, 1933.) Glasgow: Jackson. Reprinted in slightly different version (under the title "Notes on the Origin of the General Theory of Relativity") in (Einstein 1954, 285–290). Page reference to this reprint. A German manuscript is available (EA 78 668) under the title: "Einiges über die Entstehung der allgemeinen Relativitätstheorie."
———. 1949. "Autobiographical Notes." In P. A. Schilpp (ed.), *Albert Einstein: Philosopher-Scientist*. Evanston, Ill.: Library of Living Philosophers, 1–95.
———. 1954. *Ideas and Opinions*. New York: Bonanza.
Einstein, Albert, and Adriaan D. Fokker. 1914. "Die Nordströmsche Gravitationstheorie vom Standpunkt des absoluten Differentialkalküls." *Annalen der Physik* 44 (1914): 321–328, (CPAE 4, Doc. 28).
Einstein, Albert, and Marcel Grossmann. 1913. *Entwurf einer verallgemeinerten Relativitätstheorie und einer Theorie der Gravitation*. Leipzig: Teubner, (CPAE 4, Doc. 13).
———. 1914a. "Entwurf einer verallgemeinerten Relativitätstheorie und einer Theorie der Gravitation." *Zeitschrift für Mathematik und Physik* 62: 225–259. Reprint of *Einstein and Grossmann 1913* with additional "Comments" ("Bemerkungen," CPAE 4, Doc. 26).
———. 1914b. "Kovarianzeigenschaften der Feldgleichungen der auf die verallgemeinerte Relativitätstheorie gegründeten Gravitationstheorie." *Zeitschrift für Mathematik und Physik* 63: 215–225, (CPAE 6, Doc. 2).
Eisenstaedt, Jean, and A. J. Kox (eds.). 1992. *Studies in the History of General Relativity* (*Einstein Studies,* Vol. 3). Boston: Birkhäuser.
Finkelstein, David R., Andrei Galiautdinov, and James E. Baugh. 2001. "Unimodular Relativity and Cosmological Constant." *Journal of Mathematical Physics* 42: 340–346.
Gray, Jeremy (ed.). 1999. *The Symbolic Universe: Geometry and Physics, 1890–1930*. Oxford: Oxford University Press.
Hilbert, David. 1915. "Die Grundlagen der Physik. (Erste Mitteilung)." *Königliche Gesellschaft der Wissenschaften zu Göttingen. Mathematisch-physikalische Klasse. Nachrichten*: 395–407.
Holton, Gerald. 1968. "Mach, Einstein, and the Search for Reality." *Daedalus* 97: 636–673. Reprinted in Gerald Holton, *Thematic Origins of Scientific Thought: Kepler to Einstein*. Rev. Ed. Cambridge: Harvard University Press, 1988, 237–277. Page references are to this reprint.
Howard, Don, and John Stachel (eds.). 1989. *Einstein and the History of General Relativity* (*Einstein Studies*, Vol. 1). Boston: Birkhäuser.
Janssen, Michel. 1992. "H. A. Lorentz's Attempt to Give a Coordinate-free Formulation of the General Theory of Relativity." In (Eisenstaedt and Kox 1992, 344–363).
———. 1999. "Rotation as the Nemesis of Einstein's *Entwurf* Theory." In H. Goenner, J. Renn, J. Ritter, and T. Sauer (eds.), *The Expanding Worlds of General Relativity* (*Einstein Studies*, Vol. 7). Boston: Birkhäuser, 127–157.
———. 2005. "Of Pots and Holes: Einstein's Bumpy Road to General Relativity." *Annalen der Physik* 14: *Supplement* 58–85. Reprinted in J. Renn (ed.) *Einstein's Annalen Papers. The Complete Collection 1901–1922*. Weinheim: Wiley-VCH, 2005.
Kichenassamy, S. 1993. "Variational Derivations of Einstein's Equations." In (Earman et al. 1993, 185–205).
Klein, Felix. 1917. "Zu Hilberts erster Note über die Grundlagen der Physik." *Königliche Gesellschaft der Wissenschaften zu Göttingen. Mathematisch-physikalische Klasse. Nachrichten*: 469–482.

———. 1918a. “Über die Differentialgesetze für die Erhaltung von Impuls und Energie in der Einsteinschen Gravitationstheorie.” *Königliche Gesellschaft der Wissenschaften zu Göttingen. Mathematisch-physikalische Klasse. Nachrichten*: 171–189.

———. 1918b. “Über die Integralform der Erhaltungssätze und die Theorie der räumlich-geschlossenen Welt.” *Königliche Gesellschaft der Wissenschaften zu Göttingen. Mathematisch-physikalische Klasse. Nachrichten*: 394–423.

Klein, Martin J. 1970. *Paul Ehrenfest*. Vol. 1. *The Making of a Theoretical Physicist*. Amsterdam: North-Holland.

Kox, Anne J. 1988. “Hendrik Antoon Lorentz, the Ether, and the General Theory of Relativity.” *Archive for History of Exact Sciences* 38: 67–78. Reprinted in (Howard and Stachel 1989, 201–212).

———. 1992. “General Relativity in the Netherlands, 1915–1920.” In (Eisenstaedt and Kox 1992, 39–56).

Levi-Civita, Tullio. 1917. “Sulla espressione analitica spettante al tensore gravitazionale nella teoria di Einstein.” *Rendiconti della Reale Accademia dei Lincei. Atti* 26, 1st semester: 381–391.

Lorentz, Hendrik Antoon. 1915. “Het beginsel van Hamilton in Einstein’s theorie der zwaartekracht.” *Koninklijke Akademie van Wetenschappen te Amsterdam. Wis- en Natuurkundige Afdeeling. Verslagen van de Gewone Vergaderingen* 23 (1914–15): 1073–1089. Reprinted in translation as “On Hamilton’s Principle in Einstein’s Theory of Gravitation.” *Koninklijke Akademie van Wetenschappen te Amsterdam. Section of Sciences. Proceedings* 19 (1916–17): 751–765. Page reference is to the translation.

———. 1916a. “Over Einstein’s theorie der zwaartekracht. I.” *Koninklijke Akademie van Wetenschappen te Amsterdam. Wis- en Natuurkundige Afdeeling. Verslagen van de Gewone Vergaderingen* 24 (1915–16): 1389–1402. Reprinted in translation as “On Einstein’s Theory of Gravitation. I.” *Koninklijke Akademie van Wetenschappen te Amsterdam. Section of Sciences. Proceedings* 19 (1916–17): 1341–1354.

———. 1916b. “Over Einstein’s theorie der zwaartekracht. III.” *Koninklijke Akademie van Wetenschappen te Amsterdam. Wis- en Natuurkundige Afdeeling. Verslagen van de Gewone Vergaderingen* 25 (1916–17): 468–486. Reprinted in translation as “On Einstein’s Theory of Gravitation. III.” *Koninklijke Akademie van Wetenschappen te Amsterdam. Section of Sciences. Proceedings* 20 (1917–18): 2–19.

———. 1916c. “Over Einstein’s theorie der zwaartekracht. IV.” *Koninklijke Akademie van Wetenschappen te Amsterdam. Wis- en Natuurkundige Afdeeling. Verslagen van de Gewone Vergaderingen* 25 (1916–17): 1380–1396. Reprinted in translation as “On Einstein’s Theory of Gravitation. IV.” *Koninklijke Akademie van Wetenschappen te Amsterdam. Section of Sciences. Proceedings* 20 (1917–18): 20–34.

Noether, Emmy. 1918. “Invariante Variationsprobleme.” *Königliche Gesellschaft der Wissenschaften zu Göttingen. Mathematisch-physikalische Klasse. Nachrichten*: 235-257.

Norton, John D. 1984. “How Einstein Found his Field Equations, 1912–1915.” *Historical Studies in the Physical Sciences* 14: 253–316. Page references to reprint in (Howard and Stachel 1989, 101–159).

———. 1999. “Geometries in Collision: Einstein, Klein, and Riemann.” In (Gray 1999, 128–144).

———. 2000. “‘Nature is the Realisation of the Simplest Conceivable Mathematical Ideas’: Einstein and the Canon of Mathematical Simplicity.” *Studies in History and Philosophy of Modern Physics* 31: 135–170.

Rowe, David. 1999. “The Göttingen Response to General Relativity and Emmy Noether’s Theorems.” In (Gray 1999, 189–234).

Sauer, Tilman. 1999. “The Relativity of Discovery: Hilbert’s First Note on the Foundations of Physics.” *Archive for History of Exact Sciences* 53: 529–575.

Schrödinger, Erwin. 1918. “Die Energiekomponenten des Gravitationsfeldes.” *Physikalische Zeitschrift* 19: 4–7.

Stachel, John. 1989. “Einstein’s Search for General Covariance, 1912–1915.” (Based on a talk given in Jena in 1980.) In (Howard and Stachel 1989, 62–100).

Van Dongen, Jeroen. 2002. *Einstein’s Unification: General Relativity and the Quest for Mathematical Naturalness*. PhD. Thesis. University of Amsterdam.

———. 2004. “Einstein’s Methodology, Semivectors, and the Unification of Electrons and Protons.” *Archive for History of Exact Sciences* 58: 219–254.

Wald, Robert M. 1984. *General Relativity.* Chicago: University of Chicago Press, 1984.

Weyl, Hermann. 1918. *Raum–Zeit–Materie. Vorlesungen über allgemeine Relativitätstheorie*. Berlin: Springer.

INDEX

Explanatory entries are marked in boldface

The Genesis of General Relativity

BOSTON STUDIES IN THE PHILOSOPHY OF SCIENCE

VOLUME 250

The Genesis of General Relativity

Edited by Jürgen Renn

Volume 3

GRAVITATION IN THE TWILIGHT OF CLASSICAL PHYSICS: BETWEEN MECHANICS, FIELD THEORY, AND ASTRONOMY

Editors

Jürgen Renn and Matthias Schemmel
Max Planck Institute for the History of Science, Germany

Associate Editors

Christopher Smeenk
UCLA, U.S.A.

Christopher Martin
Indiana University, U.S.A.

Assistant Editor

Lindy Divarci
Max Planck Institute for the History of Science, Germany

Springer

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN-10 1-4020-3999-9 (HB)
ISBN-13 978-1-4020-3999-7 (HB)
ISBN-10 1-4020-4000-8 (e-book)
ISBN-13 978-1-4020-4000-9 (e-book)

As a complete set for the 4 volumes

Published by Springer,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

www.springer.com

Printed on acid-free paper

TABLE OF CONTENTS

Volume 3

From Heretical Mechanics to a New Theory of Relativity

Volume 4

(parallel volume)

From an Electromagnetic Theory of Matter
to a New Theory of Gravitation

Including Gravitation in a Unified Theory of Physics

JÜRGEN RENN AND MATTHIAS SCHEMMEL

GRAVITATION IN THE TWILIGHT OF CLASSICAL PHYSICS: AN INTRODUCTION

AIM AND STRUCTURE OF THIS BOOK

More than is the case for any other theory of modern physics, general relativity is usually seen as the work of one man, Albert Einstein. In taking this point of view, however, one tends to overlook the fact that gravitation has been the subject of controversial discussion since the time of Newton. That Newton's theory of gravitation assumes action at a distance, i.e., action without an intervening mechanism or medium, was perceived from its earliest days as being problematical. Around the turn of the last century, in the twilight of classical physics, the problems of Newtonian gravitation theory had become more acute. Consequently, there was a proliferation of alternative theories of gravitation which were quickly forgotten after the triumph of general relativity. In order to understand this triumph, it is necessary to compare general relativity to its contemporary competitors. As we shall see, general relativity owes much to this competition. A historical analysis of the struggle between alternative theories of gravitation and the different approaches to the problem of gravitation thus complements the analysis of Einstein's efforts. An account of the genesis of general relativity that does not discuss these competitors remains incomplete and biased. At the same time, this wider perspective on the emergence of general relativity provides an exemplary case of alternatives in the history of science, presenting a whole array of alternative theories of gravitation and the eventual emergence of a clear winner. It is thus an ideal topic for addressing long-standing questions in the philosophy of science on the basis of detailed historical evidence.

The present book, comprising volumes 3 and 4 of the series, discusses alternative theories of gravitation that were relevant to the genesis of general relativity and thus constitute its immediate scientific context. Many of these theories figured in the discussions of Einstein and his contemporaries. The set of theories covered here is not complete as far as gravitation theories in the late nineteenth and early twentieth centuries are concerned. But even a comprehensive treatment of this narrower set of theories represents a considerable challenge for the history of science. Unlike Einstein's work, many of the theories dealt with here are known only to a few specialists. This situation has only just begun to improve through recent work on the genesis of general relativity and much remains to be done in the future.

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 3
Gravitation in the Twilight of Classical Physics: Between Mechanics, Field Theory, and Astronomy.

What is presented in these volumes are two types of texts, sources and interpretations. The sources are key documents relevant to the history of general relativity. Many of these texts were originally written in German and are presented here for the first time in English translation. The interpretative texts are essays, most of them specially written for these volumes. They provide historical context and analysis of the theories presented in the sources. The book is divided into sections reflecting a classification of approaches to the problem of gravitation. Different subdisciplines of classical physics generated different ways of approaching the problem of gravitation. The emergence of special relativity further raised the number of possible approaches while creating new requirements that all approaches had to come to terms with. Each section of this book is dedicated to one of these approaches and, as a rule, consists of an historical essay and several sources.

THE UNFOLDING OF ALTERNATIVE THEORIES OF GRAVITATION

From the perspective of an epistemologically oriented history of science, the unfolding of alternative theories of gravitation in the twilight of classical physics can be interpreted as the realization of the potential embodied in the knowledge system of classical physics to address the problem of gravitation, this knowledge system eventually being transformed by the special-relativistic revolution. The dynamics of this unfolding was largely governed by internal tensions of the knowledge system rather than by new empirical knowledge, which at best played only a minor role. A central problem of the Newtonian theory of gravitation was, as already mentioned, that it assumed the action between two attracting bodies to be instantaneous and that it did not provide any explanation for the instantaneous transport of action along arbitrary distances. This characteristic feature of the gravitational force, called action at a distance, became even more dubious after the mid-nineteenth century when it was recognized that electromagnetic forces did not comply with the idea of action at a distance. This internal tension of the knowledge system of classical physics was intensified, but not created, by the advent of the theory of special relativity, according to which the notion of an instantaneous distance between two bodies as it appears in Newton's force law can no longer be accepted.

The attempts to resolve these kinds of tensions typically crystallized around mental models representing the gravitational interaction on the basis of other familiar physical processes and phenomena. A mental model is conceived here as an internal knowledge representation structure serving to simulate or anticipate the behavior of objects or processes, like imagining electricity as a fluid. Mental models are flexible structures of thinking that are suitable for grasping situations about which no complete information is available. They do so by relying on default assumptions that result from prior experiences and can be changed if additional knowledge becomes available without having to give up the model itself.

Thus, in what may be called the *gas model*, gravitation could be conceived as resulting from pressure differences in a gaseous aether. Or, in what may be called the *umbrella model*, the attraction of two bodies could be imagined to result from the mutual shielding of the two bodies immersed in an aether whose particles rush in random directions and, in collisions with matter atoms, push them in the direction of the particles' motions. Or one could think of gravitation in analogy to the successful description of electromagnetism by the *Lorentz model*, accepting a dichotomy of gravitational field on the one hand and charged particles—masses—that act as sources of the field on the other. The elaboration of these approaches, with the help of mathematical formalism, led typically to a further proliferation of alternative approaches and, at the same time, provided the tools to explore these alternatives to a depth that allowed new tensions to be revealed. The history of these alternative approaches can thus be read, in a way similar to Einstein's work, as an interaction between the physical meaning embodied in various models and the mathematical formalism used to articulate them.

THE POTENTIAL OF CLASSICAL PHYSICS

The history of treatments of gravitation in the nineteenth century reflects the transition from an era in which mechanics constituted the undisputed fundamental discipline of physics to an era in which mechanics became a subdiscipline alongside electrodynamics and thermodynamics.

From the time of its inception, the action-at-a-distance conception of Newtonian gravitation theory was alien to the rest of mechanics, according to which interaction always involved contact. This explains the early occurrence of attempts to interpret the gravitational force by means of collisions, for instance, by invoking the umbrella model described above. During these early days the comparison of the gravitational force to electric and magnetic forces had already been suggested as well. However, the analogy with electricity and magnetism became viable only after theories on these subjects had been sufficiently elaborated. There were even attempts at thermal theories of gravitation after thermodynamics had developed into an independent subdiscipline of physics. Besides providing new foundational resources for approaching the problem of gravitation, the establishment of independent subdisciplines and the questioning of the primacy of mechanics that resulted from it affected the development of the theoretical treatment of gravitation in yet another way, namely through the emergence of revisionist formulations of mechanics. This *heretical mechanics*, as we shall call it, consisted in attempts to revise the traditional formulation given to mechanics by Newton, Euler and others, and often amounted to questioning its very foundations.

Approaches to the problem of gravitation in the context of these developments of classical physics are covered by the first and last sections of the first volume of this book, *The Gravitational Force between Mechanics and Electrodynamics*, and *From Heretical Mechanics to a New Theory of Relativity.* Further stimuli for rethinking gravitation came from the development of astronomy and mathematics. This approach is addressed by the second section, *An Astronomical Road to a New Theory of Gravitation*.

The Mechanization of Gravitation

Before the advent of the special theory of relativity, the validity of Newton's law of gravitation was essentially undisputed in mainstream physics. Alternative laws of gravitation were, of course, conceivable but Newton's law proved to be valid to a high degree of precision. While the minute discrepancies between the observed celestial motions and those predicted by Newtonian theory, most prominently the advance of Mercury's perihelion, could be resolved by one of these alternatives, they could also be resolved by adjusting lower-level hypotheses such as those regarding the distribution of matter in the solar system. In any case, the empirical knowledge at that time did not force a revision of Newtonian gravitation theory. The more pressing problem of this theory was that it did not provide a convincing model for the propagation of the gravitational force.

The most elaborate theories to address this problem made use of the umbrella model. These theories start from the idea of an impact of aether particles on matter, as formulated by Le Sage in the late eighteenth century. The gravitational aether is imagined to consist of particles that move randomly in all directions. Whenever such an aether atom hits a material body it pushes the body in the direction of its movement. A single body remains at rest since the net impact of aether particles from all sides adds up to zero. However, if two bodies are present, they partly shield each other from the stream of aether particles. As a result, the impact of aether particles on their far sides outweighs that on their near sides and the two bodies are driven towards each other.

Caspar Isenkrahe, Sir William Thomson (Lord Kelvin), and others developed different theories based on this idea in the late nineteenth century. But regardless of the details, this approach suffers from a fundamental problem related to the empirical knowledge about the proportionality of the force of gravity with mass. In order to take this into account one needs to allow the aether particles to penetrate a material body in such a way that they can interact equally with all of its parts. This requirement is better fulfilled the more transparent matter is to the aether particles. But, the more transparent matter is, the less shielding it provides from the aether particles on which the very mechanism for explaining gravity is based. Hence, without shielding there is no gravitational effect; without penetration there is no proportionality of the gravitational effect to the total mass. Furthermore, in theories explaining gravitation by the mechanical action of a medium, the problem of heat exchange between the medium and ordinary matter arises (in analogy to electromagnetic heat radiation), in most approaches leading to an extreme heating of matter.

From a broader perspective, such attempts at providing a mechanical explanation of gravity had lost their appeal by the end of the nineteenth century after the successful establishment of branches of physics that could not be reduced to mechanics, such as Maxwell's electrodynamics and Clausius' thermodynamics. Nevertheless, this development led indirectly to a contribution of the mechanical tradition to solving the problem of gravitation by provoking the emergence of revised formulations of mechanics, referred to here as heretical mechanics.

Heretical Mechanics

A critical revision of mechanics, pursued in different ways by Carl Neumann, Ludwig Lange, and Ernst Mach among others, had raised the question of the definition and origin of inertial systems and inertial forces, as well as their possible relations to the distribution of masses in the universe. Through the latter issue, this revision of mechanics was also important for the problem of gravitation. It also gave rise to attempts at formulating mechanics in purely relational terms, that is, exclusively in terms of the mutual distances of the particles and derivatives of these distances. Such attempts are documented, for instance, in the texts presented in this book of Immanuel and Benedict Friedlaender and of August Föppl. As becomes clear from these texts, heretical mechanics contributed to understanding the relation between gravitational and inertial forces as both are due to the interaction of masses. According to Föppl there must be velocity-dependent forces between masses although he did not think of these forces as being gravitational. The Friedlaender brothers also conceived of inertia as resulting from an interaction between masses and did speculate on its possible relation to gravitation. In spite of such promising hints, heretical mechanics remained marginal within classical physics, in part because it lacked a framework with which one could explore the relation between gravitation and inertia. This relation was established by Einstein within the framework of field theory, first in 1907 through his principle of equivalence, and more fully with the formulation of general relativity.

Einstein's successful heuristic use of Machian ideas in his relativistic theory of gravitation encouraged the mechanical tradition to continue working toward a purely relational mechanics in the spirit of Mach. Attempts in this direction were made by Hans Reissner, Erwin Schrödinger, and, more recently, Julian Barbour and Bruno Bertotti. The success of general relativity provided a touchstone for the viability of these endeavors. At the same time, the question to which extent the issues raised by heretical mechanics, such as a relational understanding of inertia, have been settled by general relativity is still being discussed today.

From Peripheral Mathematics to a New Theory of Gravitation

The success or failure of a physical idea hinges to a large extent on the mathematical tools available for expressing it. In view of the crucial role of the mathematical concept of affine connection at a later state in the development of the general theory of relativity, it is interesting to consider the impact this tool might have had on the formulation of physical theories had it been part of mathematics by the latter half of the nineteenth century. That this counter-factual assumption is actually not that far-fetched can be seen from the work of Hermann Grassmann, Heinrich Hertz, Tullio Levi-Civita, and Elie Cartan, in part reproduced in this book. Such a fictive development might have given rise to a kind of heretical gravitation theory driven by peripheral mathematics and formulated by some "Newstein" long before the advent of special relativity. Perhaps the search for a different conceptualization of mechanics in which gravitation and inertia are treated alike, as is the case according to Einstein's equivalence principle, could have provided a physical motivation for such an alternative formulation of classical mechanics with the help of affine connections. Perhaps Heinrich Hertz's attempt to exclude forces from mechanics, replacing them by geometrical constraints, might have served as a starting point for such a development, triggering a geometrization of physics, had it not been so marginal to the mainstream of late nineteenth-century physics.

As with ordinary classical mechanics, Newstein's theory would have eventually conflicted with the tradition of electrodynamics and its implication of a finite propagation speed for physical interactions, which ultimately leads to the metrical structure of special relativity with its constraints on physical interactions. Then the problem that arose from this conflict could be—in contrast to the actual course of history—formulated directly in terms of the compatibility of two well-defined mathematical structures, the affine connection expressing the equality in essence of gravitation and inertia, and the metric tensor expressing the causal structure of spacetime. This formulation of the problem would have smoothed the pathway to general relativity considerably since the heretical aspect of Einstein's work—the incorporation of the equality in essence of gravitation and inertia—would have already been implemented in Newstein's predecessor theory. General relativity might thus have been the outcome of mainstream research.

The Potential of Astronomy

Another field of classical science that might have contributed more than it actually did to the emergence of general relativity is astronomy. This is made evident by the sporadic interventions by astronomers such as Hugo von Seeliger, who questioned the seemingly self-evident foundations of the understanding of the universe in classical science. Their work was stimulated by new mathematical developments such as the emergence of non-Euclidean geometries or by heretical mechanics insofar as it raised questions relevant to astronomy, for instance, concerning the definition of inertial systems. It was further stimulated by the recognition of astronomical deviations from the predictions of Newton's law (such as the perihelion advance of Mercury), or by the paradoxes resulting from applying classical physics to the universe-at-large when this is assumed to be infinite (such as the lack of definiteness in the expression of the gravitational force, or Olbers' paradox of the failure of the night sky to be as bright as the Sun).

Although the full extent to which these problems were connected became clear only after the establishment of general relativity, the astronomer Karl Schwarzschild, who was exceptional in his interdisciplinary outlook, addressed many of them and was even able to relate them to one another. He explored, for instance, the cosmological implications of non-Euclidean geometry and considered the possibility of an anisotropic large-scale structure of the universe in which interial frames can only be defined locally. With less entrenched disciplinary boundaries of late nineteenth-century classical science, such considerations could have had wider repercussions on the foundations of physics, perhaps giving rise to the emergence of a non-classical cosmology.

A Thermodynamic Analogy

In rejecting the assumption of an instantaneous propagation of gravitational interactions, it makes sense to modify classical gravitation theory by drawing upon analogies with other physical processes that have a finite propagation speed, such as the propagation of electromagnetic effects or the transport of heat in matter. Such analogies obviously come with additional conceptual baggage. A gravitational theory built according to the model of electrodynamic field theory, for instance, was confronted with the question of whether the gravitational analogue of electromagnetic waves really exist, or the question of why there is only one kind of charge (gravitational mass) in gravitation theory as opposed to two in electromagnetism (positive and negative charge). To avoid such complications, one could also consider amending Newtonian theory by extending the classical Poisson equation for the gravitational potential into a diffusion equation by adding a term with a first-order time derivative term, exploiting the analogy with heat transport in thermodynamics. In 1911, such a theory was proposed by Gustav Jaumann without, however, taking into account the spacetime framework of special relativity. As a consequence, it had little impact.

Electromagnetism as a Paradigm for Gravitation

Since early modern times magnetism served as a model for action at a distance as it apparently occurs between the constituents of the solar system. However, as long as there was no mathematical formulation describing magnetic forces, no quantitative description of gravitation could be obtained from this analogy. After Newton had established a quantitative description of gravitation, this could now conversely be used as a model for describing magnetic and electric forces, as realized in the laws of Coulomb, Ampère, and Biot-Savart. With the further development of electromagnetic theory as represented by velocity-dependent force laws and Maxwellian field theory, it regained its paradigmatic potential for understanding gravitation. After the striking success of Einstein's field theory of gravitation, which describes the gravitational force in terms of the geometry of spacetime, gravitation took the lead again as attempts were made that aimed at a geometrical description of electromagnetism and the other fundamental interactions with a view toward the unification of all natural forces. Such attempts are still being made today.

The motive of unification also underlay nineteenth-century attempts to reduce gravitation to electricity, such as those of Ottaviano Fabrizio Mossotti and Karl Friedrich Zöllner, who interpreted gravity as a residual effect of electric forces. They assumed that the attractive electric force slightly outweighs the repulsive one, resulting in a universal attraction of all masses built up from charged particles. Ultimately, however, this interpretation amounts to little more than the statement that there is a close analogy between the fundamental force laws of electrostatics and Newtonian gravitation.

The paradigmatic role of electromagnetism for gravitation theory was boosted dramatically when electrodynamics emerged as the first field theory of physics. A field-theoretic reformulation of Newtonian gravity modelled on electrostatics was provided by the Poisson equation for the Newtonian gravitational potential. Even though the Poisson equation was merely a mathematical reformulation of Newton's law, it had profound implications for the physical interpretation of gravitation and introduced new possibilities for the modification of Newtonian gravitation theory. The analogy with electromagnetism raised the question of whether gravitational effects propagate with a finite speed like electromagnetic effects. A finite speed of propagation further suggested the existence of velocity-dependent forces among gravitating bodies, amounting to a gravitational analogue to magnetic forces. It also suggested the possibility of gravitational waves. In short, a field theory of gravitation opened up a whole new world of phenomena that might or might not be realized in nature.

The uncertainty of the existence of such phenomena was in any case not the most severe problem that a field theory of gravitation was confronted with. If gravitation is conceived of as a field with energy content, the fact that like "charges" always attract has a number of problematic consequences. First and foremost, ascribing energy to the gravitational field itself leads to a dilemma that does not occur in the electromagnetic case. In the latter case, the work performed by two attracting charges as they approach each other can be understood to be extracted from the field, and the field energy disappears when they meet at one point. In contrast, while work can also be performed by two approaching gravitating masses, the field energy is enhanced, rather than diminished, as they come together at one point. (Accordingly no equivalent of a black hole is known in electrodynamics.) As Gustav Mie explains in his paper on the gravitational potential presented in this book, the gravitational field is peculiar in that it becomes stronger when work is released. While a similar effect occurs with the magnetic field of two current-bearing conductors, the source of the energy is obvious in this case. The energy comes from an external energy supply such as a battery. Such an external supply is missing in the case of gravitation. A plausible escape strategy was to assume that the energy of the gravitational field is negative so that, when the field becomes stronger, positive energy is released, which can be exploited as work. For the plausible option of formulating a theory of gravitation in strict analogy to electrodynamics by simply postulating Maxwell's equations with appropriately changed signs for the gravitational field, this negative energy assumption has dramatic consequences when considering dynamic gravitational fields. A minute deviation of a gravitating system from equilibrium will cause the field to release more and more energy, while the system deviates further and further from its original state of equilibrium. In fact, due to the reversed sign, gravitational induction, if conceived in analogy to electromagnetic induction, becomes a self-accelerating process. This will be referred to here as the *negative energy problem*.

Despite this problem, Hendrik Antoon Lorentz took up the thread of Mossotti and others and proposed to treat gravitation as a residual force resulting from electromagnetism. While the electromagnetic approach to gravitation offered, in principle, the possibility to account for observed deviations from Newtonian gravitation theory, the field theories actually elaborated by Lorentz and others failed to yield the correct value for the perihelion advance of Mercury, a commonly used touchstone.

All in all, the analogy of gravitation with electromagnetism, promising as it must have appeared, could not be as complete as advocated by its proponents. The considerable potential of the tradition of field theory for formulating a new theory of gravitation still needed to be explored and the key to disclosing its riches had yet to be discovered.

The attempts to subsume gravitation under the familiar framework of electromagnetism were later followed by approaches that aimed at a unification of physics on a more fundamental level, still focusing, however, on gravitation and electromagnetism. The most prominent attempts along these lines, contemporary to the genesis of general relativity, were those of Gustav Mie and David Hilbert. Their works are covered in the sections *From an Electromagnetic Theory of Matter to a New Theory of Gravitation* and *Including Gravitation in a Unified Theory of Physics*. These attempts, however, only led to a formal integration of the two forces without offering any new insights into the nature of gravity.

The key to successfully exploiting the resources of field theory for a new theory of gravitation was only found when the challenge of formulating a gravitational field theory was combined with insights from heretical mechanics. Instead of attempting a formal unification of two physical laws, Einstein combined the field theoretic approach with the idea of an equality in essence of gravitation and inertia, and eventually achieved an integration of two knowledge traditions hitherto separated due to the high degree of specialization of nineteenth-century physics.

THE CHALLENGE OF SPECIAL RELATIVITY FOR GRAVITATION

The advent of special relativity in 1905 made the need for a revision of Newtonian gravitation theory more urgent since an instantaneous propagation of gravitation was incompatible with the new spacetime framework in which no physical effect can propagate faster than the speed of light. A revision of this kind could be achieved in various ways. One could formulate an action-at-a-distance law involving a finite time of propagation as had been developed in electromagnetism, e.g. by Wilhelm Weber. Or one could formulate a genuine field theory of gravitation. The four-dimensional formulation of special relativity emerging from the work of Henri Poincaré, Hermann Minkowski, and Arnold Sommerfeld brought about a set of clearly distinguished alternative approaches for realizing such a field theory of gravitation. Eventually, however, due to the implications of special relativity not only for the kinematic concepts of space and time but also for the dynamic concept of mass, gravitation was bursting out of the framework of special relativity.

A New Law of Gravitation Enforced by Special Relativity

The simplest way to make gravitation theory consistent with special relativity was to formulate a new direct particle interaction law of gravitation in accordance with the conditions imposed by special relativity, e.g., that the speed of propagation of the gravitational force be limited by the speed of light. This kind of approach, which was pursued by Poincaré in 1906 and by Minkowski in 1909, and which is presented here in the section *A New Law of Gravitation Enforced by Special Relativity*, could rely on the earlier attempts to introduce laws of gravitation with a finite speed of propagation. However, the stricter condition of Lorentz invariance now had to be satisfied.

While the formulation of a relativistic law of gravitation could solve the particular problem of consolidating gravitation theory with the new theory of special relativity, it disregarded older concerns about Newtonian gravitation, such as those relating to action at a distance. Furthermore, questions concerning fundamental principles of physics such as that of the equality of action and reaction emerged in these formulations. It remained, in any case, unclear to which extent the modified laws of gravitation could be integrated into the larger body of physical knowledge.

Towards a Field Theory of Gravitation

More important and more ambitious than the attempts at a new direct-particle interaction law of gravitation was the program of formulating a new field theory of gravitation. As pointed out above, if gravitation—in analogy to electromagnetism—is transmitted by a field with energy content, the fact that in the gravitational case like "charges" (masses) attract has problematic consequences, such as the negative energy problem. A promising approach to the negative energy problem was the assumption that masses also have energy content defined in such a way that the energy content of two attracting masses decreases when the masses approach each other. This effect can in turn be ascribed to a direct contribution of the gravitational potential to the energy content of the masses. Hence, there is a way to infer a relation between mass and energy content by considering the negative energy problem of a gravitational field theory.

The above considerations on the negative energy problem of gravitational field theory suggest that the potential plays a greater role in such a theory than it does in classical electromagnetic field theory. How to represent the gravitational potential is further directly connected with the question of how to represent the gravitational mass, or, more generally, the source of the gravitational field, since both are related through the field equation. The following three mathematical types of potentials were considered before the establishment of general relativity with the corresponding implications on the field strengths and the sources.

- *Scalar theories*. Potential and source are Lorentz scalars and the field strength is a (Lorentz) four-vector.
- *Vector theories*. Potential and source are four-vectors and the field is what was then called a "six-vector" (an antisymmetric second-rank tensor).
- *Tensor theories*. Potential and source are symmetric second-rank tensors and the field is represented by some combination of derivatives of the potential.

From what has been said above about a theory of gravitation construed in analogy with electrodynamics, the problems of a vector theory become apparent. In contrast to the electromagnetic case, where the charge density is one component of the four-current, the gravitational mass density is not one component of a four-vector. From this it follows in particular that no expression involving the mass is available to solve the negative energy problem by forming a scalar product of source and potential in order to adjust the energy expression.

Having thus ruled out vector theories, only scalar theories and tensor theories remain. Einstein's theories, in particular the *Entwurf* theory and his final theory of general relativity, belong to the latter class. Further alternative tensor theories of gravitation were proposed, but only after the success of general relativity, which is why they are not discussed here. As concerns scalar theories, a further branching of alternatives occurs as shall be explained in the following.

Every attempt to embed the classical theory of gravitation into the framework of special relativity had to cope not only with its kinematic implications, that is, the new spacetime structure which required physical laws be formulated in a Lorentz covariant manner, but also with its dynamical implications, in particular, the equivalence of energy and mass expressed by the formula $E = mc^2$. Since, in a gravitational field, the energy of a particle depends on the value of the gravitational potential at the position of the particle, the equivalence of energy and mass suggests that either the particle's mass or the speed of light (or both) must also be a function of the potential. Choosing the speed of light as a function of the potential immediately exits the framework of special relativity, which demands a constant speed of light. It thus may seem that choosing the inertial mass to vary with the gravitational potential is preferable since it allows one to stay within that framework.

According to contemporary evidence and later recollections, Einstein in 1907 explored both possibilities, a variable speed of light and a variable mass. He quickly came to the conclusion that the attempt to treat gravitation within the framework of special relativity leads to the violation of a fundamental tenet of classical physics, which may be called *Galileo's principle*. It states that in a gravitational field all bodies fall with the same acceleration and that hence two bodies dropped from the same height with the same initial vertical velocity reach the ground simultaneously. The latter formulation generalizes easily to special relativity. If the inertial mass increases with the energy content of a physical system, as is implied by special relativity, a body with a horizontal component of motion will have a greater inertial mass than the same body without such a motion, and hence fall more slowly than the latter.

The same conclusion can be drawn by purely kinematic reasoning in the framework of special relativity. Consider two observers, one at rest, the other in uniform horizontal motion. When the two observers meet, they both drop identical bodies and watch them fall to the ground. From the viewpoint of the stationary observer, the body he has dropped will fall vertically, while the body the moving observer has dropped will fall along a parabolic trajectory. From the viewpoint of the moving observer, the roles of the two bodies are interchanged: the first body will fall along a parabolic trajectory while the second will fall vertically.

If one now assumes that, in the reference frame of the stationary observer, the bodies will touch the ground simultaneously, as is required by Galileo's principle in the above formulation, the same cannot hold true in the moving system due to the relativity of simultaneity. In other words, Galileo's principle cannot hold for both observers. Thus, the assumption of Galileo's principle leads to a violation of the principle of relativity. On the other hand, if one assumes, in accordance with the principle of relativity, that the two observers both measure the same time of fall for the body falling vertically in their respective frame of reference, the time needed for the body to fall along a parabolic path can be determined from this time by taking time dilation into account. It thus follows that the time needed for the fall along a parabolic path is longer than the time needed for the vertical fall, in accordance with the conclusion drawn from the dynamical assumption of a growth of inertial mass with energy content.

Both of the possibilities considered by Einstein, a dependence on the gravitational potential either of the speed of light or of the inertial mass, were later explored by Max Abraham and Gunnar Nordström respectively. These theories, which represented the main competitors of Einstein's theories of gravitation, are discussed in the sections *The Problem of Gravitation as a Challenge for the Minkowski Formalism* and *A Field Theory of Gravitation in the Framework of Special Relativity*, to which the remainder of this introduction is devoted.

The Problem of Gravitation as a Challenge for the Minkowski Formalism

The assumption of a dependence of the speed of light on the gravitational potential made it necessary to generalize the Minkowski formalism, although the full consequences of this generalization became clear only gradually. It was Max Abraham who took the first steps in this direction by implementing Einstein's 1907 suggestion of a variable speed of light related to the gravitational potential within this formalism. Questioned by Einstein about the consistency of the modified formalism with

Minkowski's framework, he introduced the variable line element of a non-flat four-dimensional geometry.

Abraham's theory stimulated Einstein in 1912 to resume work on a theory of gravitation. Apart from developing his own theory, Abraham also made perceptive observations on alternative options for developing a relativistic theory of gravity, on internal difficulties as well as on physical and astronomical consequences such as energy conservation in radioactive decay or the stability of the solar system.

A Field Theory of Gravitation in the Framework of Special Relativity?

While Abraham explored the implications of a variable speed of light, Nordström pursued the alternative option of a variable mass. Nordström thus remained within the kinematic framework of special relativity. As in all such approaches, however, he did so at the price of violating to some extent Galileo's principle.

More importantly, Nordström also faced the problem that in a special relativistic theory of gravitation the dynamical implications of special relativity need to be taken into account as well. These dynamical consequences suggested, for example, ascribing to energy not only an inertial but also a gravitational mass, which immediately implies that light rays are curved in a gravitational field. This conclusion, however, is incompatible with special relativistic electrodynamics in which the speed of light is constant.

Another implication of the dynamic aspects of special relativity concerns the source of the gravitational field. If any quantity other than the energy-momentum tensor of matter is chosen as a source-term in the gravitational field equation, as is the case in all scalar theories including Nordström's, gravitational mass cannot be fully equivalent to inertial mass, whose role has been taken in special relativistic physics by the energy-momentum tensor. However, while such conceptual considerations cast doubt on the viability of special relativistic theories of gravitation, they were not insurmountable hurdles for such theories. In fact, Nordström's final version of his theory remained physically viable as long as no counter-evidence was known. Einstein's successful calculation of Mercury's perihelion advance on the basis of general relativity in late 1915 undermined Nordström's theory, which did not yield the correct value. This, however, did not constitute a fatal blow as long as other astrophysical explanations of Mercury's anomalous motion remained conceivable. The fatal blow only came when the bending of light in a gravitational field was observed in 1919. Nordström's theory did not predict such an effect. For the final version of his theory this can easily be seen by observing that the trace of the energy-momentum tensor, which acts as the source of the gravitational field in that theory, vanishes for electromagnetic fields. Another way of seeing this makes use of the work of Einstein and Adriaan Fokker, who showed that Nordström's theory can be viewed as a special case of a metric theory of gravitation with the additional condition that the speed of light is a constant, thus excluding a dispersion of light waves that gives rise to the bending of light rays.

Before Nordström's theory matured to its final version, which constitutes a fairly satisfactory special relativistic theory of gravitation, several steps were necessary in which the original idea was elaborated, in particular regarding the choice of an appropriate source expression. The most obvious choice and the first considered by Nordström is the rest mass density. The problem with this quantity is, however, that it is not a Lorentz scalar. Nordström's second choice was the Lagrangian of a particle. This, however, leads to a violation of the equality of gravitational and inertial mass. While according to special relativity, kinetic energy, (e.g. the thermal motion of the particles composing a body), adds to the body's inertial mass, it is subtracted from the potential energy in the Lagrangian. If that Lagrangian hence describes the gravitational mass, the difference between the two masses increases as more kinetic energy is involved. In his final theory Nordström chose, at Einstein's suggestion, the trace of the energy-momentum tensor, the Laue scalar, thus extending the validity of the equivalence principle from mass points at rest to "complete static systems." A complete static system is a system for which there exists a reference frame in which it is in static equilibrium. In such a frame, the mechanical behavior of the system is essentially determined by a single scalar quantity. In fact, since in special relativity the inertial behavior of matter is determined by the energy-momentum tensor, the requirement of equality of inertial and gravitational mass implies that a scalar responsible for the coupling of matter to the gravitational field must be derived from the energy-momentum tensor.

The problem in choosing the Laue-scalar as a source expression is how deal with the transport of stresses in a gravitational field while maintaining energy conservation. Einstein argued that such stresses may be used—unless appropriate provisions are taken—to construct a *perpetuum mobile*, since—by creating or removing stresses—one can, so to speak, apparently switch gravitational mass on and off. In other words, while the work required for creating a stress can simply be recovered by removing it, the gravitational mass created by the stress can meanwhile be used to perform work in the presence of a gravitational field. Given that stresses depend on the geometry of the falling object under consideration, a solution can be found by appropriately adjusting the geometry, as Nordström showed. In other words, the assumption that gravitational mass can be generated by stresses led, in conjunction with the requirement of energy-momentum conservation, to the conclusion that the geometry has to vary with the gravitational potential.

According to Einstein's assessment of Nordström's final theory in his Vienna lecture, the theory satisfies all one can require from a theory of gravitation based on contemporary knowledge, which did not yet include the observation of light deflection in a gravitational field. At that time no known gravitation theory was able to explain Mercury's perihelion advance. Einstein's only remaining objection concerned the fact that what he considered to be Mach's principle—the assumption that inertia is caused by the interaction of masses—appears not to be satisfied in Nordström's theory.

But as we have seen, because of the role of stresses for gravitational mass, Nordström had to assume that the behavior of rods and clocks also depends on the gravitational potential. Indeed, as becomes clear from the hindsight of general relativity, it is arguable whether his theory really fits the special relativistic framework, corresponding as it does to a spacetime theory that is only conformally flat, i.e., based on a metric that is flat besides a scalar factor. The way that Nordström's theory stands to general relativity in that it attributes transformations to material bodies, which in the later theory are understood as transformations of spacetime, is reminiscent of the way that Lorentz's theory of the aether stands to special relativity.

ACKNOWLEDGEMENTS

For their careful reading of this text, extensive discussions, and helpful commentaries, we would like to thank Michel Janssen, Domenico Giulini, Christopher Smeenk, and John Stachel. We would also like to thank Lindy Divarci, Christoph Lehner, Carmen Hammer, Miriam Gabriel, Yoonsuhn Chung, and Dieter Hoffmann for their considerable effort and stamina during the preparation of the index.

NOTES ON THE TRANSLATIONS

The translation of sources in these volumes, if not indicated otherwise, were produced through a collective effort. The majority of the texts were first translated by Dieter Brill. Werner Heinrich also contributed considerably to the initial translations. Other texts were translated by June Inderthal and Barbara Stepansky. The translations were further reworked by Dieter Brill, Lindy Divarci, Christopher Martin, Matthias Schemmel, and Christopher Smeenk. We would also like to thank Don Salisbury for helping us to improve the translation of Max Abraham's *Eine neue Gravitationstheorie* and Erhard Scholz for helping to improve the translation of Weyl's text.

For clarification of terminology, certain words are given in the original language. The occurrence of page breaks in the original texts are marked and page numbers are given in the margin.

SOURCES

The following list contains the literature referred to in this introduction. It further contains references to a selection of the numerous contributions to the understanding of gravitation that were published in the period between the end of the nineteenth century and the advent of general relativity, without claiming to be complete. Further references as well as historical commentaries on the development of gravitational theories are found, e.g., in N. T. Roseveare *Mercury's Perihelion from Le Verrier to Einstein* (Clarendon Press, 1982); Matthew R. Edward (ed.) *Pushing Gravity: New Perspectives on Le Sage's Theory of Gravitation* (Apeiron, 2002); G. J. Whitrow and G. E. Mroduch "Relativistic Theories of Gravitation: A comparative analysis with particular reference to astronomical tests" *Vistas in Astronomy* 1: 1–67; D. Giulini "Some remarks on the notions of general covariance and background independence" arXiv:gr-qc/0603087—a contribution to *An assessment of current paradigms in the physics of fundamental interactions,* I.O. Stamatescu (ed.), (Springer, to appear); as well as the *Einstein Studies* series, edited by Don Howard and John Stachel (Birkhäuser, 1989–). Sources which are reproduced (in part or in whole) in these volumes are marked in the following list in boldface.

Abraham, Max. 1912. "Zur Theorie der Gravitation." *Physikalische Zeitschrift* 13: 1–4.

———. 1912. "Das Elementargesetz der Gravitation." *Physikalische Zeitschrift* 13: 4–5.

———. 1912. "Der freie Fall." *Physikalische Zeitschrift* 13: 310–311.

———. 1912. "Das Gravitationsfeld." *Physikalische Zeitschrift* 13: 793–797.

———. 1912. "Die Erhaltung der Energie und der Materie im Schwerkraftfelde." *Physikalische Zeitschrift* 13: 311–314.

———. 1912. "Relativität und Gravitation. Erwiderung auf eine Bemerkung des Herrn A. Einstein." *Annalen der Physik* 38: 1056–1058.

———. 1912. "Nochmals Relativität und Gravitation. Bemerkungen zu A. Einsteins Erwiderung." *Annalen der Physik* 39: 444–448.

———. 1913. "Eine neue Gravitationstheorie." *Archiv der Mathematik und Physik* 20: 193–209.

———. 1915. "Neuere Gravitationstheorien." *Jahrbuch der Radioaktivität und Elektronik* (11) 4: 470–520.

Barbour, Julian and Bruno Bertotti. 1977. "Gravity and Inertia in a Machian Framework." *Nuovo Cimento* 38B: 1–27.

———. 1982. "Mach's Principle and the Structure of Dynamical Theories." *Proceedings of the Royal Society London* 382: 295–306.

Behacker, Max. 1913. "Der freie Fall und die Planetenbewegung in Nordströms Gravitationstheorie." *Physikalische Zeitschrift* (14) 20: 989–992.

Born, Max. 1914. "Der Impuls-Energie-Satz in der Elektrodynamik von Gustav Mie." *Nachrichten der Königlichen Gesellschaft der Wissenschaften zu Göttingen* 1914, 23–36.

Cartan, Elie. 1986. *On Manifolds With An Affine Connection And The Theory Of General Relativity*. Naples: Bibliopolis, chap. 1, 31–55.

Cunningham, E. 1914. *The Principles of Relativity*. Cambridge: Cambridge University Press.

Einstein, Albert. 1907. "Über das Relativitätsprinzip und die aus demselben gezogenen Folgerungen." *Jahrbuch der Radioaktivität und Elektronik* 4, 411–462.

———. 1913. "Zum Gegenwärtigen Stande des Gravitationsproblems." *Physikalische Zeitschrift* 14: 1249–1269.

———. 1914. "Zum Relativitätsproblem." *Scientia* 15: 337–348.

———. 1920. *Äther und Relativitätstheorie*. Berlin: Springer.

Einstein, Albert, and Adriaan D. Fokker. 1914. "Die Nordströmsche Gravitationstheorie vom Standpunkt des absoluten Differentialkalküls." *Annalen der Physik* 44: 321–328.

Fokker, Adriaan D. 1914. "A Summary of Einstein and Grossmann's Theory of Gravitation." *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 29: 77–96.

Föppl, August. 1904. "Über absolute und relative Bewegung." *Sitzungsberichte der Bayerischen Akademie der Wissenschaften, mathematisch-physikalische Klasse* 34: 383–395.

———. 1905. "Über einen Kreiselversuch zur Messung der Umdrehungsgeschwindigkeit der Erde." *Königlich Bayerische Akademie der Wissenschaften, München, mathematisch-physikalische Klasse, Sitzungsberichte* 34: 5–28.

Friedlaender, Benedict and Immanuel. 1896. *Absolute oder relative Bewegung?* Berlin: Leonard Simion.

Gans, Richard. 1905. "Gravitation und Elektromagnetismus." *Physikalische Zeitschrift* 6: 803–805.

———. 1912. "Ist die Gravitation elektromagnetischen Ursprungs?" In *Festschrift Heinrich Weber zu seinem siebzigsten Geburtstag am 5. März gewidmet von Freunden und Schülern*. Leipzig, Berlin: Teubner, 75–94.

Gerber, Paul. 1917. "Die Fortpflanzungsgeschwindigkeit der Gravitation." *Annalen der Physik* 52: 415–441.

Gramatzki, H. J. 1905. *Elektrizität und Gravitation im Lichte einer mathematischen Verwandtschaft. Versuch zur Grundlage einer einheitlichen Mechanik der elektrischen gravitierenden und trägen Massen mit Hilfe der phänomenologischen Interpretation gewisser mathematischer Begriffsvorgänge*. München: Lindauer.

Grassmann, Hermann. 1844. *Die lineale Ausdehnungslehre, ein neuer Zweig der Mathematik*. Leipzig: Otto Wigand.

———. 1995. *A New Branch of Mathematics: The Ausdehnungslehre of 1844 and Other Works*. Translated by Lloyd C. Kannenberg. Chicago and LaSalle: Open Court.

Hertz, Heinrich. 1894. *Die Prinzipien der Mechanik in neuem Zusammenhange dargestellt*. Leipzig: Barth.

Hilbert, David. 1916. "Die Grundlagen der Physik. (Erste Mitteilung.)" *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-physikalische Klasse* 1915, 395–407.

———. 1917. "Die Grundlagen der Physik. (Zweite Mitteilung.)" *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-physikalische Klasse* 1916, 53–76.

Hofmann, Wenzel. 1904. *Kritische Beleuchtung der beiden Grundbegriffe der Mechanik: Bewegung und Trägheit und daraus gezogene Folgerungen betreffs der Achsendrehung der Erde und des Foucault'schen Pendelversuchs*. Wien, Leipzig: M. Kuppitsch Witwe.

Isenkrahe, Caspar. 1879. *Das Räthsel von der Schwerkraft: Kritik der bisherigen Lösungen des Gravitationsproblems und Versuch einer neuen auf rein mechanischer Grundlage*. Braunschweig: Vieweg.

Ishiwara, Jun. 1912. "Zur Theorie der Gravitation." *Physikalische Zeitschrift*13: 1189–1193.

———. 1914. "Grundlagen einer relativistischen elektromagnetischen Gravitationstheorie." Parts I and II. *Physikalische Zeitschrift 15*: 294–298, 506–510.

Jaumann, G. 1911. "Geschlossenes System physikalischer und chemischer Differentialgesetze." *Sitzungsberichte der math.-nw. Klasse der Kaiserl. Akademie der Wissenschaften*, *Wien* 120: 385–530.

———. 1912. "Theorie der Gravitation." *Sitzungsberichte der math.-nw. Klasse der Kaiserl. Akademie der Wissenschaften*, *Wien* 121: 95–182.

Lange, Ludwig. 1886. *Die geschichtliche Entwicklung des Bewegungsbegriffes und ihr voraussichtliches Endergebnis: ein Beitrag zur historischen Kritik der mechanischen Principien*. Leipzig: Engelmann.

Laue, Max von. 1917. "Die Nordströmsche Gravitationstheorie." *Jahrbuch der Radioaktivität und Elektronik* 14: 263–313.

LeSage, Georges-Louis. 1784. "Lucrèce Newtonien." *Mémoires de l'Académie royale des Sciences et Belles Lettres de Berlin, pour 1782*.

Levi-Civita, Tullio. 1916. "Nozione di parallelismo in una varietà qualunque e conseguente specificazione geometrica della curvatura riemanniana." *Circolo Matematico di Palermo. Rendiconti*, 42: 173–204.

Lorentz, Hendrik A. 1900. "Considerations on Gravitation." *Proceedings Royal Academy Amsterdam* 2: 559–574.

———. 1910. "Alte und neue Fragen der Physik." *Physikalische Zeitschrift* 11: 1234–1257.

———. 1914. "La Gravitation." *Scientia* (16) 36: 28–59.

Mach, Ernst. 1883. *Die Mechanik in ihrer Entwickelung: historisch-kritisch dargestellt*. Leipzig: Brockhaus.

Mie, Gustav. 1912. "Grundlagen einer Theorie der Materie I, II und III." *Annalen der Physik* 37: 511–534; 39 (1912), pp. 1–40; 40 (1913): 1–66.

———. 1914. "Bemerkungen zu der Einsteinschen Gravitationstheorie." *Physikalische Zeitschrift* 15: 115–122.

———. 1915. "Das Prinzip von der Relativität des Gravitationspotentials." In *Arbeiten aus den Gebieten der Physik, Mathematik, Chemie: Festschrift Julius Elster und Hans Geitel zum sechzigsten Geburtstag*. Braunschweig: Vieweg, 251–268.

Minkowski, Hermann. 1908. "Die Grundgleichungen für elektromagnetische Vorgänge in bewegten Körpern." *Nachrichten der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, 53–111.

Neumann, Carl. 1870. *Ueber die Principien der Galilei-Newton'schen Theorie*. Leipzig: Teubner.

Nordström, Gunnar. 1912. "Relativitätsprinzip und Gravitation." *Physikalische Zeitschrift* 13: 1126–1129.

———. 1913. "Träge und schwere Masse in der Relativitätsmechanik." *Annalen der Physik* 40: 856–878.

———. 1913. "Zur Theorie der Gravitation vom Standpunkt des Relativitätsprinzips." *Annalen der Physik* 42: 533–554.

———. 1914. "Über den Energiesatz in der Gravitationstheorie." *Physikalische Zeitschrift* 14: 375–380.

———. 1914. "Über die Möglichkeit, das elektromagnetische Feld und das Gravitationsfeld zu vereinigen." *Physikalische Zeitschrift* 15: 504–506.

———. 1914. "Zur Elektrizitäts- und Gravitationstheorie." *Finska Vetenskaps-Societetens Förhandlingar (LVII. A)* 4: 1–15.

Poincaré, Henri. 1906. "Sur la dynamique de l'électron." *Rendiconti del Circolo Matematico di Palermo* 21: 129–175.

Reißner, Hans. 1914. "Über die Relativität der Beschleunigungen in der Mechanik." *Physikalische Zeitschrift* 15: 371–375.

———. 1915. “Über eine Möglichkeit die Gravitation als unmittelbare Folge der Relativität der Trägheit abzuleiten.” *Physikalische Zeitschrift* (16) 9/10: 179–185.

Ritz, Walter. 1909. “Die Gravitation.” *Scientia* 5: 241–255.

Schrödinger, Erwin. 1925. “Die Erfüllbarkeit der Relativitätsforderung in der klassischen Mechanik.” *Annalen der Physik* 77, pp. 325–336.

Schwarzschild, Karl. 1897. “Was in der Welt ruht.” *Die Zeit, Vienna* Vol. 11, No. 142, 19 June 1897, 181–183.

———. 1900. “Über das zulässige Krümmungsmaass des Raumes.” *Vierteljahresschrift der Astronomischen Gesellschaft* 35: 337–347.

Seeliger, Hugo von. 1895. “Über das Newton’sche Gravitationsgesetz.” *Astronomische Nachrichten* 137: 129–136.

———. 1909. “Über die Anwendung der Naturgesetze auf das Universum.” *Sitzungsberichte der Königlichen Bayerischen Akademie der Wissenschaften, Mathematische-physikalische Klasse*, 3–25.

Sommerfeld, Arnold. 1910. “Zur Relativitätstheorie, II: Vierdimensionale Vektoranalysis.” *Annalen der Physik* 33: 649–689.

Thomson, William (Lord Kelvin). 1873. “On the ultramundane corpuscles of Le Sage.” *Phil. Mag*. 4th ser. 45: 321–332.

Wacker, Fritz. 1909. *Über Gravitation und Elektromagnetismus*. (Ph. D. thesis, Eberhard Karls Universität Tübingen). Leipzig: Noske.

Webster, D. L. 1912. “On an Electromagnetic Theory of Gravitation.” *Proceedings of the American Academy of Sciences* 47: 559–581.

Weyl, Hermann. 1918. “Reine Infinitesimalgeometrie.” *Mathematische Zeitschrift* 2: 384–411.

Zenneck, Jonathan. 1903. “Gravitation.” In Arnold Sommerfeld (ed.) *Encyklopädie der mathematischen Wissenschaften*, Vol. 5 (Physics). Leipzig: Teubner, 25–67.

Zöllner, Friedich, Wilhelm Weber, and Ottaviano F. Mossotti. 1882. *Erklärung der universellen Gravitation aus den statischen Wirkungen der Elektricität und die allgemeine Bedeutung des Weber’schen Gesetzes*. Leipzig: Staackmann.

———. 1898. “Die räumliche und zeitliche Ausbreitung der Gravitation.” *Zeitschrift für Mathematik und Physik* 43: 93–104.

THE GRAVITATIONAL FORCE BETWEEN MECHANICS AND ELECTRODYNAMICS

JÜRGEN RENN

THE THIRD WAY TO GENERAL RELATIVITY: EINSTEIN AND MACH IN CONTEXT

1. INTRODUCTION

The relationship between Einstein and Mach is often discussed as a prototypical case of the influence of philosophy on physics.[1] It is, on the other hand, notoriously difficult to accurately pinpoint such influences of philosophy on science, in particular with regard to modern physics. To a working scientist, such influences must seem to belong to a past era. There seems to be little room left for philosophy in the practice of today's physics. It plays virtually no part in the physics curriculum, and scholars who are both active physicists and philosophers are rare exceptions. In view of this situation it may be appropriate to reexamine the mythical role that philosophy played for one of the founding heroes of modern physics, Albert Einstein. It is conceivable that the disjointed remarks on philosophy that are dispersed throughout his œuvre can be integrated into a coherent image of what may then rightly be called "his philosophy." But even if such a reconstruction should be successful and yield more than an eclectic collection of occasional reflections, the more decisive question of the utility of philosophy for his science would be left unanswered. In fact, Einstein as a philosopher may have been a rather different *persona* from Einstein the physicist, and having two souls in one breast would not be an atypical state of affairs for a German intellectual. This contribution will therefore not undertake a systematic attempt at reconstructing his philosophy, but rather be limited to a case study of the interaction between philosophy and physics, reexamining the impact of Mach's philosophical critique of classical mechanics on Einstein's discovery of general relativity.[2] This reexamination is made possible by newly discovered documentary evidence concerning Einstein's research as well as by the achievements of recent studies in the history of general relativity.[3] Both factors contribute to an historical understanding of the

---

1 The literature on this subject is considerable; for more or less comprehensive accounts, see among others, (Blackmore 1992; Boniolo 1988; Borzeszkowski and Wahsner 1989, in particular, pp. 49–64; Goenner 1981; Hoefer 1994; Holton 1986, chap. 7; Norton 1993; Pais 1982, 282–288; Reichenbach 1958; Sciama 1959; Sewell 1975; Stein 1977; Torretti 1978, 1983, 194–202; Wolters 1987), as well as other literature quoted below. An earlier version of the present paper has been widely circulated in preprint form since 1994, an Italian version of this can be found in (Pisent and Renn 1994). Its themes have been taken up in various subsequent publications, see, e.g., (Barbour and Pfister 1995; CPAE 8).

2 For Mach's critique, see (Mach 1883: translated in Mach 1960).

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 3
Gravitation in the Twilight of Classical Physics: Between Mechanics, Field Theory, and Astronomy.

relation between Mach's philosophy and Einstein's physics that is not only richer in detail, but also in context, and hence able to reveal the alternatives available to the historical actors in the search for a new theory of gravitation.

The main result of the analysis presented below is that the theory of general relativity can be seen to have emerged as the result of one among several possible strategies dealing with conceptual problems of classical physics, strategies which were worked out to different degrees in the course of the historical development. Since this development was, in other words, not completely determined by the intrinsic features of the scientific problems which the historical actors confronted, it is now possible to evaluate more clearly the external factors affecting the choice between different strategies.[4] The approach pursued by Einstein can be characterized as a combination of field theoretical and mechanistic approaches shaped by his philosophical outlook on foundational problems of physics. In the following, two conclusions are drawn in particular:

i) The heuristics under the guidance of which Einstein elaborated general relativity was rooted in the heterogeneous conceptual traditions of classical physics. At least in its intermediate stages of development, the conceptual framework of Einstein's theory resembled the peculiar combination of field theoretic and mechanistic elements in Lorentz's electron theory, rather than the coherent and self-contained conceptual framework of special relativity, which had superseded the conceptual patchwork of Lorentz's theory.[5] Mach's ideas were one element in this mixture of traditional conceptual frameworks; their interpretation by Einstein depended on the context provided by the other elements. In particular, the heuristic role of Mach's ideas has to be seen in the wider context of the role that classical mechanics played for the emergence of general relativity. As with the other heuristic elements, Mach's ideas were eventually superseded by the conceptual consequences of general relativity, as Einstein saw them. In particular, Mach's concept of inertia as a property not of space but of the interaction between physical masses played a role comparable to that of the aether in Lorentz's theory of electrodynamics: it introduced a helpful heuristics that led to its own elimination, since the conceptual preconditions of the development of general relativity turned out to be incompatible with its outcome.

ii) What distinguished Einstein's early approach to the problem of gravitation from that of his contemporaries was his refusal to accept that a mechanistic and a field theoretic outlook on physics were mutually exclusive alternatives. It was his philosophical perspective on foundational problems of physics that allowed him to conceive of field theory and mechanics as complementary resources for the formulation of a new theory of gravitation. Contrary to most contemporary physicists dealing

---

3 For new evidence, see in particular, the various volumes of *The Collected Papers of Albert Einstein*. For recent historical studies of the development of general relativity, see the contributions to vols. 1 and 2 of this series and the references given therein.

4 For a similar kind of argumentation, see (Freudenthal 1986).

5 See the extensive discussion in "Classical Physics in Disarray …" and "Pathways out of Classical Physics …" (both in vol. 1 of this series).

with the problem of gravitation, he attempted to incorporate in his new theory both foundational assumptions of classical mechanics and their critical revision by Mach; and contrary to most physicists searching for a physical implementation of Mach's analysis of the foundations of mechanics, he took into account the antimechanistic philosophical intentions of this critique. Einstein's philosophical perspective is, however, not only characterized by his interest in and understanding of such philosophical intentions, but even more by his integrative outlook on the conceptual foundations of physics. His peculiar approach to the specific problem of gravitation can only be understood if one acknowledges that for him the problem of a new theory of gravitation was simultaneously the problem of developing new conceptual foundations for the entire body of physics. Although it may not be common to label such an integrative perspective as "philosophical"—in view of the predominantly metatheoretical concerns of the philosophy of science—it was also no longer a self-evident preoccupation of science at the beginning of this century, let alone of science today. Nevertheless, the fruitfulness of Einstein's approach argues for its reconsideration by both philosophy and science.

In the following, it will first be discussed how Einstein's project of generalizing the principle of relativity emerged in the context of his own research as well as in that of other contemporary approaches to the problem of gravitation (section 2). Some of the historical presuppositions of the conceptual innovation represented by general relativity will then be examined, paying particular attention to the contributions of mechanics and field theory to its development. The aim is to describe the horizon of possibilities open to the historical actors (section 3).[6] Next, the influence of Mach's critique of classical mechanics on the creation and interpretation of general relativity by Einstein (section 4) will be traced in some detail. Finally, the question of Einstein's philosophical perspective on the foundational problems of physics and its role in the emergence of general relativity (section 5) will be addressed once again.

2. A NEW THEORY OF GRAVITATION IN THE CONTEXT OF COMPETING WORLDVIEWS

2.1 A Relativistic Theory of Gravitation as a Problem of "Normal Science"

In 1907, when Einstein first dealt with the problem of a relativistic theory of gravitation, philosophical interests seemed to be far from his main concerns. Although he was employed by the Swiss patent office at that time, he was no longer an outsider to academic physics. By way of his publications, correspondence, and personal relationships, he was already becoming a well-respected member of the physics establishment. The times had passed when philosophical readings in the mock "Olympia" academy, which Einstein had founded some years earlier together with other bohemian friends, formed one of the centers of his intellectual life. Einstein was first con-

6 For the concept of horizon, see (Damerow and Lefèvre 1981).

fronted with the task of revising Newton's theory of gravitation in light of the relativity theory of 1905 when he was asked to write a review on relativity theory that would also cover its implications for various areas of physics not directly related to the field from which it originated, namely the electrodynamics of moving bodies.[7] Hence, the revision of Newton's theory of gravitation entered Einstein's intellectual horizon, not as the consequence of a philosophically minded ambition to go beyond the original special theory towards a more general theory of relativity, but as a necessary part of the usual "mopping up operation" whereby new results are integrated with the traditional body of knowledge. The necessity to modify the classical theory of gravitation appeared to Einstein and his contemporaries all the more pressing, as within the conceptual framework of classical physics an asymmetry could already be observed between the instantaneous propagation of the gravitational force and the propagation of the electromagnetic field with the finite speed of light. It therefore comes as no surprise that not only Einstein but also several of his contemporaries addressed the problem of formulating a field theory of gravitation that was to be in agreement with the principles suggested by the theory of the electromagnetic field and, most importantly, with the new kinematics of relativity theory.[8]

2.2 The Proliferation of Alternative Approaches to the Problem of Gravitation

It appears to be a phenomenon characteristic of the development of science that in a situation of conceptual conflict of this kind, alternative approaches to the solution of the conflict begin to proliferate. Among the factors accounting for this proliferation are the diverse resources upon which the alternative approaches can draw. Even after the establishment of special relativity, the instruments available for a revision of Newton's theory of gravitation had to be taken essentially from the arsenal of classical physics, in particular from classical mechanics and classical electrodynamics. As these two branches of classical physics were founded on different conceptual structures—on the one hand the direct interaction between point particles, and on the other hand, the propagation of continuous fields in time—the use of resources from one or the other branch to solve the same problem could present itself as a choice between conceptual alternatives. In this way, the problem of a new theory of gravitation assumed right from the beginning the character of a borderline problem of classical physics.[9] The choice among alternative approaches to the problem of gravitation was

7 See (Einstein 1907b, sec. V). See also Einstein's later recollections, e.g. those reported in (Wheeler 1979, 188). For a historical discussion of this paper, see (Miller 1992).

8 See, among others, the source papers in this and in vol. 4 of this series (Lorentz 1910; Minkowski 1908; Poincaré 1906), as well as the various papers by Abraham, Nordström, and Mie. See also (Abraham 1912a, 1912b; Lorentz 1910; Mie 1914; Minkowski 1911a [1908], in particular, pp. 401–404; Minkowski 1911b [1909], in particular, pp. 443–444; Nordström 1912; Poincaré 1905, in particular, pp. 1507–1508; 1906, in particular, pp. 166–175; Ritz 1909).

9 For the concept of borderline problem, see "Classical Physics in Disarray …" (in vol. 1 of this series).

therefore also related to the way in which such borderline problems were handled at that time.

Even before the turn of the century, that is, long before the great conceptual revolutions of early twentieth century physics, many physicists saw themselves at a crossroads, forced to decide between alternative conceptual foundations for their field.[10] Mechanics had long played the dual role of a subdiscipline and of an ontological foundation of physics, and at the threshold to the twentieth century, there were still physicists who adhered to the ontological primacy of mechanics, and who were therefore convinced that the entire body of physics should be built on conceptual foundations drawn from mechanics. With the formulation of classical electrodynamics by Maxwell, Hertz, and Lorentz, the difficulty of achieving such a reduction of physics to the conceptual apparatus of mechanics became increasingly evident. Although field theory itself was initially formulated in a mechanical language, towards the end of the century it came to represent an autonomous conceptual framework largely independent of that of mechanics. To some physicists, such as Wien and Lorentz, field theory even appeared to offer an alternative conceptual foundation for all of physics; they speculated about an electrodynamic worldview in which mechanics would have to be reformulated as a field theory rather than the other way around. Finally, with the development of classical thermodynamics in the mid-nineteenth century, including the formulation of the principle of conservation of energy, a third alternative conceptual foundation of physics (discussed under the name of "energetics") seemed to volunteer itself.[11] Hence, the mechanistic conception of physics, the electromagnetic worldview, and energetics distinguished themselves by the choice of the subdiscipline of classical physics to which they granted a foundational role for the entire field.

The formulation of a field theory of gravitation in analogy with, or even on the basis of, the Maxwell-Lorentz theory of the electromagnetic field was thus not a farfetched thought in the context of the electrodynamic world picture and had been approached by several authors.[12] In such a theory, gravitation, like electrodynamical interactions, would have to propagate with a finite speed. The establishment of the theory of relativity in 1905 did not make attempts in this direction obsolete; on the contrary, the issue of formulating a theory of gravity became even more urgent, since Newton's theory clearly violated one of the fundamental principles of the relativity theory—the requirement that no physical action propagates with a velocity greater than that of light. The primary task was to reformulate the experimentally well-con-

---

10 For further discussion of the conceptual foundations of classical physics at the turn of the century, see "Classical Physics in Disarray …" (in vol. 1 of this series). For a brief account of the different approaches prevalent at the turn of the century, see also (Jungnickel and McCormach 1986, chap. 24).

11 For a comprehensive historical analysis of energetics, see (Deltete 2000).

12 For contemporary reviews, see "Gravitation" (Zenneck 1903); "Recent Theories of Gravitation" (Abraham 1915), (both in this volume). For the heuristic role of electrodynamics for Einstein's formulation of a field theory of gravitation, see "Pathways out of Classical Physics …" (in vol. 1 of this series).

firmed Newtonian law of gravitation in accordance with the principles of the new kinematics, in particular with the Lorentz transformations of space and time coordinates, under which the classical law does not remain invariant. It is in fact not difficult to formulate a Lorentz covariant field equation which can be interpreted as a direct generalization of Newton's law. Around 1907 Einstein apparently pursued this line of research without, however, achieving satisfying results. Indeed, if such a Lorentz covariant generalization of Newton's theory could have been formulated without problems, there would have been no reason for Einstein to look beyond the special theory of relativity of 1905 and choose the thorny path that was to lead him to the formulation of the general theory of relativity in 1915.

One of the difficulties encountered by Einstein concerns the concept of mass, or more precisely the relation between the *two* concepts of mass in classical mechanics: gravitational and inertial. According to the special theory of relativity the inertial mass of a body depends on its energy content.[13] On the other hand, it was empirically known in the context of classical mechanics that the inertial mass is always exactly equal to the gravitational mass. In a relativistic theory of gravitation, the gravitational mass of a physical system should therefore also depend on its total energy. In a later recollection, Einstein summarized his view of this implication of classical mechanics and the special theory of relativity for a relativistic theory of gravitation:

> If the theory did not accomplish this or could not do it naturally, it was to be rejected. The condition is most naturally expressed as follows: the acceleration of a system falling freely in a given gravitational field is independent of the nature of the falling system (specially therefore also of its energy content).[14]

It was precisely this requirement, however, which turned out not to be fulfilled in the early attempts at a special relativistic theory of gravitation.[15]

In other words, a straightforward relativistic generalization of Newton's gravitational law seemed to be in conflict with "Galileo's principle," i.e., with the principle that the accelerations of bodies falling in a gravitational field are equal.[16] Quantitatively, however, the failure of the Galileo's principle may have been negligibly small, as Mie, for instance, claimed for his later special relativistic theory of gravitation.[17] Researchers such as Mie, whose outlook on this issue was shaped by the electrody-

---

13 See (Einstein 1907a), in which this conclusion is rederived in a general way, possibly with the problems of a relativistic theory of gravitation already in mind.

14 "Wenn die Theorie dies nicht oder nicht in natürlicher Weise leistete, so war sie zu verwerfen. Die Bedingung lässt sich am natürlichsten so aussprechen: die Fall-Beschleunigung eines Systems in einem gegebenen Schwerefelde ist von der Natur des fallenden Systems (speziell also auch von seinem Energie-Inhalte) unabhängig." (Einstein 1992, 64, 65)

15 For a reconstruction of Einstein's failed attempt to incorporate gravitation within the relativity theory of 1905, see "Classical Physics in Disarray …" sec. 2.9 (in vol. 1 of this series), see also "Einstein, Nordström, and the Early Demise of Scalar, Lorentz Covariant Theories of Gravitation" (in this volume). For Einstein's later recollections, see (Einstein 1992, 58–63).

16 Galileo's name is usually (but incorrectly) associated with the introduction of the principle of inertia, while the principle which is named after him here can indeed be found in his work; for historical discussion, see (Damerow et al. 2004, chap. 3).

namic worldview, were all the more willing to give up the Galileo's principle as they did not feel obliged to consider the implications of classical mechanics as foundational for physics, unless they perceived an unavoidable conflict with experimental evidence. Einstein, however, somewhat prematurely gave up this line of research. In the years 1912 to 1914, Nordström, with the help of contributions from von Laue and Einstein himself, attempted to formulate a consistent special relativistic field theory of gravitation and eventually succeeded to some extent in including the equality of gravitational and inertial mass.[18] This theory even triggered insights—e.g., that clocks and rods are affected by the gravitational field—upon which its further development in the direction of general relativity could have been based. Hence, it constituted at least the beginning of an independent road towards a theory similar to general relativity, "the route of field theory."

2.3 Mach's Critique of Mechanics and the Three Routes to General Relativity

From the conflict between classical mechanics and the special theory of relativity, which Einstein perceived in 1907, he drew a conclusion that was diametrically opposed to that of the followers of an electromagnetic worldview. For him the equality of inertial and gravitational mass was not just an empirically confirmed but otherwise marginal result of classical mechanics; he held onto it as a principle upon which a new theory of gravitation was to be based. He was therefore ready to accept that this theory would no longer fit into the framework of special relativity.[19] Hence Einstein's

---

17 See (Mie 1913, 50). Similar views are also found in other authors pursuing a special relativistic field theory of gravitation, see e.g. (Nordström 1912, 1129): "From a letter from Herr Prof. Dr. A. Einstein I learn that earlier he had already concerned himself with the possibility I used above for treating gravitational phenomena in a simple way. However, he became convinced that the consequences of such a theory cannot correspond with reality. In a simple example he shows that, according to this theory, a rotating system in a gravitational field will acquire a smaller acceleration than a non-rotating system. I do not find this result dubious in itself, for the difference is too small to yield a contradiction with experience." ("Aus einer brieflichen Mitteilung von Herrn Prof. Dr. A. Einstein erfahre ich, daß er sich bereits früher mit der von mir oben benutzten Möglichkeit befaßt hat, die Gravitationserscheinungen in einfacher Weise zu behandeln, daß er aber zu der Überzeugung gekommen ist, daß die Konsequenzen einer solchen Theorie der Wirklichkeit nicht entsprechen können. Er zeigt an einem einfachen Beispiel, daß nach dieser Theorie ein rotierendes System im Schwerkraftfelde eine kleinere Beschleunigung erhalten wird als ein nichtrotierendes. Diese Folgerung finde ich an sich nicht bedenklich, da der Unterschied zu klein ist, um einen Widerspruch mit der Erfahrung zu geben.")

18 For Nordström's work, see the section "A Field Theory of Gravitation in the Framework of Special Relativity," in particular "Einstein, Nordström and the Early Demise of Scalar, Lorentz Covariant Theories of Gravitation" (in this volume).

19 Einstein remarked with regard to the violation of Galileo's principle in Abraham's and Mie's theories of gravitation (Einstein 1914, 343): "Due to their smallness, these effects are certainly not accessible to experiments. But it seems to me that there is much to be said for taking the connection between inertial and gravitational mass to be warranted *in principle*, regardless of what forms of energy are taken into account." ("Diese Wirkungen wären zwar wegen ihrer Kleinheit dem Experiment nicht zugänglich. Aber es scheint mir viel dafür zu sprechen, dass der Zusammenhang zwischen der trägen und schweren Masse *prinzipiell* gewahrt ist, abgesehen von der Art der auftretenden Energieformen.")

further considerations did not lead him away from mechanics, but rather brought him into contact with its foundational questions, in particular with the role of inertial systems in classical mechanics.

Mach's philosophical critique of the foundations of classical mechanics suggested to Einstein that the problem of a new theory of gravitation had to be resolved in connection with a generalization of the relativity principle of classical mechanics and special relativity. Quite apart from the specific problem of gravitation, some of Mach's contemporary readers, as well as researchers who had independently arrived at similar views, had drawn the conclusion that one should look for a new, generally relativistic formulation of mechanics.[20] Their conceptual and technical resources were mostly confined to those of classical mechanics, and their chances of making contact with the more advanced results of physics at the turn of the century, which to a large extent were based on field theory (in particular, classical electrodynamics), were, at least at that time, slender. Nevertheless, the line of research that extends from the work of these early followers of Mach (discussed in more detail in the next section) to the recent work of Julian Barbour and Bruno Bertotti, Fred Hoyle and Jayant Narlikar, André Assis and others demonstrates that the project of formulating a generally relativistic theory of mechanics, including a treatment of gravitation, could be as successfully pursued as the project of a purely field theoretic approach to the problem of gravitation, as represented in particular by the work of Nordström.[21] In the following, this approach will be called "the mechanistic generalization of the relativity principle."

In view of this historical context, the heuristics that guided Einstein's formulation of the general theory of relativity can now be identified as a "third way," a peculiar mixture of field theoretical and mechanical elements. This affirmation suggests several questions, which are addressed in the following: What are the advantages and the disadvantages of the different strategies? What exactly are the contributions of the field theoretical and of the mechanical tradition to Einstein's heuristic strategy? What is the relation between the conceptual structures guiding Einstein's research and those that were newly established by it? As the development of the general theory of relativity was apparently not uniquely determined by the intrinsic nature of the problem to be solved, what then were the external factors that shaped Einstein's perspective and what role did philosophical positions play among them?

20 See, for example, the source texts in the first part of this volume "The Gravitational Force between Mechanics and Electrodynamics." For a survey of the interpretation of Mach's critique by contemporary readers, see also (Norton 1995).

21 For historical overviews of attempts to incorporate Mach's critique in physical theories, see (Assis 1995; Barbour 1993, Barbour and Pfister 1995; Goenner 1970, 1981).

3. ROOTS OF GENERAL RELATIVITY IN CLASSICAL PHYSICS

3.1 Resources and Stumbling Blocks Presented by the Tradition of Field Theory

The conceptual roots of general relativity in the tradition of field theory are more familiar than those in the tradition of mechanics. As mentioned above, not only special relativity but even the classical theory of the electromagnetic field made it plausible to conceive of gravitation as a field propagated with finite velocity. But there were also other contributions from this tradition which sooner or later found their way into the development of general relativity. Notably, field theory endows space with physical properties and thus contributes to blurring the distinction between matter and space. That this tendency (even taken by itself) could suggest the introduction of non-Euclidean geometry as a physical property of space is illustrated by the work of Georg Friedrich Bernhard Riemann and William Clifford in the nineteenth century.[22] In any case, field theory enriched the limited ontology of classical mechanics by introducing the field as a reality in its own right, an apparently trivial consequence, which, as we shall see, took considerable time to achieve a firm standing even within the development of general relativity. Field theory also suggested the existence of forces more general than the two-particle interactions usually considered in point mechanics, as is illustrated, for instance, by the transition from Coulomb forces between point charges to electrodynamic interactions such as induction; and it offered a mechanism for unifying separate forces as aspects of one more general field, as can again be illustrated by the example of electrodynamics conceived of as a unification of electric and magnetic interactions. It was therefore natural for those who pursued the program of formulating a field theory of gravitation either on the basis of, or in analogy to electrodynamics to search for the dynamic aspects of the gravitational field, considering Newton's law (in analogy to Coulomb's) as a description of the field's static aspects only. But the knowledge of the Newtonian special case could, and also did serve as a touchstone for any attempt at a more general theory—including Einstein's general theory of relativity, in whose development the question of the "Newtonian limit" was to play a crucial role.[23] The mature formulation of electrodynamic field theory by H. A. Lorentz also suggested a model for the essential elements of a field theory of gravitation and for their interplay: a field equation was needed to describe the effect of sources on the field, and an equation of motion was needed to describe the motion of bodies in the field.[24] Finally, those who looked for an "electromagnetic" theory of gravitation were also very clear about the

---

22 See (Clifford 1976 [1889]; Riemann 1868). On p. 149 of his paper, Riemann claims that non-Euclidean geometry could be important in physics if the concept of body should turn out not to be independent of that of space. He expected this consideration to be of relevance for a future microphysics.

23 See (Norton 1989b) and "Pathways out of Classical Physics …" (in vol. 1 of this series).

24 For a discussion of the historical continuity between Lorentz's electron theory and Einstein's theory of general relativity, see (McCormmach 1970) and "Pathways out of Classical Physics…" (in vol. 1 of this series).

experimental evidence to be accounted for by the new theory: the explanation of the perihelion shift of Mercury was in fact mentioned as an empirical check in almost all discussions of electromagnetic theories of gravitation, which, in this sense, can be said to have left a very tangible patrimony to general relativity in pointing to one of its classical tests.[25]

But as much as the tradition of field theory was able to contribute to the conceptual development of general relativity, it did not determine a heuristic strategy that clearly outlined the way to a satisfactory solution of the problem of gravitation. What is more, in hindsight, from the perspective of the completed theory of general relativity, it becomes evident that the tradition of classical field theory also included conceptual components that must be considered as stumbling blocks on the way to such a solution. In first turning to the problem of the heuristic ambiguity of field theory, as mentioned above, there were indeed several different lines to follow in formulating a field theory of gravitation within this tradition.[26] One of the factors accounting for this proliferation of alternatives lay in the uncertainty as to which principles of mechanics were to be maintained in a field theory of gravitation, given the necessity of revising at least some of them. The electromagnetic approach to the problem of gravitation tended, in any case, to ignore the foundational problems of mechanics, as long as this seemed experimentally acceptable. An early example of this tendency, characteristic of the electromagnetic world picture, is provided by the stepmotherly way in which, before the advent of special relativity, the principle of relativity and the principle of the equality of action and reaction was treated in Lorentz's electron theory. The same attitude characterized his later attempts to integrate gravitation into the conceptual framework of field theory. For instance, in a 1910 review paper (Lorentz 1910) Lorentz seemed unperturbed by the fact that the relativistic law of gravitation he proposed violated the principle of the equality of action and reaction. This difficulty is just one representative example of the problems associated with the task of reconstructing the shared knowledge accumulated in mechanics on the basis of purely field theoretic foundations. In addition to these problems, there was little experimental guidance in how to proceed in building the new theory of gravitation—apart from the speculations about the perihelion shift of Mercury mentioned above. To use a metaphor employed by Einstein (1913, 1250): the task of constructing a field theory of gravitation was similar to finding Maxwell's equations exclusively on the basis of Coulomb's law of electrostatic forces, that is, without any empirical knowledge of non-static gravitational phenomena.

Let us now address the problem of the conceptual stumbling blocks. Their evaluation naturally depends on the point of view one takes. In view of the conceptual framework of the finished general theory of relativity, classical field theory must have been misleading in several respects. One obvious aspect is the linearity of the classi-

---

25 See (Zenneck 1903). For a contemporary survey of the problem of gravitation and the role of the perihelion shift, see (Roseveare 1982).

26 See note 8 above.

cal theory in contrast to the non-linearity of the field equations of general relativity. A related aspect is the independence of the field equation and the equation of motion from each other in the classical theory, as opposed to their interdependence in general relativity. Closely associated with these more structural aspects—and perhaps even more important—are the conceptual changes with respect to classical physics brought about by general relativity. These changes include the introduction of new concepts of space and time, the new role of the gravitational field acting as its own source, and the changes of the concepts of energy and force manifested, for instance, by the absence of a gravitational stress-energy tensor in general relativity, in contrast to the existence of such a stress-energy tensor for the electromagnetic field in classical field theory. These changes could not have been anticipated on the basis of classical field theory; furthermore, in the search for a new theory of gravitation, classical field theory necessarily engendered expectations which were flatly contradicted by the outcome of that search.

3.2 The Foundational Critique of Mechanics and the Mechanistic Generalization of the Relativity Principle

The heuristic contributions of classical physics to the development of general relativity as well as the conceptual stumbling blocks it presented for this development obviously require a more detailed treatment and should be discussed in particular in the context of the concrete theories which are subsumed here under the rather general heading of "classical physics." For the purposes of the present contribution, an examination of this kind will be attempted only for the tradition of mechanics, for which one particular strand was of primary influence on the development of general relativity—both directly and as an alternative to Einstein's theory. This strand was represented by a reevaluation of mechanics, which was the outcome of a debate on its foundations in the second half of the nineteenth century. In this period some basic concepts of classical mechanics had ceased to be as self-evident as they had once appeared in the Newtonian tradition.

A central example is Newton's claim that even a single body in an otherwise empty universe possesses inertia, a claim which—in spite of its metaphysical character—played a crucial role in his argument in favor of the existence of absolute space.[27] This argument involves a bucket filled with water, which is considered first in a state in which the bucket rotates but the water is at rest and its surface flat, and second in a state in which both the bucket and the water rotate, producing a curved surface. According to Newton's interpretation of this experiment, the second case represents an absolute rotation, whereas the first case represents only a relative motion between water and bucket that does not cause physical effects. The conclusion that this argument provides evidence for the existence of absolute space is, however, only legitimate if other physical causes of the curvature of the water in the

27 This has been shown in detail in (Freudenthal 1986), on which also the following remarks are based.

second case can be excluded; in other words, the argument is convincing only under the physically unrealizable assumption that a rotational motion of the water in an otherwise empty universe would also give rise to the same effect. This assumption in turn is based on the metaphysical premise that a system is composed of parts which carry their essential properties (such as inertia in the case of a material system) even when they exist in isolation in empty space. It was also on this premise that Newton considered gravitation—in contrast to inertia—to be a universal but not an essential property of a material body.[28]

In the middle of the nineteenth century, a motivation for revisiting such metaphysical foundations of mechanics was provided by the establishment of non-mechanical theories such as electrodynamics and thermodynamics as mature subdisciplines of classical physics.[29] As a consequence of this development, mechanics not only lost its privileged status as the only conceivable candidate providing a conceptual basis for the entire building of physics, a status which was often associated with a claim to *a priori* truth, but also the conceptual foundation of mechanics itself could now be critically reexamined, including, for instance, the concept of absolute space and its justification by Newton. This revision of the status of the fundamental concepts of mechanics alone helped to prepare the conditions for a change of these concepts, should such a change become necessary in view of the growing body of knowledge.[30]

In any case, the critical reevaluation of the conceptual presuppositions of mechanics led to a proliferation of alternative approaches to the problem of gravitation, much as with the proliferation of alternative approaches within the framework of field theory. It was possible to elaborate more clearly the presuppositions upon which classical Newtonian mechanics was built, to revise the theory by attempting to eliminate those assumptions, which now appeared to be no longer acceptable (without any other substantial changes), or to formulate a new theory altogether. Carl Neumann's paper "On the Principles of the Galilean-Newtonian Theory" of 1869 provides an example of the first alternative: in order to replace Newton's concept of absolute space, he introduced the "body alpha" as the material embodiment of an absolute reference frame, comparing it with the luminiferous aether of electrodynamics as a like-

28 See the explanation of *Regula* III in (Newton 1972 [1726], 389).

29 Compare also the sequence in which Einstein, in his *Autobiographical Notes* (Einstein 1992), treats the *external* criticism of mechanics (the critique of mechanics as the basis of physics, pp. 22–23) and the "internal," conceptual criticism (pp. 24–31).

30 Compare e.g. the remark by Carl Neumann in 1869 (Neumann 1993 [1870], 367): "Finally, just as the present theory of electrical phenomena may perhaps one day be replaced by *another* theory, and the notion of an electric fluid could be removed, it is also the case that it is not an absolute impossibility that the Galilean-Newtonian theory will be supplanted one day by another theory, by some other picture, and the body alpha be made superfluous." ("Ebenso endlich, wie die gegenwärtige Theorie der elektrischen Erscheinungen vielleicht dereinst durch eine *andere* Theorie ersetzt, und die Vorstellung des elektrischen Fluidums beseitigt werden könnte; ebenso ist es wohl auch kein Ding der absoluten Unmöglichkeit, dass die Galilei-Newton'sche Theorie dereinst durch eine andere Theorie, durch ein anderes Bild verdrängt, und jener Körper Alpha überflüssig gemacht werde.") For the "body alpha" see below.

wise hypothetical, yet legitimate, conceptual element of the theory.[31] Nevertheless, by this reformulation Neumann did not intend to change the substance of Newton's theory, in particular not with respect to the question of relative and absolute motion, as the following passage illustrates:

> This seems to be the right place for an observation which forces itself upon us and from which it clearly follows how unbearable are the contradictions that arise when motion is conceived as something relative rather than something absolute.
>
> Let us assume that among the stars there is one which is composed of fluid matter and is somewhat similar to our terrestrial globe and that it is rotating around an axis that passes through its center. As a result of such a motion, and due to the resulting centrifugal forces, this star would take on the shape of a flattened ellipsoid. We now ask: *What shape will this star assume if all remaining heavenly bodies are suddenly annihilated (turned into nothing)?*
>
> These centrifugal forces are dependent only on the state of the star itself; they are totally independent of the remaining heavenly bodies. Consequently, this is our answer: These centrifugal forces and the spherical ellipsoidal form dependent on them will persist regardless of whether the remaining heavenly bodies continue to exist or suddenly disappear.[32]

The critical examinations of the foundations of classical mechanics in (Lange 1886) and (Mach 1960) correspond to the second alternative mentioned above, since both authors were guided by the intention to revise mechanics by eliminating problematic assumptions. They may be considered as attempts to provide a conceptual reinterpretation of the existing formalism of classical mechanics (possibly even including

---

31 "But a further question arises, whether this body exists—really, concretely, as the earth, the sun, and the remaining heavenly bodies do. We may answer this question, as I see it, by saying that its existence can be stated with the same right, with the same certainty, as the existence of the luminiferous ether or the electrical fluid." ("Aber es erhebt sich die weitere Frage, ob jener Körper denn eine wirkliche, concrete Existenz besitze gleich der Erde, der Sonne und den übrigen Himmelskörpern. Wir könnten, wie mir scheint, hierauf antworten, dass seine Existenz mit demselben Recht, mit derselben Sicherheit behauptet werden kann wie etwa die Existenz des Licht-Aethers oder die des elektrischen Fluidums.") (Neumann 1993 [1870], 365).

32 "Es mag hier eine Betrachtung ihre Stelle finden, welche sich leicht aufdrängt, und aus welcher deutlich hervorgeht, wie unerträglich die Widersprüche sind, welche sich einstellen, sobald man die Bewegung nicht als etwas Absolutes, sondern nur als etwas Relatives auffasst.
Nehmen wir an, dass unter den Sternen sich einer befinde, der aus *flüssiger* Materie besteht, und der—ebenso etwa wie unsere Erdkugel—in rotirender Bewegung begriffen ist um eine durch seinen Mittelpunkt gehende Axe. In Folge einer solchen Bewegung, infolge der durch sie entstehenden Centrifugal kräfte wird alsdann jener Stern die Form eines abgeplateten Ellipsoids besitzen. *Welche Form wird*—fragen wir nun—*der Stern annehmen, falls plötzlich alle übrigen Himmelskörper vernichtet (in Nichts verwandelt) würden?*
Jene Centrifugalkräfte hängen nur ab von dem Zustande des Sternes selber; sie sind völlig unabhängig von den übrigen Himmelskörpern. Folglich werden—so lautet unsere Antwort—jene Centrifugalkräfte und die durch sie bedingte ellipsoidische Gestalt ungeändert *fortbestehen*, völlig gleichgültig ob die übrigen Himmelskörper fortexistiren oder plötzlich verschwinden.") (Neumann 1993 [1870], 366, n. 8)

minor adjustments of this formalism), with no ambition to formulate a new theory or to cover new empirical ground. Lange's approach is today the less well known, precisely because his contribution was the introduction of the concept of an inertial system, a contribution that was successful in becoming part of the generally accepted conceptual interpretation of classical mechanics. Mach's widely discussed critique of the foundations of classical mechanics, on the other hand, is characterized by vacillation between more or less successful attempts to reformulate Newtonian mechanics on a clearer and leaner conceptual basis and suggestions to create a new theory. It seems plausible to assume that this ambiguity was actually not in conflict with Mach's intentions, as the principal aim of his reformulation of elements of classical mechanics was to stress and clarify the dependence of this theory on experience, and hence to open up the possibility of revising the theory if required by new empirical evidence.[33]

One of the principal targets of Mach's critique was Newton's interpretation of the bucket experiment as evidence in favor of the existence of absolute space.[34] To Newton's argument, according to which the curvature of the surface of the rotating water is a physical effect of the rotation with respect to absolute space, he objected that in our actual experience this rotation can be considered as a relative rotation, namely with respect to the fixed stars:

> Try to fix Newton's bucket and rotate the heaven of fixed stars and then prove the absence of centrifugal forces. (Mach 1960, 279)

Thus, Mach questioned the fundamental metaphysical presupposition of Newton's conclusion that physical effects of absolute space would also occur if the rotation took place in an otherwise empty universe, i.e. the presupposition that all elements of a system retain their essential properties independently of their relation to the composite system:

> Nature does not begin with elements, as we are obliged to begin with them. (Mach 1960, 287–288)

On the grounds of his different philosophical view, Mach demanded that the entire corpus of mechanics should be reformulated in terms of the motion of material bodies relative to each other. For instance, he introduced a new definition of the concept of mass based on the mutual accelerations of bodies with respect to each other. He also suggested that inertial frames of reference should be determined on the basis of the observable relative motions of bodies in the universe, e.g. by determining a frame of reference in which the average acceleration of a mass with respect to other—ide-

---

33 This seems to be the most natural explanation for Mach's rather indifferent reaction to the controversy about the purpose of his critique as observed in (Norton 1995). Compare Mach's remarks on his revised principle of inertia: "It is impossible to say whether the new expression would still represent the true condition of things if the stars were to perform rapid movements among one another. The general experience cannot be constructed from the particular case given us. We must, on the contrary, *wait* until such an experience presents itself." (Mach 1960, 289)

34 See (Mach 1960, chap. 2, sec. 6), in particular, pp. 279–284.

ally all—bodies in the universe vanishes. On the one hand, Mach's proposals for a reformulation of classical mechanics clearly presuppose the validity of classical mechanics: both his new definition of mass by mutual accelerations and his idea of introducing increasingly improved inertial frames of reference by taking into account more and more bodies, over whose relative motion an average can be taken, assume that the concept of an inertial frame makes sense exactly as understood in classical mechanics. In other words, his proposal presupposes that there is indeed such a privileged class of reference frames and that they can be realized physically with sufficient approximation.[35] However, Mach's analysis also indicated the limits of the validity of classical mechanics, in particular by explicitly relating the concept of inertial frame to the motion of cosmic masses. Thus, without changing the substance of classical mechanics, he succeeded nevertheless in making clear—by proposing an alternative formulation based on different philosophical presuppositions—that the range of application of classical mechanics may be more limited than hitherto assumed, and that the theory might have to be changed eventually, for instance, in light of growing astronomical knowledge. Only on the basis of such an increased knowledge could it then be decided whether Mach's suggestion to reformulate classical mechanics in terms of relative motions would actually amount to proposing a new theory, substantially different from Newton's.

Attempts to formulate such a new theory, even in the absence of new empirical evidence, form a third alternative reaction to the critical reevaluation of the foundations of mechanics in the second half of the nineteenth century. Whether these attempts were stimulated by Mach or not, their common starting point was the rejection of Newton's philosophical presupposition that the properties of the elements of a physical system could be ascribed to each one of them even if they existed alone in empty space. It was this presupposition that enabled Newton to infer from the nature of the inertial effects present in the bucket experiment to that of the inertial behavior of a single particle in empty space, and from there, to the physical reality of empty space. Only by introducing an entity such as "absolute space" did Newton succeed in distinguishing between the kinematical and dynamical aspects of motion. Hence, if now this presupposition had become questionable, so had the entire relation between dynamics and kinematics. In particular, the distinction between force-free motions and those explained by the action of forces had to be given a new grounding in terms of relative motions between ponderable bodies. While Mach had essentially presupposed the validity of classical mechanics and attempted to reconstruct its achievements on this new basis, it was also conceivable to start from first principles and reformulate dynamics from the beginning in terms of relative motions between ponderable bodies, possibly even without using the concept of an inertial frame in the sense of classical mechanics. Attempts in this direction of a *mechanistic generalization of the relativity principle* were first undertaken around the turn of the century by Benedict and Imman-

35 See the penetrating analysis in (Wahsner and von Borzeszkowski 1992, 324–328).

uel Friedlaender, August Föppl, and Wenzel Hofmann, then decades later by Reissner and Schrödinger, and in our days by Barbour, Bertotti and others.[36]

Physicists of at least the first generation in this genealogy were confronted with the difficulty of taking up once again many of the foundational questions of mechanics discussed centuries earlier by Galileo Galilei, René Descartes, Isaac Newton, Gottfried Wilhelm Leibniz, and Christiaan Huyghens, and they attempted to recreate mechanics essentially from scratch. Indeed, apart from the foundational role given to the concept of relative motion even in dynamics and the known laws of classical mechanics, this approach of a mechanistic generalization of the relativity principle had few heuristic clues to go on. One of these clues was directly related to the criticism of Newton's interpretation of the bucket experiment: if it is indeed true that the curvature of the rotating water in the bucket is due to an interaction between the water and the distant cosmic masses, then a similar but smaller effect should be observable if large but still manipulable terrestrial masses are brought into rotation with respect to a test body. Experiments along these lines were suggested by several of these researchers and conducted by, among others, the Friedlaender brothers and Föppl—all with a negative result.[37] Nevertheless, the theoretical efforts continued—even though they remained marginal—and eventually found additional resources and inspiration in the theory of general relativity formulated in 1915 by Einstein.

3.3 Resources and Stumbling Blocks Presented by the Tradition of Mechanics

After this discussion of the historical roots of the mechanistic generalization of the relativity principle, we can now summarize some of the principal heuristic contributions and obstacles which the tradition of mechanics presented to the development of general relativity, as in the beginning of this section for field theory. First and foremost it was the idea of abolishing the privileged status of the inertial frame, which emerged from the foundational critique of mechanics in the nineteenth century, that proved to be an essential component of both Einstein's early research concerning generalized relativity, and of the considerations surrounding the competing traditions to provide a mechanistic generalization of the relativity principle. In fact, if separable material bodies are to be the ultimate basis of reality, as they are in the approach of a mechanistic generalization of the relativity principle, each material body should be equally well suited to defining a reference frame, and therefore should enter into the laws of physics on the same level with all other bodies. The idea of abolishing the privileged status of inertial frames was associated with the interpretation of the so-called inertial forces—such as those acting on the rotating water in Newton's bucket—as aspects of a new, yet to be discovered, velocity-dependent physical interaction between masses in relative motion with respect to each other. Under the desig-

36 See, e.g., (Friedlaender and Friedlaender 1896; Föppl 1905a, 1905b; Hofmann 1904; Reissner 1914, 1915; Schrödinger 1925, Barbour and Bertotti 1977).

37 See (Friedlaender and Friedlaender 1896; Föppl 1905a).

nation of “dragging effects,” such interactions became an important theme of the later general theory of relativity; there they can be understood as a new aspect of the gravitational interaction between masses, which was unknown in Newtonian mechanics. Finally, Mach’s definition of inertial mass in terms of the accelerations that two bodies induce in each other brought the concept of inertial mass even closer to the concept of gravitational mass than their quantitative identity in classical mechanics had so far. It follows from this definition that, contrary to Newtonian mechanics, inertial mass, in contrast to gravitational mass, can no longer be considered as a property that bodies possess independently of their interaction with each other. The search for effects of the presence of other bodies on the inertial mass of a test body became a component of the heuristics guiding Einstein’s research on a generalized theory of relativity.

While these were the specific and crucial contributions of the foundational critique of mechanics, other aspects of mechanics in the nineteenth century also contained important heuristic hints and conceptual resources for the development of general relativity; these resources, however, cannot be dealt with here systematically. In particular, the introduction of laws of motion expressed in generalized coordinates, the formulation of mechanics for non-Euclidean geometry, and the attempts at an elimination of the concept of force all represent resources which could be, and in part were, exploited in the development of general relativity.[38] The study of motion constrained to curved surfaces in classical mechanics provided, for instance, the blueprints for the formulation of the geodesic law of motion as a generalization of the principle of inertia in general relativity: in both cases the essential assertion is that motion not subject to external forces follows a geodesic line.

But unlike the foundational critique of mechanics, these other aspects of the development of classical mechanics did not themselves constitute another independent research program for formulating a substantially new mechanics that could lead to a theory comparable to general relativity. Rather their heuristic contribution to formulating such a new theory became relevant only in the context of Einstein’s later attempt to solve the problem of gravitation, and only on the basis of results that lay outside their scope. For instance, Hertz’s mechanics (Hertz 1894) is a reformulation of classical mechanics in which the elimination of the concept of force requires the introduction of hypothetical invisible masses acting as constraints for the visible motions. Not only does its formalism and, in particular, its generalized geodesic law of motion bear a number of similarities with the formalism of general relativity, but also the general approach of replacing the concept of force by geometrical concepts is common to both theories.[39] Although, even in the context of classical mechanics, the concept of force can be eliminated in the specific case of the gravitational interaction without introducing Hertz’s speculative entities—merely on the basis of Galileo’s principle, that is, by realizing that all bodies move with equal speeds in a gravitational

---

38 For a historical account of these developments, see (Lützen 1993).

39 The geometrical interpretation of general relativity is, however, a largely post-1915 development.

field—a systematic exploitation of formal results such as Hertz's required not only a restriction of mechanics to the special case of gravitational interaction, but also the introduction of Minkowski's reformulation of special relativity uniting the time with the space-coordinates into one spacetime continuum. Only under these presuppositions did the formal achievements of nineteenth-century mechanics become a resource for the insight that force-free motion in a gravitational field can also be understood as geodesic motion in a non-Euclidean spacetime continuum.

Considered in hindsight, however, these contributions to the development of general relativity that were rooted in the tradition of classical mechanics also presented conceptual obstacles to its development. First of all, as in the case of field theory discussed above, there was much ambiguity in the research program of a mechanistic generalization of the relativity principle. It is impossible to assess the direction that this program would have taken by itself without the guidance of Einstein's achievement, since the general theory of relativity was formulated in 1915—long before an elaborate and more or less successful realization of this program emerged. Around 1915 it was far less advanced than the attempts to solve the problem of gravitation in the context of field theory. The papers proposing a mechanistic generalization of the relativity principle are mostly in the form of programmatic treatises. They contain few technical details, and show, even by their style, that they deal with foundational problems of mechanics as commonly discussed by Galileo and his contemporaries in early modern times. In particular, in exploring the postulated velocity-dependent interaction, the mechanistic approach had few tools comparable to the powerful methods developed in the field-theoretic context: of particular interest were the tools for coping with the interaction of electrically charged masses in motion with respect to each other. Also on the experimental level, the mechanistic generalization of the relativity principle failed to identify evidence in favor of this new interaction between moving masses. It is therefore not surprising that the followers of a mechanistic generalization of the relativity principle were few and played only a marginal role in contemporary discussions. In addition to its weaknesses as an independent program of research, the idea of a mechanistic generalization of the relativity principle included aspects that, if judged from the perspective of the accomplished theory of general relativity, were both stimulating and misleading: while the ideal of a theory in which all physical aspects of space are derived from the relations between separable material bodies was an essential motivation for the search for a general theory of relativity, it turned out to be incompatible with its outcome since in general relativity the gravitational field has an existence in its own right, which cannot be reduced to the effects of matter in motion.

3.4 The Example of Benedict and Immanuel Friedlaender

The opportunities and difficulties presented by a mechanistic generalization of the relativity principle are best illustrated by the contribution of the Friedlaender brothers. Their philosophical starting point is the critique of the concept of motion of a single body in an otherwise empty space, on which, as we have seen, Newton's argument for absolute space was founded:

> Now think (if you can) of a progressive motion of a *single* body in a universal space that is otherwise imagined to be totally empty; how can the motion be detected, i.e. distinguished from rest? *By nothing* we should think; indeed the whole idea of such an absolute, progressive motion is meaningless.[40]

Like other critics of Newtonian mechanics, Immanuel and Benedict Friedlaender question the meaning of inertial frames and postulate a new velocity-dependent interaction between moving masses. But unlike other representatives of a mechanistic generalization of the relativity principle, they explicitly link this new interaction to gravitation:

> If this phenomenon was verifiable, this would be the incentive for a reformulation of mechanics and at the same time a further insight would have been gained into the nature of gravity, since these phenomena must be due to the actions at a distance of masses, and here in particular to the dependence of these actions at a distance on relative rotations.[41]

How far they went in anticipating the relation between gravitation and inertia as understood in general relativity becomes clear from speculations presented towards the end of their paper:

> It is also apparent that according to our conception the motions of the bodies of the solar system could be seen as pure *inertial motions*, whereas according to the usual view the inertial motion, or rather its permanent gravitationally modified tendency, would strive to produce a rectilinear-tangential motion.[42]

And, in another suggestive passage:

> But it seems to me that the correct formulation of the law of inertia will be found only when *relative inertia* as an effect of masses on each other, and gravity, which is after all also an effect of masses on each other, are reduced to a *unified law*.[43]

At a first glance, the insight formulated by the Friedlaenders into the relation between velocity-dependent inertial forces and gravitation seems to contradict the claim that a

---

40 "Nun denke man sich aber, (wenn man kann,) eine fortschreitende Bewegung eines *einzigen* Körpers in dem als sonst völlig leer gedachten Weltenraume; woran wäre die Bewegung bemerklich, d.h. von Ruhe unterscheidbar? An *Nichts* sollten wir meine; ja, die ganze Vorstellung einer solchen absoluten, fortschreitenden Bewegung ist sinnleer." The first part of their jointly published booklet, pp. 5–17, is by Immanuel Friedlaender and the second part, pp. 18–35, by Benedict Friedlaender. (Friedlaender and Friedlaender 1896, 20)

41 "War diese Erscheinung nachzuweisen, so war der Anstoß zu einer Umformung der Mechanik gegeben und zugleich ein weiterer Ausblick in das Wesen der Gravitation gewonnen, da es sich ja dabei nur um Fernwirkungen von Massen und zwar hier der Abhängigkeit dieser Fernwirkungen von relativen Rotationen handeln kann." (Friedlaender and Friedlaender 1896, 15)

42 "Es ist auch leicht ersichtlich, daß nach unsrer Auffassung die Bewegungen der Körper des Sonnensystems als reine *Beharrungsbewegungen* angesehen werden könnten; während nach der üblichen Anschauung die Beharrungsbewegung, oder vielmehr deren fortwährend durch die Gravitation abgeänderte Tendenz eine geradlinig-tangentiale Bewegung hervorzurufen bestrebt wäre." (Friedlaender and Friedlaender 1896, 33)

mechanistic generalization of the relativity principle did not possess tools comparable to those used in the electromagnetic tradition to treat the interaction of charged masses in motion with respect to one another. However, a footnote to the same passage makes it clear that the source of this insight into a possible relation between gravity and inertia actually is the *combination* of the introduction of velocity-dependent forces by the mechanistic generalization of the relativity principle and of the treatment of velocity-dependent forces in the electromagnetic tradition:

> For this it would be very desirable to resolve the question whether Weber's law applies to gravity, as well as the question concerning gravity's speed of propagation.[44]

The reference is to Wilhelm Weber's fundamental law for the force between electric point charges, which is a generalization of Coulomb's law for the electrostatic force, taking into account also the motion of the charges. By including velocity-dependent terms, Weber's law represents an attempt to cover electrodynamic interactions too, while maintaining the form of an action-at-a-distance law, that is, of a direct interaction between the point charges without an intervening medium. In other words, the Friedlaenders established a connection between their foundational critique of mechanics and the contemporary discussions about an electromagnetic theory of gravitation.[45]

By the time their paper was published, however, action-at-a-distance laws such as Weber's had been largely superseded by the field-theoretic approach to electromagnetism taken by Maxwell, Hertz, and others, who assumed a propagation of the electromagnetic force by an intervening medium, the aether.[46] The Friedlaenders themselves seem to have entertained considerations along these lines, without, however, drawing any technical consequences from them:

> No mind thinking scientifically could ever have permanently and seriously believed in unmediated action at a distance; the apparent action at a distance can be nothing other than the result of the action of forces that are transmitted in some way by the medium being situated between the two gravitating bodies.[47]

---

43 "Mir will aber scheinen, daß die richtige Fassung des Gesetzes der Trägheit erst dann gefunden ist, wenn die *relative Trägheit* als eine Wirkung von Massen auf einander und die Gravitation, die ja auch eine Wirkung von Massen auf einander ist, auf ein *einheitliches Gesetz* zurückgeführt sein werden." (Friedlaender and Friedlaender 1896, 17)

44 "Es wäre dazu sehr zu wünschen, daß die Frage, ob das Webersche Gesetz auf die Gravitation anzuwenden ist, sowie die nach der Fortpflanzungsgeschwindigkeit der Schwerkraft gelöst würden." (Friedlaender and Friedlaender 1896, 17)

45 Hints to such a connection are also found in other authors, even if they are less explicit; see, e.g., (Föppl 1905b, 386–394; Mach 1960, 296), with reference to the Friedlaender brothers and Föppl. For a discussion of Mach's position, see (Wolters 1987), in particular, pp. 37–70.

46 For the role of Weber's law in the later tradition of generally relativistic mechanics, see (Assis 1989, 1995; see also Barbour 1992, 145).

47 "An die unvermittelte *Fernwirkung* kann kein naturwissenschaftlich denkender Kopf jemals andauernd und ernstlich geglaubt haben; die scheinbare *Fernkraft* kann nichts anderes sein, als das Resultat von Kraftwirkungen, die durch das zwischen beiden gravitirenden Körpern befindliche Medium in irgend einer Art vermittelt werden." (Friedlaender and Friedlaender 1896, 19)

But whether in the field-theoretic or in the action-at-a-distance form, it was the tools of the electromagnetic tradition of classical physics which allowed the Friedlaenders to establish the link between the new understanding of inertia and gravitation. It is therefore not surprising that they treat the dragging effects of masses in relative motion to each other in analogy with electromagnetic induction:

> ... only in order to indicate the extent to which the problem of motion that we have raised and hypothetically solved here is related to that of the nature of gravity, but at the same time that comes rather close to the known effects of electric forces, will the following parallel be pointed out: a body that approaches a second one or moves away from it would be without influence on the latter as long as the velocity of approach (to be taken either with a positive or a negative sign) remains unchanged; any change of this velocity would entail the above-demonstrated [dragging] effect.
>
> As is well known, the presence of a current in a conductor is not sufficient for the generation of induction effects, either the magnitude of the current or the distance must vary; in our case the change of distance, i.e. the motion, would not suffice for the generation of the attractive or repulsive effects, but rather the velocity itself has to change.[48]

3.5 The Historical Horizon Before Einstein's Contribution

In summary, we have identified and discussed two entirely different strategies—both pursued at the time when Einstein began to work seriously on a relativistic theory of gravitation—for dealing with important conceptual issues at the foundations of mechanics and gravitation theory. The field theoretic approach to the problem of gravitation was, around this time, mainly stimulated by the incompatibility between Newton's theory of gravitation and the special theory of relativity, while the starting point of the mechanistic generalization of the relativity principle was a philosophical critique of the foundations of Newtonian mechanics based on newly established branches of classical physics. Their relation can be understood in the context of the two principal competing worldviews of classical physics around the turn of the century, the electromagnetic worldview and the mechanical worldview. In particular, these worldviews determined the different conceptual resources from which the two strategies drew rather exclusively, those of field theory and of classical mechanics respectively. Whereas the mechanistic generalization of the relativity principle

---

48 "… und nur, um anzudeuten, in wie fern das hier angeregte und hypothetisch gelöste Bewegungsproblem mit demjenigen des Wesens der Gravitation zusammenhängt, sich zu gleicher Zeit aber den bekannten Wirkungsweisen elektrischer Kräfte einigermaßen nähert, sei auf folgende Parallele hingewiesen: Ein Körper, der sich einem zweiten nähert oder von ihm entfernt, würde ohne Einfluß auf diesen sein, solange die positiv oder negativ zu nehmende Annäherungsgeschwindigkeit unverändert bleibt; jede Aenderung der Geschwindigkeit hingegen würde die vorher gezeigte Wirkung ausüben.
Das Vorhandensein eines Stromes in einem Leiter genügt bekanntlich zur Erzeugung von Induktionswirkungen nicht, es muß entweder die Stromstärke oder die Entfernung wechseln; in unserem Falle würde nun zur Erzeugung der anziehenden oder abstoßenden Wirkungen auch die Entfernungsänderung, d.h. die Bewegung nicht ausreichen, es muß vielmehr die Geschwindigkeit selbst sich ändern." (Friedlaender and Friedlaender 1896, 30)

remained in the margin of contemporary physics, the field theoretic approach to gravitation, at least for a while, played a larger part in contemporary discussions, and both strategies were pursued in ignorance of one another.

The two strategies encountered problems which, in hindsight, can be recognized as being closely related to each other. On a general level, the difficulties of the two strategies were in an inverse relation to each other: those following the field theoretic approach were confronted with the problem of reconstructing on a new conceptual basis the shared knowledge accumulated in classical mechanics, e.g. the insight into the equality of gravitational and inertial mass. The followers of a mechanistic generalization of the relativity principle, on the other hand, had to face the task of keeping up in their terms with the immense contribution of field theory to the progress of physics in the nineteenth century, a formidable challenge even for current attempts to pursue the tradition of the mechanistic generalization of the relativity principle. But on the specific level of the gravitational and inertial interactions of masses, the problems faced by the two approaches are better characterized as being complementary to each other: on the basis of concise theoretical considerations, the electromagnetic approach to the problem of gravitation required the existence of a velocity-dependent gravitational interaction in analogy to electromagnetic induction, for which there was, however, little, if any, experimental evidence; the mechanistic generalization of the relativity principle, on the other hand, postulated a new velocity-dependent interaction between inertial masses in order to explain well-known observations such as the curvature of the water's surface in Newton's bucket experiment, but failed to develop a theoretical framework for its systematic treatment. Since the two traditions remained isolated from each other—with the remarkable but inconsequential exception of the Friedlaender brothers—their complementary strengths were not exploited before Einstein's contribution.

4. MACH'S PRINCIPLE: BETWEEN A MECHANISTIC GENERALIZATION OF THE RELATIVITY PRINCIPLE AND A FIELD THEORY OF GRAVITATION

4.1 The Emergence of a Link Between Einstein's Research on Gravitation and Mach's Critique of Mechanics in 1907

The problems of a field theory of gravitation, from which Einstein had started in 1907, pointed in two ways to Mach's critique of Newton's mechanics, namely, to his redefinition of the concept of mass and to his rejection of absolute space as a foundation for the understanding of inertial motion. As discussed in the previous section, the concept of inertial mass and the concept of absolute space were in fact connected through Newton's assumption that the essential properties of the elements of a system are independent of these elements' part in the larger (composite) system. The rejection of this assumption shattered both Newton's distinction between inertial and gravitational mass as essential and non-essential properties of a body respectively, and his demonstration of absolute space. Einstein had been familiar with Mach's critique of Newton's mechanics since his student days[49] and probably reread the corresponding chapters of the *Mechanik* after his first attack on the problem of gravitation in 1907.[50]

The physical asymmetry between inertial and gravitational mass, which, as perceived by Einstein in 1907, was at the heart of the conflict between a special relativistic theory of gravitation and classical mechanics, may have directed his attention to their more general asymmetry in Newtonian mechanics. According to Newtonian mechanics, inertial mass is a property that can also be ascribed to a single body in an otherwise empty universe, whereas gravitational mass can only be conceived as a property of a system of bodies. Mach's analysis of the concept of inertial mass can be considered as an attempt to remove just this asymmetry, at least on the level of an operational definition of inertial mass. According to this definition, inertial mass is determined, as we have seen, on the basis of the mutual accelerations within a system of bodies, i.e. not as the independent property of a single body. Although Mach's intention was probably only to give a more concise account of classical mechanics without changing its content, nevertheless his definition makes it clear that, in principle, the interaction between two masses, and hence their magnitude, may depend on the presence of other masses in the world (recall that the inertial frame within which the accelerations are measured is, according to Mach's critique of absolute space, determined by the distribution of masses in the universe). In any case, according to Mach, inertial mass and gravitational mass both depend upon interactions between bodies. This lends additional strength to Einstein's conclusion that the equivalence of inertial and gravitational mass in classical mechanics points to a deeper conceptual unity that is to be preserved also in a new theory of gravitation.

Einstein's introduction of the principle of equivalence, which expresses the equality of inertial and gravitational mass independent of the specific laws of motion of classical mechanics, indicated a connection to Mach's critical discussion of Newton's purported demonstration of absolute space. The successful use of a uniformly accelerated frame of reference to describe the behavior of bodies falling in a constant gravitational field must naturally have raised questions about the relation between arbitrarily accelerated reference frames and more general gravitational fields. For Einstein, such questions pointed in particular to the problem of the privileged role of inertial frames in classical mechanics, as he confirms in the recollection already quoted in the first section:

---

49 For an early reference to Mach, see Einstein to Mileva Marić, 10 September 1899 (CPAE 1, Doc. 54; Renn and Schulmann 1992, 14, 85). For later recollections mentioning Mach, see (Einstein 1933, 1954b, 1992).

50 For contemporary evidence of Einstein's rereading, see p. 58 of Einstein's Scratch Notebook 1910–1914? (CPAE 3, 592, app. A), where Einstein wrote the title of the crucial sec. 6 of chap. 2 of Mach's *Mechanics* (Mach 1960); pp. 7–8 of Einstein's *Lecture Notes for an Introductory Course on Mechanics* at the University of Zurich, Winter semester 1909/1910, (CPAE 3, 15–16, discussed in more detail below); and the discussion of Mach's ideas in a notebook on Einstein's Course on Analytical Mechanics, Winter semester 1912/13, by Walter Dällenbach, (for a brief description, see (CPAE 4, app. A).

> So, if one considers pervasive gravitational fields, not *a priori* restricted by spatial boundary conditions, physically possible, then the concept of 'inertial system' becomes completely empty. The concept of 'acceleration relative to space' then loses all meaning and with it the principle of inertia along with the paradox of Mach.[51]

In other words, the appearance of accelerated frames of reference in an argument concerning gravitation made it possible to relate to each other two theoretical traditions which had until then led essentially separate existences, the tradition of a field theory of gravitation in the sense of electrodynamics and the tradition of foundational critique of mechanics in the sense of what is called here "mechanistic generalization of the relativity principle." In the previous section, we have seen that the idea of including accelerated frames of reference on an equal footing with inertial systems was as alien to the tradition of field theory as the idea of a field theory of gravitation was to the tradition of the mechanistic generalization of the relativity principle.

Now, however, Mach's critical examination of the privileged role of inertial frames in classical mechanics provided Einstein with the context for considering his introduction of an accelerated frame of reference in the equivalence principle argument, not only as a technical trick to deal with a specific aspect of the problem of formulating a field theory of gravitation, but as a hint to the solution of a foundational problem of classical mechanics. But while Mach's critique justified the consideration of arbitrary frames of reference as a basis for the description of physical processes, and hence the extension of the equivalence principle argument to include more general accelerated frames, such as the rotating frame of Newton's bucket,[52] it did not provide Einstein with the conceptual tools for dealing with the strange effects encountered in such frames. The tradition of field theory, in the context of which he had first approached the problem of gravitation, offered him, on the other hand, just the conceptual tools that allowed him to interpret the inertial forces in accelerated frames of reference as aspects of a more general notion of a gravito-inertial field, in the same sense that electromagnetic field theory makes it possible to conceive induction as an aspect of a more general notion of an electric field.

In other words, instead of attempting to resolve Mach's paradox of the privileged role of inertial frames in the context of a revised version of classical mechanics, as did the adherents of a mechanistic generalization of the relativity principle, Einstein was now able to address this foundational problem of mechanics in the context of a field theory of gravitation in which inertial forces could be understood as an aspect of

---

51 "Wenn man also das Verhalten der Körper inbezug auf das letztere Bezugssystem als durch ein "wirkliches" (nicht nur scheinbares) Gravitationsfelder als möglich betrachtet, so wird der Begriff des Inertialsystems völlig leer. Der Begriff "Beschleunigung gegenüber dem Raume" verliert dann jede Bedeutung und damit auch das Trägheitsprinzip samt dem Mach'schen Paradoxon." (Einstein 1992, 62–63)

52 For the particular role of rotating frames in motivating this generalization, compare Einstein's later remark concerning an objection against the privileged role of inertial frames in classical mechanics and in special relativity: "The objection is of importance more especially when the state of motion of the reference-body is of such a nature that it does not require any external agency for its maintenance, e.g. in the case when the reference body is rotating uniformly." (Einstein 1961, 72)

a unified gravito-inertial field. By establishing a "missing link" between the traditions of a mechanistic generalization of the relativity principle and field theory, he had found the key to the problems which appeared to be unsolvable within each of the two traditions taken separately. Where the followers of a field theory of gravitation searched in vain for an empirical clue which could have guided them beyond "Coulomb's law" of static gravitation (i.e. Newton's law) to a gravitational dynamics, Einstein succeeded with the help of Mach's critique in recognizing in the inertial effects of a rotating system, such as Newton's bucket, the case of a stationary gravitational field caused by moving masses. He interpreted this case as a gravitational analogue to a magnetostatic field in electrodynamics which can also be conceived as being caused by moving (in this case: electrical) masses.[53] And vice versa, where the adherents of a mechanistic generalization of the relativity principle searched in vain for new effects that could reveal more about the mysterious interaction between distant masses in relative motion with respect to each other, which in the only case known to them was responsible for the curvature of the water's surface in Newton's bucket, Einstein had no qualms about identifying this force as a dynamical aspect of universal gravitation and thus relating the unknown force to a well-explored domain of classical physics. In summary, Einstein's experiences with a field theory of gravitation and his familiarity with the foundational problems of mechanics had set the stage for his reception of whatever these two traditions had to offer for his program to build a relativistic theory of gravity that was also to be a theory of general relativity. What had previously seemed to be mutually exclusive approaches, to some extent now became, from his perspective, complementary.

4.2 Hints at a Machian Theory of Mechanics in Einstein's Research on Gravitation Between 1907 and 1912

The following is a brief account of those features of Einstein's heuristics that reflect the complementary influence of the two traditions in the sense outlined above. While there is no direct contemporary evidence for the role of Mach's critique of mechanics on Einstein's 1907 formulation of what later became known as the equivalence principle, such an influence very likely forms the background for Einstein's reaction to the problems of a relativistic theory of gravitation.[54] Beyond shaping this reaction and opening the perspective towards a generalization of relativity theory, Mach's influence on the further development of this theory remained secondary, even when Einstein began to elaborate his original insight into the equivalence principle in papers published in 1911 and 1912.[55] The principal reason for this secondary status is that, in

---

53 See Einstein to Paul Ehrenfest, 20 June 1912 (CPAE 5, Doc. 409), discussed below.

54 See, in particular, (Einstein 1954b) for evidence that Einstein's perspective was indeed shaped at a very early stage by Mach's critique of mechanics. For a discussion of the relation between equivalence principle and Mach's interpretation of the bucket experiment, see "Classical Physics in Disarray …" (in vol. 1 of this series).

55 See, in particular, (Einstein 1911, 1912a, 1912b, 1912c).

this period, he drew mainly on the resources of field theory with the aim of constructing a field equation—analogous to the classical field equation for Newton's gravitational field—for the static gravitational field of his elevator thought experiment.[56]

Nevertheless, between 1907 and 1912 Einstein seems also to have collected hints pointing at a Machian theory of mechanics. For instance, he made use of Mach's analysis of the conceptual foundations of mechanics in preparing a course on classical mechanics at the University of Zurich for the winter semester 1909/1910[57] and referred to it in connection with his research on gravitation in correspondence to Ernst Mach of the same period.[58] At about the same time, he wrote the following in a letter to a friend:

> I am just now lecturing on the foundations of that poor, dead mechanics, which is so beautiful. What will its successor look like? With that question I torment myself ceaselessly.[59]

In the notes Einstein prepared for his lecture course he introduces Mach's definition of mass.[60] He emphasized the close relation between gravitational and inertial mass, following from the independence of gravity from material properties:

> The fact that the force of gravity is independent of the material demonstrates a close kinship between inertial mass on the one hand and gravitational action on the other hand.[61]

The dependence of inertial mass on the entire system of bodies in the universe implicit in Mach's definition of mass made it conceivable for Einstein that the inertial mass of a given body may also be a function of the system of other bodies, which varies with their distribution around the given body.[62] In (Einstein 1912c), he partially confirmed this conclusion by calculating the effect on the inertial mass of a body due to the presence of a massive spherical shell around it; the paper also deals with the effect on this body by an accelerated motion of the spherical shell. This paper, dedicated to Einstein's theory of the static gravitational field, is not only the first paper in

56 For an extensive discussion, see "Classical Physics in Disarray …" (in vol. 1 of this series).

57 See Einstein's *Lecture Notes for an Introductory Course on Mechanics* at the University of Zurich, Winter semester 1909/1910 in (CPAE 3).

58 See Einstein to Ernst Mach, 9 August 1909 (CPAE 5, Doc. 174, 204) and Einstein to Ernst Mach, 17 August 1909 (CPAE 5, Doc. 175, 205).

59 "Ich lese gerade die Fundamente der armen gestorbenen Mechanik, die so schön ist. Wie wird ihre Nachfolgerin aussehen? Damit plage ich mich unaufhörlich." Einstein to Heinrich Zangger, 15 November 1911 (CPAE 5, Doc. 305, 349).

60 See pp. 7–8 of Einstein's *Lecture Notes for an Introductory Course on Mechanics* at the University of Zurich, Winter semester 1909/1910 (CPAE 3, 15–16).

61 "Die Thatsache, dass die Kraft der Schwere vom Material unabhängig ist, zeigt eine nahe Verwandschaft zwischen träger Masse einerseits und Gravitationswirkung andererseits." See p. 15 of Einstein's *Lecture Notes for an Introductory Course on Mechanics* at the University of Zurich, Winter semester 1909/1910 (CPAE 3, 21; my translation)

62 This is in disagreement with the claim expressed in (Barbour 1992, 135), that Einstein was not justified in maintaining that he was a following a stimulation by Mach in considering a dependence of inertial mass on the presence of other masses in the universe.

which he publicly mentions Mach's critique as a heuristic motivation behind his search for a generalized theory of relativity, but it also carries a title expressing the translation of this heuristics into the language of field theory: "Is there a gravitational effect which is analogous to electrodynamic induction?"

In 1912 Mach's critique gained a new importance for Einstein's work on gravitation for yet another reason. After convincing himself that he had found a more or less satisfactory theory of the static gravitational field, he turned to what he considered to be the next simple case, the stationary field represented by the inertial forces in a rotating frame. In other words, after exhausting, at least for the time being, the heuristic potential of the "elevator," he now turned to that of the "bucket." His contemporary correspondence confirms that he considered this case both from the perspective of field theory and from that of the mechanistic generalization of the relativity principle. In a letter to Ehrenfest from 20 June 1912, with reference to his theory of the static gravitational field and to the generalization necessary to cope with situations such as that of a rotating ring, he wrote:

> In the theory of electricity my case corresponds to the electrostatic field, while the more general static case would further include the analogue of the static magnetic field. I am not yet that far.[63]

In a letter to Besso dated 26 March 1912, Einstein remarked—probably referring to the same topic, i.e., the treatment of the inertial forces in a rotating frame as generalized gravitational effects in a frame at rest—in the spirit of Mach's remark on Newton's bucket: "You see that I am still far from being able to conceive rotation as rest!".[64] Not only Einstein's publications and correspondence but also his private research notes document the influence of both traditions—electrodynamics and mechanics—on the terminology in which he expressed the heuristics of his theory. Thus, we can exclude the possibility that his choice of words was merely a matter of making himself understood by his audience.[65]

4.3 Einstein's Machian Heuristics in his Research on a Relativistic Theory of Gravitation between 1912 and 1915

Einstein found it difficult to accomplish the transition from his treatment of the static special case to a more general theory that included the dynamical aspects of the gravitational field. In the summer of 1912, however, he attained the insight into the crucial role of non-Euclidean geometry for formulating the gravitational field theory he

63 "Mein Fall entspricht in der Elektrizitätstheorie dem elektrostatischen Felde, wogegen der allgemeine[r]e statische Fall noch das Analogon des statischen Magnetfeldes mit einschliessen würde." Einstein to Paul Ehrenfest, 20 June 1912, (CPAE 5, Doc. 409, 486).

64 "Du siehst, dass ich noch weit davon entfernt bin, die Drehung als Ruhe auffassen zu können." Einstein to Michele Besso, 26 March 1912, (CPAE 5, Doc. 377, 436).

65 See, in particular, Einstein's comments on his calculation of the effect of rotation and linear acceleration of a massive shell on a test particle in his and Michele Besso's May 1913 "Manuscript on the Motion of the Perihelion of Mercury" in (CPAE 4, Doc. 14).

searched for, an insight which, in spite of the many difficulties still to be resolved, paved the way for the final theory of general relativity published in 1915. This insight provides an important example of the fruitfulness of the combined heuristics of "elevator" (i.e., Einstein's equivalence principle) and "bucket" (i.e., Newton and Mach's bucket in Einstein's interpretation).[66] The heuristics of the bucket, i.e. the Machian idea to consider water in a bucket as constituting a frame at rest, first provided the qualitative insight into a possible role of non-Euclidean geometry in a rotating frame of reference.[67] The heuristics of the elevator, i.e. the elaboration of the theory of the static gravitational field, then prepared, in combination with Minkowski's four-dimensional formalism, the technical environment for the concrete application of this insight to the problem of gravitation. The crucial link between the general idea and this technical environment was provided by Gaussian surface theory, which made it possible to interpret the equation of motion suggested by the formalism of the static theory as the geodesic equation of a non-Euclidean geometry. It was only possible, however, to exploit the formal similarity between the two equations because of the deeper conceptual similarity between the problem of motion in a gravitational field and the problem of inertial motion in an accelerated frame of reference, as suggested by Einstein's Machian interpretation of inertia. This conceptual similarity may have helped Einstein to think of Gaussian surface theory in the first place, as he had been familiar since his student days with the relation in classical mechanics between motion constrained to a surface without external forces—which also can be conceived of as generalized inertial motion—and the geodesic equation in Gaussian surface theory.[68]

But even after Einstein recognized that the gravitational potential of his static theory could be interpreted as a component of the metric tensor in four-dimensional geometry, he would nevertheless have been, at least in principle, in the same situation as those searching for a dynamic theory of the gravitational field starting from Newton's theory as the only known special case. It was his "Machian" insight that the inertial effects in accelerated frames can be considered as an aspect of a more general gravito-inertial field that provided him with an entire class of examples supporting the relation between the equation of motion, metric tensor, and gravito-inertial field, which emerged from the generalization of the static theory. In fact, Einstein could

66 Compare also Einstein's Kyoto Lecture (Ishiwara 1971, 78–88).

67 This was first stressed in (Stachel 1989). For a more extensive reconstruction, see "Classical Physics in Disarray …" and "The First Two Acts" (both in vol. 1 of this series).

68 This is suggested by the similarity between a page in the Zurich Notebook by Einstein (p. 41R of "Research Notes on a Generalized Theory of Relativity," dated ca. August 1912, in (CPAE 4, Doc. 10) and p. 88 of the student notes on Geiser's lecture course on infinitesimal geometry, taken by Einstein's friend Marcel Grossmann in 1898 (Eidgenössische Technische Hochschule, Zürich, Bibliothek, Hs 421: 16); for Einstein's attendance of this course in the summer semester 1898, see (CPAE 1, 366); for his later recollections on the significance of this course for his work on general relativity, see (Ishiwara 1971, 78–88). The connection between Einstein's research notes and Grossmann's student notes was identified by Tilman Sauer; see also (Castagnetti et al. 1994) and "Commentary" (in vol. 2 of this series).

easily show that the inertial motion of a particle in an arbitrarily accelerated frame of reference can be described by the same type of equation as that published in May 1912 for a static gravitational field,[69] involving not just one variable but indeed a 4-by-4 metric tensor.

The introduction of the metric tensor provided Einstein with the framework for capturing the resources of the two traditions, field theory and the mechanistic generalization of the relativity principle, as well as those of the mathematical tradition established, among others, by Riemann and Christoffel. The tradition of field theory suggested, for instance, that—following the model of Poisson's equation for the gravitational potential in classical physics—some second-order differential operator was to be applied to the metric tensor in order to yield the left-hand-side of a gravitational field equation.[70] It therefore comes as no surprise to find that the first entries in the Zurich Notebook, in which Einstein tackled the problem of gravitation, reflect his attempt to translate the field equation of the theory for the static field into a second order differential equation for the metric tensor.[71] However, the construction of a satisfactory field equation for the gravitational field was an incredibly difficult task that would demand Einstein's attention for the following three years. In his search, he could rely on the tradition of the mechanistic generalization of the relativity principle which offered him concrete examples for metric tensors to be covered by the new theory, such as the metric tensor for Minkowski space described from the perspective of a rotating frame of coordinates. The inertial forces arising in such a rotating frame are well known from classical physics and could hence serve as criteria for the theory to be constructed.

In the course of Einstein's long search for a gravitational field equation, he continued to exploit the heuristics of the "elevator" and "bucket" in particular, and the traditions of field theory and mechanics in general, in order to build up a considerable "machinery" of formalisms, mathematical techniques, and conceptual insights. This machinery eventually developed a dynamics of its own and led to a "conceptual drift"; i.e., to results that were not always compatible with Einstein's heuristic starting points, whether they were rooted in field theory or in the mechanistic generalization of the relativity principle.[72]

One of the first indications of such a conceptual drift was a revision published in 1912 of the theory of the static gravitational field, which conflicted with the "heuristics of the elevator," and also with an expectation raised by traditional field theory.[73]

---

69 See (Einstein 1912b).

70 For more extensive discussion, see "Pathways out of Classical Physics …" (in vol. 1 of this series).

71 See p. 39L of "Research Notes on a Generalized Theory of Relativity" (dated ca. August 1912) in (CPAE 4, Doc. 10). See also "Facsimile and Transcription of the Zurich Notebook" (in vol. 1 of this series).

72 See (Elkana 1970).

73 For Einstein's first theory, see (Einstein 1912a), for his second, revised theory, see (Einstein 1912b). For historical discussion, see "The First Two Acts" and "Pathways out of Classical Physics …" (both in vol. 1 of this series).

The revision of Einstein's static theory became necessary after he found out that his theory violated the principle of the equality of action and reaction. The non-linearity of the revised field equation turned out to be incompatible with the equivalence principle as formulated by Einstein in 1907. The homogeneous static gravitational field which he replaced by a uniformly accelerated frame of reference was simply no longer a solution of the revised, non-linear field equation.[74] In other words, after the revision, the theory of the static gravitational field contradicted its own heuristic starting point. Consequently, Einstein had to restrict the principle of equivalence. But from the perspective of our present discussion, the most significant implication of this episode was that the gravitational field had entered the scene in its own right, on a par with the material bodies acting as its source. Hence it became, at least in principle, conceivable that non-trivial gravito-inertial fields could exist without being caused by material bodies. Einstein, however, remained hesitant to accept this conclusion—which is in obvious contradiction with the Machian requirement that all inertial effects are due to ponderable masses—even after he had formulated the final theory of general relativity.

During Einstein's work on his generalized theory of relativity in the years 1912 and 1913, the "heuristics of the bucket" did not fare much better. In Einstein's research notes from this period, one encounters again and again the metric tensor representing the Minkowski space as seen from a rotating frame of reference.[75] However, it remained unclear for some time whether or not the field equation of the preliminary theory of gravitation, which Einstein published in 1913 with his mathematician friend Marcel Grossmann (Einstein and Grossmann 1913), satisfied this requirement of incorporating the Machian bucket. Einstein's eventual discovery that the "*Entwurf*" theory conflicted with this expectation was an important motive for discarding this theory and for beginning anew the search for a theory that promised to fulfill his original goals.[76] In this way, the "heuristics of the bucket" once more played a crucial role in the genesis of the general theory of relativity.

4.4 Attempts at a Machian Interpretation of General Relativity in the Period 1915–1917

After Einstein had formulated his theory in 1915, the tension between his original heuristics and the implications of the new theory remained unresolved; this tension continued to characterize the further development of the theory until at least 1930. Initially, one motive behind Einstein's emphasis on epistemological arguments based

74 For an extensive evaluation of Einstein's principle of equivalence, see (Norton 1989a), and, in particular, for the present discussion, p. 18.

75 See, e.g., pp. 42R, 43L, 11L, 12L, 12R, 24R, and 25R of "Research Notes on a Generalized Theory of Relativity" (dated ca. August 1912) in (CPAE 4, Doc. 10). See also "Facsimile and Transcription of the Zurich Notebook" (in vol. 1 of this series).

76 See, e.g., Einstein to Arnold Sommerfeld, 28 November 1915, (CPAE 8, Doc. 153). For historical discussion, see (Janssen 1999) and "What Did Einstein Know …" (in vol. 2 of this series).

on the relation between the new theory and its Machian heuristics may have been his desire to make his achievement acceptable to the scientific community. In fact, an important element of the empirical confirmation of the theory was only supplied when the eclipse expedition of 1919 confirmed the bending of light in a gravitational field. In 1913 Einstein had written to Mach that the agreement which he had found between the consequences of his then preliminary theory of gravitation and Mach's critique of Newtonian mechanics was practically the only argument he had in its favor.[77] Also in his early publications on the final theory he continued to insist on its epistemological advantages, which provided additional support for its claim of superiority with regard to competing theories.[78]

But Einstein's insistent pursuit of the Machian aspects of general relativity in the early years after its formulation was determined less by tactical motives than by the perceived need for a physical interpretation of the technical features of the new theory in light of his original heuristics. For instance, Einstein soon realized that, as a rule, the field equation determines the gravitational field only if, in addition to the matter distribution, boundary conditions are specified. This technical feature of the theory had to be brought together with his intention to realize a generally relativistic theory and his Machian hopes of explaining inertial behavior by material bodies only.[79] For some time in 1916 and early 1917, he attempted to formulate boundary conditions that would somehow comply with his original intentions.[80] He searched, for example, for boundary conditions in which the components of the metric tensor take on degenerate values since he assumed that a singular metric tensor would remain invariant under general coordinate transformations, and thus make it possible to maintain the requirement of general covariance even at the boundary region of spacetime. He also searched for a way to define a region outside the system of masses that constituted the physical universe in which a test body would possess no inertia, so that he might then be able to claim that inertia is indeed created by the physical system circumscribed by this empty boundary region.[81] In the course of these attempts, the expectation that general relativity was to provide a Machian explanation of inertia began to be silently transformed from a requirement concerning the nature of the theory to a criterion to be applied to specific solutions of the theory. Since Minkowski's flat spacetime, with inertial properties familiar from classical mechanics and special relativity, was a solution to the vacuum field equations of general rela-

---

77 See Einstein to Ernst Mach, second half of December 1913 (CPAE 5, 583–584).

78 See, e.g., (Einstein 1916a, 771–772).

79 See Einstein to Lorentz, 23 January 1915 (CPAE 8, Doc. 47), and the extensive historical discussion in (Kerszberg 1989a, 1989b), as well as in (Hoefer 1994), on which the following account is based. See also the introduction to (CPAE 8); "The Einstein-de Sitter-Weyl-Klein Debate," (CPAE 8, 351–357).

80 See, e.g., Einstein to Michele Besso, 14 May 1916, (CPAE 8, Doc. 219).

81 See Einstein to de Sitter, 4 November 1916 (CPAE 8, Doc. 273) and Einstein to Gustav Mie, 8 February 1918 (Doc. 460).

tivity, it simply could not be true in general that in this theory inertial effects are explained by the presence of matter.

After Einstein's failure to find a satisfactory treatment of the supposed Machian properties of general relativity along the road of singular boundary conditions, in (Einstein 1917) he advanced a completely different proposal for dealing with the cosmological aspects of the theory. He introduced a spacetime that satisfied all his expectations concerning the constitution of the universe, including the explanation of its inertial properties by the masses acting as sources of the gravitational field, but at the price of modifying the field equations to which this spacetime was a solution. As Einstein's cosmological paper of 1917 has been discussed a number of times, it may suffice to briefly emphasize its place in the development of the tensions between Einstein's Machian heuristics and the implications of the new theory.[82] The solution to the field equations—modified by the introduction of a "cosmological constant"—which Einstein considered in 1917 describes a spatially closed, static universe with a uniform matter distribution. It therefore entirely avoided the problem of specifying appropriate boundary conditions and, at the same time, was believed by him to correspond to a more or less realistic picture of the universe as known at that time. In general, though, Einstein tended to neglect the relation between the new theory and astronomy, as well as the exploration of the properties of the solutions to its field equations. In contrast, Willem de Sitter—at the time Einstein's principal opponent in the discussion of the allegedly Machian features of general relativity— repeatedly emphasized the astronomical consequences of the various solutions to the field equations.[83] In any case, Einstein not only hoped that his radical step of modifying the field equations of general relativity allowed him to find at least one acceptable solution to the field equations, but he also assumed that he would succeed in excluding altogether empty space solutions in which inertial properties are present in spite of the absence of matter.[84] It was therefore an unpleasant surprise—which he found difficult to digest and at first attempted to refute—when de Sitter demonstrated shortly after the publication of Einstein's paper that even the modified field equations allow such an empty space solution.[85] In 1918 Einstein published a critical note on de Sitter's solution in which he wrote:

> If de Sitter's solution were valid everywhere, then it would be thereby shown that the purpose which I pursued with the introduction of the λ-term [the cosmological con-

---

82 See, in particular, (Hoefer 1994) for a detailed discussion of this paper from the point of view of Mach's influence on Einstein.

83 See, e.g., Einstein to Willem de Sitter, before 12 March 1917, (CPAE 8, Doc. 311), where he referred to his solution as a "Luftschloss," (castle in the air) having the principal purpose of showing that his theory is free of contradictions. See also Einstein to Michele Besso, 14 May 1916, for the Machian motivations of Einstein's construction. For a historical account of the controversy between Einstein and de Sitter on the implementation of Machian ideas and cosmological considerations in general relativity, see (Kerszberg 1989a, 1989b).

84 See Einstein to de Sitter, 24 March 1917, (CPAE 8, Doc. 317).

85 See de Sitter to Einstein, 20 March 1917, (CPAE 8, Doc. 313).

> stant–J R.] has not been reached. In my opinion the general theory of relativity only forms a satisfactory system if according to it the physical qualities of space are *completely* determined by matter alone. Hence no $g_{\mu\nu}$ -field must be possible, i.e., no space-time-continuum, without matter that generates it.[86]

4.5 The Introduction of "Mach's principle" in 1918

The increasing tension between Einstein's original intentions and the ongoing exploration of the consequences of the new theory was accompanied by attempts to rephrase the criteria of what it meant to satisfy the philosophical requirements corresponding to the heuristics that had guided the discovery of the theory. Characteristically, Einstein (Einstein 1918a, 241–242) introduced and defined the very term "Mach's principle" in the context of a controversy over whether or not the general theory of relativity in fact represented a realization of his intention to implement a generalization of the relativity principle of classical mechanics and special relativity. His paper of 1918 was a response to the argument by Kretschmann that the general covariance of the field equations of general relativity does not imply such a generalization of the relativity principle, but can be considered as a mathematical property only. Einstein argued that he had so far not sufficiently distinguished between two principles, which he now introduced as the principle of relativity and Mach's principle.[87]

The first principle, as defined by Einstein, states that the only physically meaningful content of a relativistic theory are coincidences of physical events at points of space and time. Since the occurrence of these point coincidences is independent of whether they are described in one or the other coordinate frame, their most appropriate description is by a generally covariant theory. This principle had, of course, not been the starting point of Einstein's search for a generally relativistic theory of gravitation, but rather constitutes a result of his reflection on complications encountered in a long, but eventually successful, search for such a theory.[88] For our purposes here, it is particularly remarkable that this formulation of the principle of relativity no longer appeals to the intuition of a world of isolated bodies distributed in an otherwise empty space whose physical interactions should depend only on their relative distances, velocities, etc. As we have seen, this intuition was characteristic of the mechanistic generalization of the relativity principle, and was at the root of Einstein's search for a generalized theory of relativity.

---

86 "Bestände die De Sittersche Lösung überall zu Recht, so würde damit gezeigt sein, daß der durch die Einführung des "λ- Gleides" von mir beabsichtigte Zweck nicht erreicht wäre. Nach meiner Meinung bildet die allgemeine Relativitätstheorie nämlich nur dann ein befriedigendes System, wenn nach ihr die physikalischen Qualitäten des Raumes allein durch die Materie vollständig bestimmt werden. Es darf also kein $g_{\mu\nu}$ -Feld, d. h. dein Raum-Zeit-Kontinuum, möglich sein ohne Materie, welche es erzeugt." (Einstein 1918b, 271)

87 For historical discussions of this paper and its context, on which the following account is based, see (Norton 1992a, in particular, pp. 299–301, 1993, 806–809; Rynasiewicz 1999).

88 See the various discussions of Einstein's "hole argument" in the recent literature, e.g. in (Norton 1989b, sec. 5). See also the discussion in "Untying the Knot …" (in vol. 2 of this series).

This original intuition in fact included Mach's suggestion to conceive of inertial effects as the result of physical interactions between the bodies of such a world. Now, however, the causal link between inertial effects and matter suggested by Mach's critical analysis of the foundations of classical mechanics needed to be reinterpreted in light of the newly developed formalism of general relativity. According to this formalism, inertial effects are described by the metric tensor representing the gravito-inertial field, while matter is described by the energy-momentum tensor representing the source term of the field equations for the gravitational field. It was therefore natural for Einstein to translate the supposed causal nexus between inertial forces and matter into the requirement that the gravitational field be entirely determined by the energy-momentum tensor. It is this requirement which he chose in 1918 to call "Mach's principle."[89] Certainly, this was not a mathematically concise criterion to determine whether general relativity as a theory, or, as particular solutions of the theory, do or do not satisfy Mach's principle. Two aspects of this principle are nevertheless clear. First, the translation of Mach's original suggestion into the language of general relativity transferred it from the conceptual context of mechanics into that of field theory, as both terms in Einstein's 1918 definition of Mach's principle—the gravitational field and the energy-momentum tensor—are basically field theoretical concepts. Second, it is obvious from the context of this definition, discussed above, that whatever was precisely intended, Einstein considered empty space solutions of the gravitational field equations as a violation of this principle.

4.6 The Conceptual Drift from Mach's Principle to "Mach's Aether" (1918–1920)

Ironically, both these aspects of Einstein's first explicit definition of Mach's principle in his writings contributed to the preparation for its eventual rejection. As a first step towards this rejection, de Sitter established that not only Einstein's gravitational field equations of 1915, but even the equations modified by the introduction of the cosmological constant, admit empty space solutions. As a consequence, Mach's principle now definitely took on the role of a selection principle for solutions to the field equations. It seems that one interpretative reaction by Einstein to this serious defeat of his principle was to extend the field theoretical interpretation of general relativity at the expense of the emphasis on the mechanical roots of his original heuristics. By 1920, the 1918 attempt to define Mach's principle in terms of the conceptual building blocks of his theory had been complemented by the introduction of a "Machian aether" as a means of capturing its conceptual implications.[90] In a lecture given in Leiden, Einstein (1920) exploited the time-honored concept of an aether, to which

89 "Mach's principle: The *G*-field is *completely* determined by the masses of bodies. Since mass and energy are identical in accordance with the results of the special theory of relativity and the energy is described formally by means of the symmetric energy tensor ($T_{\mu\nu}$), the *G*-field is conditioned and determined (*bedingt und bestimmt*) by the energy tensor of the matter." See (Einstein 1918a, 241–242), quoted from (Barbour 1992, 138).

Lorentz had given the definitive form in the realm of electrodynamics, in order to explain the new concept of space which had emerged with general relativity. He now directly turned against Mach's interpretation of inertial effects as caused by cosmic masses, because this interpretation presupposed an action at a distance, a notion incompatible with both field theory and relativity theory. Instead, contrary to his original heuristics, Einstein (1920, 11–12) associated these inertial effects with the nature of space, which he now conceived as being equipped with physical qualities, and which he hence appropriately called aether. Contrary to Lorentz's aether, however, Mach's aether, which Einstein thought of as being represented by the metric tensor, was supposed not only to condition but also to be conditioned, at least in part, by matter. This capacity of being influenced by the presence of matter was, apparently, the last resort for the Machian idea of the generation of inertial effects by the interaction of material bodies in Einstein's conceptual framework.

For the time being, however, two aspects of the relation between matter and space remained open problems: with space—under the name Machian aether—taking on the role of an independent physical reality, the question presented itself of whether matter had not lost all claims to primacy in a causal nexus between space and matter. In his Leiden lecture, Einstein (1920, 14) noted that it was possible to imagine a space without an electromagnetic field, but not one without a gravitational field, as space is only constituted by the latter; he concluded that matter, which for him was represented by the electromagnetic field, appears to be only a secondary phenomenon of space. In (Einstein 1919), he had made an attempt at a derivation of the properties of matter from the gravitational and the electrodynamic field, an attempt which he considered as still being unsatisfactory but which, for him, constituted the beginning of a new line of research in the tradition of the electrodynamic—or field theoretical—worldview. This kind of research program held out the possibility not only of reintroducing the concept of an aether in order to represent the physical qualities of space, but also of providing a theoretical construction of matter as an aspect of this aether. The other outstanding question concerning the relation between matter and space, which was left unclarified even after Einstein's introduction of a Machian aether, was the astronomical problem of the distribution of masses and of the large-scale spatial structure of the universe. Both questions, the theoretical as well as the empirical, turned out to be significant, not only for Einstein's further exploration of general relativity, but indirectly for the fate of Mach's principle as well.

---

90 For historical discussions, see (Illy 1989; Kox 1989; Kostro 1992, 2000; Renn 2003). Probably under the influence of Lorentz, Einstein had begun to reconsider the concept of aether already in 1916. On 17 June of this year he had written to H. A. Lorentz: "I admit that the general theory of relativity is closer to the aether hypothesis than the special theory." ("Ich gebe Ihnen zu, dass die allgemeine Relativitätstheorie der Aetherhypothese näher liegt als die spezielle Relativitätstheorie." (CPAE 8, Doc 226; English translation in Kostro 1992, 262.) At that time, however, as the same letter suggests, Einstein took it for granted that the aether is entirely determined by material processes. The transition to the aether concept as explained in the following seems to be complete by the end of 1919, see Einstein to H. A. Lorentz, 15 November 1919, (CPAE 9, Doc. 165).

4.7 Mach's Principle: From the Back Burner to Lost in Space (1920–1932)

The program of interpreting general relativity along the lines of Mach's philosophical critique of classical mechanics ceased to play a significant role in Einstein's research after 1920. In addition to the difficulty of implementing Machian criteria in the elaboration of the theory, his exploration during the twenties of the heuristic potential that general relativity offered for the formulation of a unified theory of gravitation and electrodynamics was probably responsible for this shift of interest.[91] As this heuristic potential for a further unification of physics was associated with the field theoretic aspects of general relativity, the relation of the theory to the foundational problems of mechanics naturally faded into the background. Nevertheless, on several occasions during his ongoing research on a unified theory of gravitation and electromagnetism, Einstein hoped that he could link the program of a unified field theory with a satisfactory solution of the cosmological problem in the sense of his Machian heuristics. In 1919, for example, he emphasized that his new theory had the advantage that the cosmological constant appears in the fundamental equations as a constant of integration, and no longer as a universal constant peculiar to the fundamental law; he made a point of showing that again a spherical world results from his new equations (Einstein 1919, 353; 1923b, 36). An additional reason for not completely rejecting Mach's principle may have been Einstein's awareness, in a period which saw the triumph of quantum mechanics, that, after all, the corpuscular foundation of physics and not the field theoretic might prevail in the end; fields would then indeed have to be conceived as epiphenomena of matter, like the gravitational field according to Mach's principle.[92]

There was also a rather mundane reason why Mach's principle did not figure prominently in Einstein's publications of this period, while not being entirely dismissed by him. More than its definition in 1918, its association with the cosmological model of 1917 had brought the principle to an end point of its theoretical development, to a point where the question of whether or not Mach's principle could be implemented in general relativity had become a question of its confirmation or refutation by astronomical data. In 1921 Einstein remarked, with reference to the possibility of explaining inertia in the context of his cosmological model:

> Experience alone can finally decide which of the two possibilities is realized in nature. (Einstein 1922a, 42)[93]

In any case, for the time being, he remained convinced that astronomical research on the large systems of fixed stars would bear a model of the universe compatible with his Machian expectations. In 1921 he also wrote:

---

91 See (Pais 1982, 287–288); see also the extensive discussion in (Vizgin 1994).

92 See, in particular, Einstein's views expressed in connection with theoretical and experimental studies of radiation in this period, for example: "It is thus proven with certainty that the wave field has no real existence, and that the Bohr emission is an instantaneous process in the true sense." Einstein to Max Born, 30 December 1921, my translation; see also the discussion in (Vizgin 1994, 176).

93 The German original is (Einstein 1921a).

> A final question has reference to the cosmological problem. Is inertia to be traced to mutual action with distant masses? And connected with the latter: Is the spatial extent of the universe finite? It is here that my opinion differs from that of Eddington. With Mach, I feel that an affirmative answer is imperative, but for the time being nothing can be proved. (Einstein 1921b, 784)[94]

In other words, although in the period between 1920 and 1930 Einstein invested his hopes and his research efforts mainly in the creation of a unified field theory, he nevertheless kept Mach's principle on the back burner as long as it was not contradicted by astronomical data.

Einstein's firm conviction made him sceptical with respect to the possibility of alternative cosmological models. In 1922 he criticized, among other proposals, Friedmann's paper on solutions to the original field equations which correspond to a dynamical universe.[95] He mistakenly identified a calculational error in Friedmann's solution, which he had viewed with suspicion from the beginning. In another paper of the same year (Einstein 1922b, 437), he explicitly criticized a cosmological model for its incompatibility with "Mach's postulate." In 1923, however, Einstein recognized that he had committed an error in rejecting Friedmann's dynamical solutions. He published a retraction (Einstein 1923c) of his earlier criticism and henceforth no longer expected an astronomical confirmation of his Machian cosmology with the same certainty as before. The change of Einstein's attitude is apparent from a comparison between the published retraction of his criticism with a manuscript version that has been preserved. In the manuscript version Einstein wrote:

> It follows that the field equations, besides the static solution, permit dynamic (that is, varying with the time coordinate) spherically symmetric solutions for the spatial structure, to which a physical significance can hardly be ascribed.

In the published paper, on the other hand, Einstein omitted the last half-sentence.[96] In another paper of the same year, Einstein referred with scepticism to "Mach's postulate" and to the modification of the field equations that it required, because the introduction of the cosmological constant was not founded on experience. He concluded:

> For this reason the suggested solution of the 'cosmological problem' can, for the time being, not be entirely satisfactory.[97]

Nevertheless, until the end of the twenties Einstein did not give up hope that Mach's principle could be maintained as a feature of a cosmologically plausible solution of the field equations of general relativity. When he discussed the "aether" of general

---

94 Einstein's astronomical views in this period were strongly under the influence of his Machian belief, see, e.g., (Einstein 1922b, 436).

95 See (Einstein 1922d); for Einstein's criticism of other proposals, see (Einstein 1922b, 1922c).

96 This has been noted by John Stachel. For the translation of the passage, see also (Stachel 1986, 244).

97 "Aus diesem Grunde kann die angedeutete Lösung des kosmologischen Problems einstweilen nicht völlig befriedigen." (Einstein 1923a, 8; my translation.) He also modified an earlier version of an attempt to formulate a unified field theory by omitting the cosmological constant, see (Vizgin 1994, 192–193).

relativity in (Einstein 1924, 90), he added that it is determined by ponderable masses and that this determination is complete if the world is spatially finite and closed in itself. In the same paper, he dealt both with the possibility that a unification of gravitation and electrodynamics can be achieved by field theory and with the possibility that an understanding of the quantum problem can be achieved without field theoretical components.[98] As suggested above, it is conceivable that this ambivalence as to which of the foundational concepts—field or corpuscle—would eventually prevail may have reinforced the role of Mach's principle in Einstein's thinking. In 1926, he discussed the cosmological implications of general relativity in line with his earlier arguments in favor of a finite static universe.[99] In 1929 he wrote:

> Nothing certain is known of what the properties of the space-time continuum may be as a whole. Through the general theory of relativity, however, the view that the continuum is infinite in its time-like extent but finite in its space-like extent has gained in probability. (Einstein 1929, 107)

Around 1930, however, things began to change. Primarily driven by his strong intellectual engagement in the program to formulate a unified field theory, Einstein expressed himself even more definitely than earlier in favor of a causal primacy of space in relation to matter—in sharp contrast to his original Machian heuristics. He would still ask the question:

> If I imagine all bodies completely removed, does empty space still remain?

and suggest a negative answer. But now this question was not so much intended to refer to the constitution of the universe, but was rather an epistemological inquiry regarding the construction of the concept of space:

> But how is the concept of space itself constructed? If I imagine all bodies completely removed, does empty space still remain? Or is even this concept to be made dependent on the concept of body? Yes, certainly, I reply.[100]

While in the sequel of the paper, Einstein develops at length his reasons for suggesting a *cognitive* primacy of the concept of physical object with respect to the concept of space, he concludes his discussion of the state of research on the foundations of physics with the following remark:

> Space, brought to light by the corporeal object, made a physical reality by Newton, has in the last few decades swallowed ether and time and seems about to swallow also the field and the corpuscles, so that it remains as the sole medium of reality.[101]

---

98 See (Einstein 1924), in particular, pp. 92–93.

99 See (Einstein 1926–1927) and, for historical discussion, (Vizgin 1994, 212–213).

100 "Wie kommt aber der Raumbegriff selbst zustande? Wenn ich die Körper allesamt weggenommen denke, bleibt doch wohl der leere Raum über? Soll etwa auch dieser vom Körperbegriff abhängig gemacht werden? Nach meiner Überzeugung ganz gewiß!" (Einstein 1930a, 180)

101 "Der Raum, ans Licht gebracht durch das körperliche Objekt, zur physikalischen Realität erhoben durch Newton, hat in den letzten Jahrzehnten den Äther und die Zeit verschlungen und scheint im Begriffe zu sein, auch das Feld und die Korpuskeln zu verschlingen, so daß er als alleiniger Träger der Realität übrig bleibt." (Einstein 1930a, 184)

In a lecture given in 1930 Einstein formulated his view even more drastically:

> The strange conclusion to which we have come is this—that now it appears that space will have to be regarded as a primary thing and that matter is derived from it, so to speak, as a secondary result. Space is now turning around and eating up matter. We have always regarded matter as a primary thing and space as a secondary result. Space is now having its revenge, so to speak, and is eating up matter. (Einstein 1930b, 610)

In the course of his work on unified field theory, assisted by his epistemological reflections, Einstein had come a long way from believing that a successful implementation of Mach's principle would entail a synthesis of physics in which the concept of matter would play a primary and the concept of space a secondary role. Nevertheless, as the development of Mach's principle in his thinking had become so closely associated with his cosmological ideas, the question of Mach's principle remained open precisely to the extent that the decision about Einstein's static universe was left open by observational cosmology. In the period between 1917 and 1930, a prevailing problem debated by researchers in this field was whether de Sitter's or Einstein's static universe is a better model of reality, while the question of expanding universes, raised by Friedmann in 1922 and by Lemaître in 1927, largely remained outside the horizon of observational cosmology.[102] The range of theoretical alternatives taken into account by contemporary researchers testifies to the persistent role of Einstein's Machian interpretation of general relativity for cosmology, even if this interpretation gradually became a mere connotation of one of the cosmological alternatives rather than being the primary issue.

With the stage thus set for an observational decision on Mach's principle, a definitive blow to Einstein's belief in it came with the accumulation of astronomical evidence in favor of an expanding universe, the decisive contribution being Hubble's work published in 1929.[103] Einstein became familiar with these results early in 1931 during a stay at the California Institute of Technology. As is suggested by an entry in his travel diary from 3 January 1931, Richard Tolman convinced Einstein that his doubts about the correctness of Tolman's arguments in favor of the role of nonstatic models for a solution of the cosmological problem were not justified.[104] In March of the same year Einstein wrote to his friend Michele Besso:

> The Mount Wilson Observatory people are excellent. They have recently found that the spiral nebulae are spatially approximately uniformly distributed and show a strong Doppler effect proportional to their distance, which follows without constraint from the theory of relativity (without cosmological constant).[105]

---

102 See (Ellis 1989, 379–380).

103 For historical discussion, see (Ellis 1989, 376–378).

104 "Doubts about correctness of Tolman's work on cosmological problem. Tolman, however, was in the right." Quoted from (Stachel 1986, n. 53, 249); for a discussion of Tolman's contribution, see (Ellis 1989, 379–380).

105 Einstein to Michele Besso, 1 March 1931, quoted from (Stachel 1986, 245).

Almost immediately after his return to Berlin, Einstein published a paper (Einstein 1931b) on the cosmological problem in which he stated that the results of Hubble had made his assumption of a static universe untenable. As it was even easier for general relativity to account for Hubble's results than for a static universe—because no modification of the field equations by the introduction of a cosmological constant was required—his earlier static solution now appeared unlikely to Einstein, given the empirical evidence (Einstein 1931b, 5).

In a lecture given in October of 1931, he still mentioned his static solution in connection with the implementation of Mach's ideas in general relativity, but, in spite of the numerous remaining difficulties of the dynamical conception of the universe, he now had definitely given up his belief in a Machian world (Einstein 1932). In 1932, in a joint paper with de Sitter—his main antagonist in the earlier controversy about a Machian explanation of inertia—Einstein himself presented an expanding universe solution to the unmodified field equations. In this paper, the original Machian motivation for Einstein's static universe solution is no longer even mentioned:

> Historically the term containing the "cosmological constant" λ was introduced into the field equations in order to enable us to account theoretically for the existence of a finite mean density in a static universe. It now appears that in the dynamical case this end can be reached without the introduction of λ. (Einstein and de Sitter 1932, 213)

In other words, in the course of the evolution of Einstein's cosmological views, from his adherence to a static world to his acceptance of an expanding universe, Mach's principle had simply disappeared.

4.8 Reflections in the Aftermath of Mach's Principle

Although Einstein continued to acknowledge the role of Mach's critique of classical mechanics for the emergence of general relativity even after 1930, one can nevertheless notice a tendency to reinterpret even the heuristics which had originally guided his formulation of the theory. In his later accounts of the conceptual foundations of general relativity, he appealed to the field concepts in order to point out those weaknesses of classical physics that he had discussed earlier in the spirit of Mach's critique of mechanics. He emphasized, for instance, that it was due to the introduction of the field concept that the standpoint of considering space and time as independent realities had been surmounted.[106] Or he argued (Einstein 1961, app. V, 153) that the principle of equivalence, which had originally motivated the extension of the relativity principle beyond the special theory of relativity, already demonstrated the existence of the field as a reality in its own right, that is, independent of matter, since for the field experienced by an observer in an accelerated frame of reference the question of sources does not arise.

When the occasion presented itself, Einstein also became quite explicit about his rejection of his earlier Machian heuristics. In a letter to Felix Pirani, for instance, he

106 See, e.g., (Einstein 1961, app. V, 144).

explains with reference to Mach's principle, as he himself had earlier defined it, that he no longer finds it plausible that matter represented by the energy-momentum tensor could completely determine the gravitational field, since the specification of the energy-momentum tensor itself already presupposes knowledge of the metric field. In the same letter Einstein explicitly revokes Mach's principle:

> In my view one should no longer speak of Mach's principle at all. It dates back to the time in which one thought that the "ponderable bodies" are the only physically real entities and that all elements of the theory which are not completely determined by them should be avoided. (I am well aware of the fact that I myself was long influenced by this *idée fixe*.)[107]

He similarly explains in his *Autobiographical Notes*:

> Mach conjectures that in a truly reasonable theory inertia would have to depend upon the interaction of the masses, precisely as was true for Newton's other forces, a conception that for a long time I considered in principle the correct one. It presupposes implicitly, however, that the basic theory should be of the general type of Newton's mechanics: masses and their interaction as the original concepts. Such an attempt at a resolution does not fit into a consistent field theory, as will be immediately recognized. (Einstein 1992, 27)

In summary, this section has shown that Mach's critique of classical mechanics was a crucial element in the heuristics guiding Einstein's way to the formulation of the general theory of relativity. It played this role as one among several aspects of the tradition of classical physics and was, just as many of these other elements, eventually superseded by the development of general relativity. At the outset, it opened up Einstein's perspective towards a generalization of the relativity principle and towards an explanation of inertial effects, and hence of the physical properties of space, by material bodies. By conceptualizing inertial forces as an interaction of bodies in motion, it provided a decisive complement to the prospect of a dynamical theory of gravitation, which was suggested by the conceptual tradition of field theory, but lacked an empirical substantiation that could offer orientation among a variety of possible research directions. The results which Einstein accumulated in the course of his search for a general theory of relativity enforced several adjustments and reformulations of his original heuristics. Eventually, it became impossible for him to bring the progress of general relativity into agreement with these heuristics.[108] Here we have seen that this is the case for those aspects of his heuristics which were founded on the stimulation received from Mach's critique of mechanics. It turns out, however, that the incompatibility between the conceptual framework that shaped Einstein's original heuristics and that which emerged from the final theory can be demonstrated more generally.[109]

---

107 Einstein to Felix Pirani, 2 February 1954 (my translation). (Einstein Archives: call number 17 - 447.00.)

108 See also the systematic discussions of the relation between Mach's principle and the progress of general relativity in (Goenner 1970, 1981; Torretti 1983, 199–201).

109 For extensive discussion, see "Pathways out of Classical Physics …" (in vol. 1 of this series)

5. EINSTEIN'S PHILOSOPHICAL PERSPECTIVE ON THE FOUNDATIONAL PROBLEMS OF PHYSICS

5.1 Einstein's Route to General Relativity: Between Physics and Philosophy

The account given in the previous section of the impact of Mach's critique on the development of general relativity seems to provide a strong case in point for an influence of philosophy on physics. Einstein himself confirms in many contemporary comments as well as in later recollections that he conceived the emergence of general relativity at least in part as a response to Mach's analysis of the foundations of classical mechanics.[110] He indeed continued his search for such a response even when more simple alternative approaches to the problem of gravitation seemed to be available and when only epistemological arguments could motivate the continuation of his search for a generalization of the relativity principle.[111] The fact that also the followers of a mechanistic generalization of the relativity principle could refer to Mach's analysis as the philosophical background of their enterprise, however, raises some doubts as to how significant the contribution of philosophy to Einstein's particular approach actually was. The starting point of Einstein's revision of the foundations of mechanics was in fact, as we have seen, in contrast to that of these "Machians," not a general philosophical concern but a concrete problem which he encountered in the course of his research. It was not that the principle of equivalence had been formulated as a consequence of Einstein's search for a generalization of the principle of relativity, but vice versa, that the introduction of the equivalence principle in the context of a problem of "normal science" had opened up the perspective towards the foundational questions of mechanics. In a recollection from 1919 Einstein laconically states with reference to the emergence of general relativity:

> The epistemological urge begins only in 1907.[112]

There is, however, a crucial distinction between the reaction of Einstein and that of the adherents of a mechanistic generalization of the relativity principle to Mach's critique of the foundations of mechanics. In Einstein's view, the primary philosophical force of Mach's critique was directed against precisely what seemed to be for the "Machian relativists"—at least within the context of this particular research problem—an undisputed presupposition of their thinking, namely the mechanistic ontol-

---

110 For contemporary evidence, see, e.g., Einstein's correspondence with Mach quoted above, for a later recollection, see, e.g., (Einstein 1954a, 133–134). The significance of Mach's philosophical critique of mechanics for Einstein is exhaustively treated in (Wolters 1987, chap. 1).

111 See (Einstein 1914, 344), where Einstein comments on Nordström's competing theory.

112 "Das erkenntnistheoretische Bedürfnis beginnt erst 1907." Einstein to Paul Ehrenfest, 4 December 1919 (CPAE 9, Doc. 189 - my translation). See also (Wheeler 1979, 188), for a later recollection by Einstein, according to which he recognized the significance of the equality of inertial and gravitational mass only as a consequence of his failure to formulate a special relativistic theory of gravitation. For a different interpretation, see (Barbour 1992, 130, 133).

ogy on the basis of which they attempted a generalization of the relativity principle. Einstein himself later remembered that questioning the self-evident character of the concepts of mechanics was one of the principal effects that Mach's philosophy had upon him:

> We must not be surprised, therefore, that, so to speak, all physicists of the last century saw in classical mechanics a firm and definitive foundation for all physics, yes, indeed, for all natural science, and that they never grew tired in their attempts to base Maxwell's theory of electromagnetism, which, in the meantime, was slowly beginning to win out, upon mechanics as well. ... It was Ernst Mach who, in his *History of Mechanics*, shook this dogmatic faith; this book exercised a profound influence upon me in this regard while I was a student.[113]

In other words, in contrast to those physicists whose reception of Mach's critique of mechanics was shaped only by the perspective of this one subdiscipline of physics, Einstein read Mach as a philosopher and understood the central philosophical intention behind Mach's historical and critical account of mechanics, which was directed against the special status which mechanics had had for a long time among the subdisciplines of physics.

We may therefore ask whether it was this philosophical sensibility with regard to the epistemological character of some of the foundational problems of classical physics which protected Einstein from the temptation to seek a solution to these problems within one of the subdisciplines of classical physics, as for instance, the adherents of a mechanistic generalization of the relativity principle. Indeed, there can be little doubt that Einstein's thinking was characterized by such a sensibility, which was heightened by his reading of philosophical authors such as Kant, Hume, Helmholtz, Mach, and Poincaré.[114] But it seems doubtful, on the other hand, whether philosophical scepticism with regard to false pretensions of a conceptual system is sufficient to overcome its limitations. At the turn of the century, philosophical critics of the privileged status of classical mechanics, often associated as it was with the pretension of an *a priori* character, may themselves serve as counter examples. Neither Mach nor Poincaré built the foundations of a new mechanics upon the basis of their respective epistemological critiques, let alone the foundations of a new conceptual framework for all of physics. As late as 1910, Poincaré—who had emphasized the conventional character of scientific concepts—was nevertheless of the opinion that the principles of mechanics may turn out to be victorious in their struggle with the new theory of relativity, and

113 "Wir dürfen uns daher nicht wundern, dass sozusagen alle Physiker des letzten Jahrhunderts in der klassischen Mechanik eine feste und endgültige Grundlage der ganzen Physik, ja der ganzen Naturwissenschaft sahen, und dass sie nicht müde wurden zu versuchen, auch die indessen langsam sich durchsetzende Maxwell'sche Theorie des Elektromagnetismus auf die Mechanik zu gründen. […] Ernst Mach war es, der in seiner *Geschichte der Mechanik* an diesem dogmatischen Glauben rüttelte; dies Buch hat gerade in dieser Beziehung einen tiefen Einfluss auf mich als Student ausgeübt." (Einstein 1992, 20–21) See also (Holton 1986, chap. 7, 237–277, in particular, p. 241; 1988, chap. 4, 77–104; Wolters 1987, chap. 1, 20–36).

114 For a list of some of Einstein's philosophical readings, see the introduction to (CPAE 2).

that it was hence unjustified to prematurely abandon these principles.[115] Mach (1960, 295–296) had left it open, as we have seen, that new empirical evidence may require a modification of the principles of mechanics. Contrary to Einstein, he speculated that an electromagnetic worldview may provide a new universal conceptual framework for the entire body of physics, while his own contributions to such a unity remained rather on the level of a metatheoretical reflection on science.[116] Einstein, in any case, was convinced that one should not attempt to identify Mach's crucial contribution in what can also be found in the works of Bacon, Hume, Mill, Kirchhoff, Hertz, or Helmholtz, but rather in his concrete analysis of scientific content.[117]

In addition, it can be historically documented that Einstein's scepticism, with respect to the competing worldviews based on mechanics, electrodynamics, or thermodynamics, was rooted in his precise knowledge of their respective scientific failings and not only in his epistemological awareness.[118] Shortly after the turn of the century, when the electromagnetic worldview still appealed to many physicists as the most promising starting point for a new conceptual foundation of physics, Einstein had already recognized the devastating consequences which the discovery by Planck of the law of heat radiation had for classical electrodynamics and hence for the conceptual backbone of a worldview based on traditional field theory. But does this observation not imply that Einstein's philosophical perspective on the foundational problems of physics simply dissolves, in the end, into technical competence in physics? This conclusion would only be justified if one accepted the conceptual distinction between philosophy of physics and physics as accepted today, that is, as a distinction between a methodological, epistemological, or metaphysical—in any case, a metatheoretical—study of physics and the concrete occupation with its scientific problems. In order to respond to the question of the philosophical character of Einstein's perspective, we therefore have to examine briefly the historical situation of the relation between physics and philosophy in Einstein's time.

5.2 The Historical Context of Einstein's Philosophical Perspective on Physics

At the turn of the century, the separation between philosophy of science and science in the sense accepted today had long been complete. The more recent history of this separation can be understood as a consequence of the failure of traditional philosophy to integrate the natural sciences into its reflective enterprise. This failure is partly due to the explosive growth of the shared knowledge of the various disciplines, and partly

---

115 See (Poincaré 1911); see also (Cuvaj 1970, 108) for a historical discussion.

116 For an extensive discussion of Mach's attitude with respect to the electromagnetic worldview, see (Wolters 1987, 29–36). For Mach's attempt to integrate mechanics into the body of physics on the level of methodological reflections, see (Mach 1960, chap. 5).

117 See his remarks to this effect in his obituary for Mach, (Einstein 1916b, 154–155).

118 See, in particular, Einstein's own account in his *Autobiographical Notes* (Einstein 1992), in particular, pp. 42–45, which is confirmed by contemporary evidence such as Einstein's letters to Mileva Marić, see (Renn and Schulmann 1992; CPAE 1).

to the change of the cultural and political role which philosophy, and philosophy of science in particular, underwent in the nineteenth century. In German academic philosophy of the second half of the nineteenth century, for instance, neo-Kantianism, which saw itself as a critical reaction to the philosophy of German idealism, played a weighty role.[119] Its stance was that of a politically neutral epistemology which—in contrast to the natural philosophy of German idealism—often anything but politically neutral—no longer issued any prescriptions for science but just attempted to capture the epistemological and methodological structures that made scientific progress possible. Although neo-Kantianism and the tradition in philosophy of science that pursued its metatheoretical concerns took the natural sciences as their guidepost, they did not, however, offer a theoretical framework that enabled scientists to reflect upon the body of scientific knowledge in its totality, let alone to discuss the social and cultural conditions and implications of science.

On the other hand, since the middle of the nineteenth century, the intrinsic necessity of dealing with science as a social and cultural phenomenon had been approached primarily on a pragmatic level, as is witnessed by the increasing role in the development of the large-scale structure of science played by state science and education policy and the creation of funding agencies and scientific organizations (such as the Kaiser-Wilhelm-Gesellschaft in Germany). Attempts to achieve an intellectual integration of scientific knowledge, for instance in the form of a scientific worldview, remained in the shadow of this development towards a practical control of the sciences as a social system, and was only later supplemented by theoretical studies of science policy and the sociology of science.[120] As a consequence of this complex dynamics of the social and the intellectual development of science, the transfer of knowledge beyond disciplinary boundaries, and the establishment of connections between disparate branches of the body of knowledge, remained a process largely left to chance and to the initiative of the individual researcher. Only to a small degree was this process systematically furthered by the requirements of the intellectual integration of science for the purposes of education, to mention one extreme, and in the context of a few, themselves highly specialized interdisciplinary research projects, to mention the other extreme. The lack of a global intellectual synthesis of scientific knowledge was, on the other hand, only poorly compensated for by a popular scientific literature whose aim was often less the distribution and mediation of scientific knowledge than its mystification.

The lack of a systematic place in the social system of the sciences and of academic philosophy for reflection on the contents of science beyond the narrow requirements of disciplinary specialization lent a particular importance to the philosophical efforts by scientists themselves. For Einstein's intellectual development it is in fact clear that the writings of scientists such as Mach, Duhem, Poincaré, and Helmholtz

119 For this and the following, see the detailed study, (Köhnke 1986).

120 For an attempt to assess this historical situation from the point of view of a systematic historical epistemology, see (Damerow and Lefèvre 1994).

had a greater impact on his philosophical reflection on science than the works of contemporary academic philosophers, precisely because they often dealt with the philosophical implications of concrete problems at the forefront of research. Nevertheless, it would be misleading to consider Einstein's own philosophical contribution only as a continuation of the tradition of epistemological and methodological reflections by nineteenth-century philosopher-scientists. Although this view is naturally suggested by the separation of physics and philosophy as understood today, it is too restrictive to capture the peculiar way in which research in physics and philosophical reflection are intertwined in Einstein's work. In fact, Einstein's scientific contributions to many branches of physics, from thermodynamics to statistical mechanics, from the theory of relativity to quantum physics, cannot be understood without assuming the background of a scientific world picture holding together otherwise disparate chunks of knowledge. As student, Einstein already possessed an extraordinary overview of the state of physics of his time. This enabled him to recognize foundational questions of physics in problems which others preferred to see only from the point of view of their area of specialization.[121] In comparison to Einstein's perception of the entire body of physics and its conceptual incongruences, the claim of those who undertook the construction of, say, an electromagnetic world picture almost appears to be an attempt to conceal the limitations of a specialists' outlook. In any case, Einstein's perspective distinguished itself profoundly—and with significant consequences—from the mutual ignorance that characterized the field theoretical approach to the problem of gravitation and the approach of a mechanistic generalization of the relativity principle, as we have seen earlier.

5.3 Einstein and the Culture of Science

From this sketch of the historical relation between physics and philosophy, it should be clear that the roots of the scientific worldview, which shaped Einstein's perception of physics at the beginning of his career, could only have been of a highly eclectic and backward character. From what is known of his early biography, it is clear that his reading of popular scientific books, together with his exposure to the technical culture associated with the business activities of his family, played a crucial role in the early development of his scientific worldview.[122] The popular scientific books that he devoured as an adolescent combined an easily accessible and conceptually organized overview of scientific knowledge with the claim that the enterprise of science also serves as a model for the development of moral and political standards.[123] These works represented an attempt to transmit the values of democracy and of political and technological progress (which had been defeated on the political scene with

121 For a reconstruction of Einstein's discoveries of 1905 along these lines, see (Renn 1993). See also (Holton 1988, chap. 4).

122 For evidence, see (Einstein 1992), as well as the documents collected in (CPAE 1); for historical discussion, see (Damerow and Lefèvre 1994; Gregory 2000; Pyenson 1985; Renn 1993).

123 See, in particular, (Bernstein 1867–1869).

the failure of the revolution of 1848) in the medium of popular science.[124] Einstein's scientific worldview, which apparently had some of its roots in his early fascination with these popular scientific books, has indeed much in common with their image of science as a substitute for religion, with their appeal to the moral and also political ideals of science, and with their effort to achieve a conceptual unification of scientific knowledge beyond its disciplinary boundaries.[125]

The conceptual framework that formed the basis of this effort was a rather primitive combination of remnants of the old natural philosophy from the beginning of the nineteenth century, and of scientific results roughly on the level of the state of knowledge at the middle of the century. It was, however, apparently sufficient to provide the young Einstein with a global perspective on science to which he could then assimilate a broad array of detailed knowledge without committing himself to a premature specialization. In any case, during his entire scientific career he pursued the idea of a conceptual unity of physics, whose first primitive image he may have encountered in his early reading of popular scientific literature. The history of Einstein's formulation of the special theory of relativity, for instance, illustrates not only that he saw, even at the start of his career, in the conceptual diversity of mechanics and field theory a challenge to this unity of physics, but also that he was aware that neither of the two subdisciplines alone could provide the basis for a solution of this conflict. The foundation of the special theory of relativity—the principle of relativity being rooted in classical mechanics, and the principle of constancy of the speed of light in the tradition of field theory—makes it clear that the conceptual innovation represented by this theory presupposed an integration of the knowledge accumulated in these two branches of classical physics.[126]

In Einstein's reaction to the clash between classical mechanics and field theory in the case of gravitation it is now possible to recognize an intellectual attitude that was deeply rooted in his scientific worldview and shaped by his experience with the creation of the special theory of relativity.[127] The approach of a mechanistic generalization of the relativity principle had a function for the emergence of general relativity which is similar to the function mechanics had for the development of special relativity: it supported the principle of relativity with a network of arguments that went beyond the narrow scope of the specific questions under examination, whether these concerned the electrodynamics of moving bodies or the integration of Newton's theory of gravitation into a relativistic field theory. Although Einstein's perspective on the foundational problems of physics encompassed the entire range of classical phys-

---

124 The biographical background of Bernstein, the author of the book which apparently played a key role for Einstein's early intellectual development, is discussed in (Gregory 2000). For more on the relation between popular scientific literature and political developments in the nineteenth century, see (Gregory 1977); see also (Lefèvre 1990).

125 For a systematic analysis of the role of "images of science" as a mediator between science and its external influences, see (Elkana 1981). For a discussion of the religious dimensions of Einstein's scientific worldview, see (Renn 2005).

126 For further discussion, see "Classical Physics in Disarray …" (in vol. 1 of this series).

ics, there can be no doubt that it was dominated by the tension between its two major conceptual strands, field theory and mechanics. In 1931, for instance, he wrote:

> In a special branch of theoretical physics the continuous field appeared side by side with the material particle as the representative of physical reality. This dualism, though disturbing to any systematic mind, has today not yet disappeared.

He then added with specific reference to Lorentz's theory of electrons, as well as with respect to the special and general theories of relativity:

> The successful physical systems that have been set up since then represent rather a compromise between these two programmes, and it is precisely this character of compromise that stamps them as temporary and logically incomplete, even though in their separate domains they have led to great advances. (Einstein 1931a, 69–70, 72)

For Einstein, the insight into the need to overcome the dualism of matter and field was not just paying lip service to the conceptual unity of physics, but one of the principal determinants of his research program. While his perspective was broader than that of many contemporary physicists, it was also limited by this same program. The extent to which Einstein's intellectual horizon was actually circumscribed by the problem of reconciling the fundamental, conceptual conflict he perceived at the heart of classical physics can be seen in his role, up to the twenties, in the exploration of the consequences of the theory of general relativity. Contrary to other researchers who took part in this research, Einstein's interest focused almost exclusively on what might be called the "philosophical closure" of the new theory. Whether concerning boundary conditions for the gravitational field, or the exact solutions to the field equations, his interest in these emerging research topics was not guided by a program of exploring new features of the theoretical structures he had created, nor by comparing these structures with the empirical results of astronomy, but rather by the question of whether or not a deeper understanding of general relativity would reveal its agreement with the heuristics that had guided its discovery. This interest merely reflects the perspective which had accompanied Einstein's work on general relativity since its inception: he had not searched for a theoretical foundation of cosmology, but rather for a contribution to the conceptual unification of classical physics and, in particular, a synthesis of the field theoretical and mechanical aspects of gravitation.

---

127 Einstein himself compared the heuristics which motivated his search for a general theory of relativity with that guiding his formulation of special relativity: "The theory has to account for the equality of the inertial and the gravitational mass of bodies. This is only achieved if a similar relation is established between inertia and gravitation as that [which is established] by the original theory of relativity between Lorentz's electromotive force and the action of electrical field strength on an electrical mass. (Depending on the choice of the frame of reference, one is dealing with one or the other.)" ("Die Theorie muss Rechenschaft geben von der Gleichheit der trägen und schweren Masse der Körper. Dies wird nur erzeilt, wenn zwischen Trägheit und Schwerkraft eine ähnliche Beziehung hergestellt wird, wie durch die urprüngl. Relativitätstheorie zwischen Lorentz'scher elektromotorischer Kraft und Wirkung elektrischer Feldstärke auf eine elektrische Masse. (Je nach der Wahl des Bezugssystems liegt das eine oder das andere vor.)") Einstein to H. A. Lorentz, 23 January 1915 (CPAE 8, Doc. 47 - my translation).

In spite of these limitations of Einstein's perspective, and in spite of the conflict between his heuristic expectations and the conceptual implications of what he had found, it is remarkable that in the course of his work on general relativity he was nevertheless gradually able to overcome his own preconceived expectations and to adapt the interpretation of his theory to new results. This contrasts with many other cases of conceptual innovation in science, in which the crucial step of conceptual innovation takes place at a generational transition, in the transmission of knowledge from "master" to "disciple" so to speak, as was actually the case in the transformation of Lorentz's electrodynamics into Einstein's special theory of relativity.[128] Einstein's own significant contribution to the conceptual understanding of general relativity is related to the fact that, from his earliest efforts to formulate such a theory until the end of his life, he expounded unceasingly the conceptual presuppositions and consequences of his research in accounts accessible also to the non-specialist. Einstein was himself one of the great authors of popular scientific literature. With only minimal technical content, his writings made the intellectual core of his scientific problems accessible to readers. That Einstein's general accounts of the theory of relativity functioned not only to disseminate expert knowledge to the layman, and that they also formed a medium for his own reflection on the conceptual aspects of scientific problems, are facts often overlooked by philosophers of science. But the gradual adaptation of Einstein's Machian heuristics to the implications of general relativity, and its eventual definitive abandonment in the light of these implications, provide a vivid illustration of the impact these reflective accounts had on Einstein's own understanding of general relativity.

In general, it is hardly possible to overlook the significance that the effort to explain scientific knowledge to laymen had for Einstein's intellectual biography, in particular his capacity to address foundational questions beyond the limits imposed by disciplinary specialization. In Bern, as well as in Zurich, he shared his ideas with a group of friends, most of whom were not physicists. We know with certainty that Einstein was indebted to Michele Besso for a decisive inspiration which made possible the breakthrough in the formulation of the special theory of relativity.[129] Einstein also belonged to amateur science societies in Bern and in Zurich that offered an institutional framework for an exchange of ideas which transgressed the usual academic and social boundaries. Even before studying physics in Zurich, he attended an unusual high school in Aarau whose intellectual atmosphere presented no sharp demarcation between research and education, and in which he could experience the spirit of a *res publica litterarum*. Teachers, who were also scientists, such as the physicists Conrad Wüest and August Tuchschmid, or the linguist Jost Winteler, must have confirmed Einstein's conviction that science could offer the foundation for making a life, and not just an intellectual life.[130]

128 See (Damerow et al. 2004; Renn 1993).
129 See the acknowledgement in (Einstein 1905) as well as the recollection in (Ishiwara1971).

To conclude: a culture of science which includes the effort of explanation as well as the search for conceptual unity in the diversity of scientific knowledge, that is, a "culture of scientific mediation," forms an essential background for Einstein's philosophical perspective on the foundational problems of physics. The historical preconditions that made this perspective possible were already fragile at the time: evidently, neither popular scientific literature nor amateur science societies could halt the disciplinary fragmentation of scientific knowledge and the loss of possibilities for a single individual to achieve a comprehensive overview. Despite the claim by many physicists of Einstein's generation of the proximity of their field to philosophy, Einstein was in fact already part of a small minority who continually attempted to reflect upon the whole of physics and to search for its conceptual unity. The isolation in which he worked on his later attempts to create a unified field theory testify to his failure to achieve a unity of physics along these lines. But considering how much a single individual could accomplish, even on the basis of inadequate presuppositions, we can read the history of Einstein's achievements as the challenge and the encouragement to work on a culture of science that responds to the needs of today.

ACKNOWLEDGEMENTS

For permission to quote from unpublished Einstein documents I am grateful to Ze'ev Rosenkranz, former curator of the Albert Einstein Archives of the Hebrew University, Jerusalem. I would like to thank my colleagues from the Eidgenössische Technische Hochschule in Zurich—where part of the research for this paper was done—for their hospitality and friendly support, in particular Elmar Holenstein. For their careful reading of earlier versions of this text, as well as for helpful suggestions I am grateful to Met Bothner, Leo Corry, Peter Damerow, Yehuda Elkana, Gideon Freudenthal, Hubert Goenner, John Norton, Wolfgang Lefèvre, Peter McLaughlin, Fiorenza Renn, Ted Richards, Tilman Sauer, and Gereon Wolters. To Frederick Gregory, Michel Janssen, and Tilman Sauer I am particularly indebted for making preliminary versions of their papers accessible to me.

REFERENCES

Abraham, Max. 1912a. "Zur Theorie der Gravitation." *Physikalische Zeitschrift* 13: 1–4. (English translation in this volume.)

———. 1912b. "Das Elementargesetz der Gravitation." *Physikalische Zeitschrift* 13: 4–5.

———. 1915. "Neuere Gravitationstheorien." *Jahrbuch der Radioaktivität und Elektronik* 11: 470–520. (English translation in this volume.)

Assis, A. K. T. 1989. "On Mach's Principle." *Foundation of Physics Letters* 2: 301–318.

———. 1995. "Weber's Law and Mach's Principle." In (Barbour and Pfister 1995).

Barbour, Julian B. 1990. "The Part Played by Mach's Principle in the Genesis of Relativistic Cosmology." In R. Balbinot, B. Bertotti, S. Bergia, and A. Messina (eds.), *Modern Cosmology in Retrospect*. Cambridge, New York, Port Chester, Melbourne, Sydney: Cambridge University Press, 47–66.

---

130 See the documents collected in (CPAE 1; CPAE 5). For historical discussion, see (Pyenson 1985) and the introduction to (Renn and Schulmann 1992).

———. 1992. "Einstein and Mach's Principle." In J. Eisenstaedt and A. J. Kox (eds.), *Studies in the History of General Relativity,* (*Einstein Studies,* vol. 3). Boston: Birkhäuser, 125–153.

———. 1993. "The Search for True Alternatives and its Unexpected Outcome: General Relativity is Perfectly Machian." In *International Conference "Mach's Principle: From Newton's Bucket to Quantum Gravity.*" Tübingen: Preprint.

Barbour, Julian B. and Bruno Bertotti. 1977. "Gravity and Inertia in a Machian Framework." *Il Nuovo Cimento* B 38: 1–27.

Barbour, Julian B. and Herbert Pfister (eds.). 1995. *Mach's Principle: From Newton's Bucket to Quantum Gravity.* (*Einstein Studies,* vol. 6.) Boston: Birkhäuser.

Bernstein, Aaron. 1867–1869. *Naturwissenschaftliche Volksbücher.* 20 vols. Berlin: Duncker.

Blackmore, John (ed.). 1992. *Ernst Mach: A Deeper Look. Documents and New Perspectives. Boston Studies in the Philosophy of Science*, vol 143. Dordrecht: Kluwer Academic Publishers.

Boniolo, Giovanni. 1988. *Mach e Einstein*. Rome: Armando.

Borzeszkowski, Horst-Heino von and Wahsner, Renate. 1989. *Physikalischer Dualismus und dialektischer Widerspruch: Studien zum physikalischen Bewegungsbegriff*. Darmstadt: Wissenschaftliche Buchgesellschaft.

Castagnetti, Giuseppe, Peter Damerow, Werner Heinrich, Jürgen Renn, and Tilman Sauer. 1994. *Wissenschaft zwischen Grundlagenkrise und Politik. Einstein in Berlin*. Berlin: Max-Planck-Institut für Bildungsforschung. Forschungsbereich Entwicklung und Sozialisation. Arbeitsstelle Albert Einstein.

Clifford, William K. 1976. "On the Space Theory of Matter." In M. Capek (ed.), *The Concepts of Space and Time*. Dordrecht/Boston: Reidel Publishing Company, 295–296.

CPAE 1. 1987. *The Collected Papers of Albert Einstein*, vol. 1: *The Early Years 1879–1902*, ed. J. Stachel, R. Schulmann, D. Cassidy, and J. Renn. Princeton: Princeton University Press.

CPAE 2. 1989. *The Collected Papers of Albert Einstein*, vol. 2: The Swiss Years: Writings, 1900–1909, ed. J. Stachel, R. Schulmann, D. Cassidy, J. Renn, D. Howard, and A. J. Kox. Princeton: Princeton University Press.

CPAE 3. 1993. *The Collected Papers of Albert Einstein*, vol. 3: The Swiss Years: Writings, 1909–1911, ed. M. Klein, A. J. Kox, J. Renn, and R. Schulmann. Princeton: Princeton University Press.

CPAE 4. 1995. *The Collected Papers of Albert Einstein*, vol. 4: The Swiss Years: Writings, 1912–1914, ed. M. J. Klein, A. J. Kox, J. Renn, and R. Schulmann. Princeton: Princeton University Press.

CPAE 5. 1993. *The Collected Papers of Albert Einstein*, vol. 5: The Swiss Years: Correspondence, 1902–1914, ed. M. Klein, A. J. Kox, J. Renn, R. Schulmann, et al. Princeton: Princeton University Press.

CPAE 6. 1996. A. J. Kox, Martin J. Klein, and Robert Schulmann (eds.), *The Collected Papers of Albert Einstein*. Vol. 6. *The Berlin Years: Writings, 1914–1917*. Princeton: Princeton University Press.

CPAE 8. 1998. Robert Schulmann, A. J. Kox, Michel Janssen, and József Illy (eds.), *The Collected Papers of Albert Einstein,* vol. 8. *The Berlin Years: Correspondence, 1914–1918*. Princeton: Princeton University Press.

CPAE 9. 2004. Diana Kormos Buchwald, Robert Schulmann, Jósef Illy, Daniel J. Kennefick, and Tilman Sauer (eds.). *The Collected Papers of Albert Einstein, v*ol. 9: *The Berlin Years: Correspondence, January 1919–April 1920*. Princeton: Princeton University Press.

Cuvaj, Camillo. 1970. "A History of Relativity: The Role of Henri Poincaré and Paul Langevin." Ph. D., Yeshiva University.

Damerow, Peter, Gideon Freudenthal, Peter McLaughlin, and Jürgen Renn. 2004. *Exploring the Limits of Preclassical Mechanics*, 2nd ed. New York: Springer Verlag.

Damerow, Peter, and Wolfgang Lefèvre. 1981. *Rechenstein, Experiment, Sprache: historische Fallstudien zur Entstehung der exakten Wissenschaften*. Stuttgart: Klett-Cotta.

———. 1994. "Wissenssysteme im geschichtlichen Wandel." In F. Klix and H. Spada (eds.), *Enzyklopädie der Psychologie*. Band G, Themenbereich C: Theorie und Forschung - Serie II: Kognition.

Deltete, Robert J. (ed.). 2000. *The Historical Development of Energetics by Georg Helm*. (Translated from the German and with an introductory essay by Robert J. Deltete.) Dordrecht: Kluwer.

Ehrenfest, Paul. 1909. "Gleichförmige Rotation starrer Körper und Relativitätstheorie." *Physikalische Zeitschrift* 10: 918.

Einstein, Albert. 1905. "Zur Elektrodynamik bewegter Körper." *Annalen der Physik* 17: 891–921, (CPAE 2, Doc. 23).

———. 1907a. "Über die vom Relativitätsprinzip geforderte Trägheit der Energie." *Annalen der Physik* 23: 371–384, (CPAE 2, Doc. 45).

———. 1907b. "Über das Relativitätsprinzip und die aus demselben gezogenen Folgerungen." *Jahrbuch für Radioaktivität und Elektronik* 4: 411–462, (CPAE 2, Doc. 47).

———. 1911. "Über den Einfluß der Schwerkraft auf die Ausbreitung des Lichtes." *Annalen der Physik* 35: 898–908, (CPAE 3, Doc. 23).

———. 1912a. "Lichtgeschwindigkeit und Statik des Gravitationsfeldes." *Annalen der Physik* 38: 355–369, (CPAE 4, Doc. 3).

———. 1912b. “Zur Theorie des statischen Gravitationsfeldes” and “Nachtrag zur Korrektur.” *Annalen der Physik* 38: 443–458, (CPAE 4, Doc. 4).
———. 1912c. “Gibt es eine Gravitationswirkung, die der elektrodynamischen Induktionswirkung analog ist?” *Vierteljahrsschrift für gerichtliche Medizin und öffentliches Sanitätswesen* 44: 37–40, (CPAE 4, Doc. 7).
———. 1913. “Zum gegenwärtigen Stande des Gravitationsproblems.” *Physikalische Zeitschrift* 14: 1249–1262, (CPAE 4, Doc. 17). (English translation in this volume.)
———. 1914. “Zum Relativitätsproblem.” *Scientia* 15: 337–348, (CPAE 4, Doc. 31). (English translation in this volume.)
———. 1916a. “Die Grundlagen der allgemeinen Relativitätstheorie.” *Annalen der Physik* 49: 769–822, (CPAE 6, Doc. 30).
———. 1916b. “Ernst Mach.” *Physikalische Zeitschrift* 17: 101–104, (CPAE 6, Doc. 29).
———. 1917. “Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie.” *Preussische Akademie der Wissenschaften. Sitzungsberichte* 142–152, (CPAE 6, Doc. 43).
———. 1918a. “Prinzipielles zur allgemeinen Relativitätstheorie.” *Annalen der Physik* 55: 241–244, (CPAE 7, Doc. 4).
———. 1918b. “Kritisches zu einer von Hrn. De Sitter gegebenen Lösung der Gravitationsgleichungen.” *Preussische Akademie der Wissenschaften, Sitzungsberichte* 270–272, (CPAE 7, Doc. 5).
———. 1919. “Spielen Gravitationsfelder im Aufbau der materiellen Elementarteilchen eine wesentliche Rolle?” *Preussische Akademie der Wissenschaften, Sitzungsberichte* 349–356, (CPAE 7, Doc. 17).
———. 1920. *Äther und Relativitätstheorie. Rede. Gehalten am 5. Mai 1920 an der Reichs-Universität zu Leiden*. Berlin: Julius Springer, (CPAE 7, Doc. 38; Meyenn 1990, 111–123). (English translation in this volume.)
———. 1921a. “Geometrie und Erfahrung.” *Preussische Akademie der Wissenschaften, Sitzungsberichte* 123–130, (CPAE 7, Doc. 52).
———. 1921b. “A Brief Outline of the Development of the Theory of Relativity.” *Nature* 106: 782–784, (CPAE 7, Doc. 53).
———. 1922a. “Geometry and Experience.” In A. Einstein (ed.), *Sidelights of Relativity.* London: Methuen, 27–56.
———. 1922b. “Bemerkungen zu der Franz Seletyschen Arbeit ‘Beiträge zum kosmologischen System’.” *Annalen der Physik* 69: 436–438.
———. 1922c. “Bemerkungen zu der Abhandlung von E. Trefftz ‘Das statische Gravitationsfeld zweier Massenpunkte in der Einsteinschen Theorie’.” *Preussische Akademie der Wissenschaften, Sitzungsberichte* 448–449.
———. 1922d. “Bemerkung zu der Arbeit von A. Friedman ‘Über die Krümmung des Raumes’.” *Zeitschrift für Physik* 11: 326.
———. 1923a. “Grundgedanken und Probleme der Relativitätstheorie.” In *Les Prix Nobel en 1921–1922*. Stockholm: Imprimerie Royale. (Lecture held in Gotenburg on 11 July 1923.)
———. 1923b. “Zur allgemeinen Relativitätstheorie.” *Preussische Akademie der Wissenschaften, Sitzungsberichte* 32–38.
———. 1923c. “Notiz zu der Arbeit von A. Friedman ‘Über die Krümmung des Raumes’.” *Zeitschrift für Physik* 12: 228.
———. 1924. “Über den Äther.” *Verhandlungen der Schweizerischen Naturforschenden Gesellschaft* 105: 85–93.
———. 1926–27. “Über die formale Beziehung des Riemann’schen Krümmungstensors zu den Feldgleichungen der Gravitation.” *Mathematische Annalen* 97: 99–103.
———. 1929. “Space-time.” In *Encyclopedia Britannica*, vol. 21. 105–108.
———. 1930a. “Space, Ether and the Field in Physics.” In *Forum Philosophicum*, vol. 1. 180–184. (Translated from “Raum, Äther und Feld in der Physik” by Edgar S. Brightman.)
———. 1930b. “Address at the University of Nottingham.” *Science* 71: 608–610. (Address given in German and translated by Dr. I. H. Brose.)
———. 1931a. “Maxwell’s Influence on the Conception of Physical Reality.” In *James Clerk Maxwell: A Commemoration Volume*. Cambridge: Cambridge University Press, 66–73.
———. 1931b. “Zum kosmologischen Problem der allgemeinen Relativitätstheorie.” *Preussische Akademie der Wissenschaften. Sitzungsberichte* 235–237.
———. 1932. “Der gegenwärtige Stand der Relativitätstheorie.” *Die Quelle* 82: 440–442.
———. 1933. *The Origins of the General Theory of Relativity. Being the First Lecture of the Georg A. Gibson Foundation in the University of Glasgow Delivered on June 20th, 1933*. Glasgow: Jackson, Wylie and Co.
———. 1954a. “On the Theory of Relativity.” In *Ideas and Opinions*. New York: Crown, 246–249.

———. 1954b. “Notes on the Origin of the General Theory of Relativity.” In *Ideas and Opinions*. New York: Crown, 285–290.

———. 1961. *Relativity. The Special and General Theory. A Popular Exposition*. New York: Bonanza Books.

———. 1992. *Autobiographical Notes*. La Salle, Illinois: Open Court.

Einstein, Albert, and Willem de Sitter. 1932. “On the Relation between the Expansion and the Mean Density of the Universe.” *Proceedings of the National Academy of Sciences* 18: 213–214.

Einstein, Albert, and Marcel Grossmann. 1913. *Entwurf einer verallgemeinerten Relativitätstheorie und einer Theorie der Gravitation*. Leipzig: B. G. Teubner, (CPAE 4, Doc. 13).

Elkana, Yehuda. 1970. “Helmholtz’ “Kraft’: An Illustration of Concepts in Flux.” *Historical Studies in the Physical Sciences* 2: 263–298.

———. 1981. “A Programmatic Attempt at an Anthropology of Knowledge.” In E. Mendelsohn and Y. Elkana (eds.), *Sciences and Cultures. Sociology of the Sciences*, vol. 5. Dordrecht: Reidel, 1–76.

Ellis, George F. R. 1989. “The Expanding Universe: A History of Cosmology from 1917 to 1960.” In (Howard and Stachel 1989, 367–431).

Föppl, August. 1905a. “Über einen Kreiselversuch zur Messung der Umdrehungsgeschwindigkeit der Erde.” *Königlich Bayerische Akademie der Wissenschaften, München, mathematisch-physikalische Klasse, Sitzungsberichte* (1904) 34: 5–28.

———. 1905b. “Über absolute und relative Bewegung.” *Königlich Bayerische Akademie der Wissenschaften, München, mathematisch-physikalische Klasse*, *Sitzungsberichte* (1904) 34: 383–395. (English translation in this volume.)

Freudenthal, Gideon. 1986. *Atom and Individual in the Age of Newton*. Dordrecht: Reidel.

Friedlaender, Benedict and Immanuel Friedlaender. 1896. *Absolute oder Relative Bewegung?* Berlin: Leonhard Simion. (English translation in this volume.)

Goenner, Hubert. 1970. “Mach’s Principle and Einstein’s Theory of Gravitation.” In R. S. Cohen and R. J. Seeger (eds.), *Ernst Mach. Physicist and Philosopher. Boston Studies in the Philosophy of Science*, vol. 6. Dordrecht: Reidel, 200–215.

———. 1981. “Machsches Prinzip und Theorien der Gravitation.” In N. Jürgen, J. Pfarr and E.-W. Stachow (eds.), *Grundlagenprobleme der modernen Physik*. Mannheim/Wien/ Zürich: Bibliographisches Institut, 83–101.

Gregory, Frederick. 1977. *Scientific Materialism in 19th-Century Germany. Studies in the History of Modern Science*, vol. 1. Dordrecht: Reidel.

———. 2000. “The Mysteries and Wonders of Natural Science: Aaron Bernstein’s ‘Naturwissenschaftliche Volksbücher’ and the Adolescent Einstein.” In D. Howard and J. Stachel (eds.), *Einstein’s Formative Years*. (*Einstein Studies*, vol. 8.) Boston: Birkhäuser,

Hertz, Heinrich. 1894. *Die Prinzipien der Mechanik in neuem Zusammenhange dargestellt*. Leipzig: Johann Ambrosius Barth.

Hoefer, Carl. 1994. “Einstein and Mach’s Principle.” *Studies in History and Philosophy of Science* 25: 287–335.

Hofmann, Wenzel. 1904. *Kritische Beleuchtung der beiden Grundbegriffe der Mechanik: Bewegung und Trägheit und daraus gezogene Folgerungen betreffs der Achsendrehung der Erde und des Foucault’schen Pendelversuchs*. Wien, Leipzig: M. Kuppitsch Witwe.

Holton, Gerald. 1986. *The Advancement of Science, and its Burdens: The Jefferson Lecture and Other Essays*. Cambridge/New York/Sydney: Cambridge University Press.

———. 1988. *Thematic Origins of Scientific Thought. Kepler to Einstein* (rev. ed.). Cambridge/London: Harvard University Press.

Howard, Don and John Stachel (eds.). 1989. *Einstein and the History of General Relativity.* (*Einstein Studies*, vol. 1.) Boston: Birkhäuser.

Illy, József. 1989. “Einstein Teaches Lorentz, Lorentz Teaches Einstein. Their Collaboration in General Relativity, 1913–1920.” *Archive for History of Exact Sciences* 39: 247–289.

Ishiwara, Jun. 1971. *Einstein Kyozyo-Koen-roku*. Tokyo: Kabushika Kaisha.

Janssen, Michel. 1999. “Rotation as the Nemesis of Einstein’s ‘*Entwurf*’ Theory.” In H. Goenner, J. Renn, J. Ritter, and T. Sauer (eds.), *The Expanding Worlds of General Relativity.* (*Einstein Studies*, vol. 7.) Boston: Birkhäuser: Birkhäuser, 127–157.

Jungnickel, Christa, and Russel McCormmach. 1986. *The Now Mighty Theoretical Physics. Intellectual Mastery of Nature*, vol. 2. Chicago and London: The University of Chicago Press.

Kerszberg, Pierre. 1989a. “The Einstein-de Sitter Controversy of 1916–1917 and the Rise of Relativistic Cosmology.” In (Howard and Stachel 1989, 325–366).

———. 1989b. *The Invented Universe. The Einstein-De Sitter Controversy (1916–17) and the Rise of Relativistic Cosmology.* Oxford: Clarendon Press.

Köhnke, Klaus Christian. 1986. *Entstehung und Aufstieg des Neukantianismus. Die deutsche Universitätsphilosophie zwischen Idealismus und Positivismus*. Frankfurt am Main: Suhrkamp.

Kostro, Ludwik. 1992. "An Outline of the History of Einstein's Relativistic Ether Conception." In J. Eisenstaedt and A. J. Kox (eds.), *Studies in the History of General Relativity,* (*Einstein Studies,* vol. 3). Boston: Birkhäuser, 260–280.

———. 2000. *Einstein and the Ether*. Montreal: Apeiron.

Kox, A. J. 1989. "Hendrik Antoon Lorentz, the Ether, and the General Theory of Relativity." In (Howard and Stachel 1989, 201–212).

Lange, Ludwig. 1886. *Die geschichtliche Entwicklung des Bewegungsbegriffes und ihr voraussichtliches Endergebnis. Ein Beitrag zur historischen Kritik der mechanischen Principien*. Leipzig: Wilhelm Engelmann.

Lefèvre, Wolfgang. 1990. "Die wissenschaftshistorische Problemlage für Engels 'Dialektik der Natur'." In H. Kimmerle, W. Lefèvre, and R. Meyer (eds.), *Hegel-Jahrbuch* 1989. Bochum: Germinal, 455–464.

———. 1994. "La raccomandazione di Max Talmey - L'esperienza formativa del giovane Einstein." In (Pisent and Renn 1994).

Lorentz, H. A. 1910. "Alte und neue Fragen der Physik." *Physikalische Zeitschrift* 11: 1234–1257. (English translation in this volume.)

Lützen, Jesper. 1993. *Interactions between Mechanics and Differential Geometry in the 19th Century.* Preprint Series, vol. 25. København: Københavns Universitet. Matematisk Institut.

Mach, Ernst. 1883. *Die Mechanik in ihrer Entwickelung*. Leipzig: Brockhaus.

———. 1960. *The Science of Mechanics*. La Salle, Illinois: Open Court.

McCormmach, Russel. 1970. "Einstein, Lorentz and the Electron Theory." *Historical Studies in the Physical Sciences* 2: 41–87.

Meyenn, Karl von (ed.). 1990. *Albert Einsteins Relativitätstheorie: die grundlegenden Arbeiten*. Braunschweig: Vieweg.

Mie, Gustav. 1913. "Grundlagen einer Theorie der Materie. Dritte Mitteilung." *Greifswald, Physical Institute* 40: 1–65. (English translation in vol. 4 of this series.)

———. 1914. "Bemerkungen zu der Einsteinschen Gravitationstheorie. I und II." *Physikalische Zeitschrift* 14: 115–122, 169–176. (English translation in vol. 4 of this series.)

Miller, Arthur I. 1992. "Albert Einstein's 1907 *Jahrbuch* Paper: The First Step from SRT to GRT." In J. Eisenstaedt and A. J. Kox (eds.), *Studies in the History of General Relativity,* (*Einstein Studies,* vol. 3). Boston: Birkhäuser, 319–335.

Minkowski, Hermann. 1908. "Mechanik und Relativitätspostulat," appendix to "Die Grundgleichungen für elektromagnetischen Vorgänge in bewegten Körpern." *Nachrichten der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, 53–111. (English translation in this volume.)

———. 1911a. "Die Grundgleichungen für die elektromagnetischen Vorgänge in bewegten Körpern." In *Gesammelte Abhandlungen*, vol. 2. Leipzig: Teubner, 352–404.

———. 1911b. "Raum und Zeit." In *Gesammelte Abhandlungen*, vol. 2. Leipzig, 431–444.

Neumann, Carl. 1993. "The Principles of the Galilean-Newtonian Theory." *Science in Context* 6: 355–368. (Taken from a lecture held at the University of Leipzig on 3 November 1869 [1870] *Ueber die Principien der Galilei-Newton'schen Theorie*. Leipzig: Teubner.)

Newton, Isaac. 1972. *Philosophiae Naturalis Principia Mathematica* (1726 ed.), A. Koyré and I. B. Cohen (eds.). Cambridge: Harvard University Press.

Nordström, Gunnar. 1912. "Relativitätsprinzip und Gravitation." *Physikalische Zeitschrift* 13: 1126–1129. (English translation in this volume.)

Norton, John. 1989a. "What Was Einstein's Principle of Equivalence?" In (Howard and Stachel 1989, 5–47).

———. 1989b. "How Einstein found his Field Equations: 1912–1915." In (Howard and Stachel 1989, 101–159).

———. 1992a. "The Physical Content of General Covariance." In J. Eisenstaedt and A. J. Kox (eds.), *Studies in the History of General Relativity,* (*Einstein Studies,* vol. 3). Boston: Birkhäuser, 281–315.

———. 1993. "General Covariance and the Foundations of General Relativity: Eight Decades of Dispute." *Reports on Progress in Physics* 56: 791–858.

———. 1995. *Mach's Principle before Einstein*. In (Barbour and Pfister1995, 9–57).

Pais, Abraham. 1982. *Subtle is the Lord. The Science and Life of Albert Einstein*. Oxford/New York/Toronto/Melbourne: Oxford University Press.

Pisent, Gualtiero, and Jürgen Renn (eds.). 1994. *L'eredità di Einstein. Percorsi della scienza storia testi problemi*, vol. 4. Padova: il poligrafo.

Poincaré, Henri. 1905. "Sur la dynamique de l'électron." *Comptes rendus des séances de l'académie des sciences* 140: 1504–1508.

———. 1906. "Sur la dynamique de l'électron." *Rendiconto del Circolo Matematico di Palermo* 21: 129–175. (English translation in this volume.)

———. 1911. "Die neue Mechanik." *Himmel und Erde* 23: 97–116.

Pyenson, Lewis. 1985. *The Young Einstein: The Advent of Relativity.* Bristol: Hilger.

Reichenbach, Hans. 1958. *The Philosophy of Space and Time*. Translated by Marie Reichenbach and John Freund. New York: Dover.

Reissner, Hans. 1914. "Über die Relativität der Beschleunigungen in der Mechanik." *Physikalische Zeitschrift* 15: 371–375.

———. 1915. "Über eine Möglichkeit die Gravitation als unmittelbare Folge der Relativität der Trägheit abzuleiten." *Physikalische Zeitschrift* 16: 179–185.

Renn, Jürgen. 1993. "Einstein as a Disciple of Galileo: A Comparative Study of Conceptual Development in Physics." In M. Beller, R. S. Cohen and J. Renn (eds.), *Einstein in Context*. A special issue of *Science in Context*, 311–341.

———. 2003. "Book Review: Einstein and the Ether by Ludwik Kostro." *General Relativity and Gravitation* (35) 6: 1127–1130.

———. 2005. "Wissenschaft als Lebensorientierung: Eine Erfolgsgeschichte?" In E. Herms (ed.), *Leben: Verständnis. Wissenschaft. Technik*. (*Veröffentlichungen der Wissenschaftlichen Gesellschaft für Theologie*, vol. 24.) Tübingen: Gütersloher Verlagshaus, 15–31.

Renn, Jürgen and Robert Schulmann (eds.). 1992. *Albert Einstein - Mileva Marić: The Love Letters*. Princeton: Princeton University Press.

Riemann, Bernard. 1868. "Ueber die Hypothesen, welche der Geometrie zu Grunde liegen." *Abhandlungen der Königlichen Gesellschaft der Wissenschaft zu Göttingen* 13: 133–150.

Ritz, Walter. 1909. "Die Gravitation." *Scientia* 5: 241–255.

Roseveare, N. T. 1982. *Mercury's Perihelion from Le Verrier to Einstein*. Oxford: Clarendon Press.

Rynasiewicz, Robert. 1999. "Kretschmann's Analysis of Covariance and Relativity Principles." In H. Goenner, J. Renn, J. Ritter, and T. Sauer (ed.), *The Expanding Worlds of General Relativity.* (*Einstein Studies*, vol. 7.) Boston: Birkhäuser, 431–462.

Schrödinger, Erwin. 1925. "Die Erfüllbarkeit der Relativitätsforderung in der klassischen Mechanik." *Annalen der Physik* 77: 325–336.

Sciama, Dennis W. 1959. *The Unity of the Universe*. Garden City, New York: Doubleday & Co.

Sewell, William Clyde. 1975. "Einstein, Mach, and the General Theory of Relativity." Ph. D., Case Western Reserve University.

Sommerfeld, Arnold. (ed). 1903–1926. *Physik*, 3 vols. (*Encyklopädie der mathematischen Wissenschaften mit Einschluss ihrer Anwendungen,* vol. 5.) Leipzig: Teubner.

Stachel, John. 1986. "Eddington and Einstein." In E. Ullmann-Margalit (ed.), *The Prism of Science*. Dordrecht and Boston: Reidel, 225–250.

———. 1989. "The Rigidly Rotating Disk as the 'Missing Link' in the History of General Relativity." In (Howard and Stachel 1989, 48–62).

Stein, Howard. 1977. "Some Philosophical Prehistory of General Relativity." In C. Glymour, J. Earman and J. Stachel (eds.), *Foundations of Space-Time Theories*. Minneapolis: University of Minnesota Press, 3–49.

Torretti, Roberto. 1978. *Philosophy of Geometry from Riemann to Poincaré*. Dordrecht/Boston/London: Reidel.

———. 1983. *Relativity and Geometry.* Oxford: Pergamon Press.

Vizgin, Vladimir P. 1994. *Unified Field Theories in the First Third of the 20th Century. Science Networks, Historical Studies*, E. Hiebert and H. Wussing (eds.), vol. 13. Basel/Boston/Berlin: Birkhäuser.

Wahsner, Renate, and Horst-Heino von Borzeszkowski. 1992. *Die Wirklichkeit der Physik: Studien zu Idealität und Realität in einer messenden Wissenschaft*. Europäische Hochschulschriften, Frankfurt am Main/Berlin/Bern/New York/Paris/Wien: Lang.

Wheeler, John A. 1979. "Einstein's Last Lecture." In G. E. Tauber (ed.), *Albert Einstein's Theory of General Relativity.* New York: Crown.

Wolters, Gereon. 1987. *Mach I, Mach II, Einstein und die Relativitätstheorie: Eine Fälschung und ihre Folgen*. Berlin/New York: de Gruyter.

Zenneck, Jonathan. 1903. "Gravitation." In (Sommerfeld 1903–1926), 1: 25–67. (Printed in this volume.)

JONATHAN ZENNECK

GRAVITATION

Originally published as the entry "Gravitation" in the Encyklopädie der mathematischen Wissenschaften, 5th Volume: Physics, A. Sommerfeld, ed. B. G. Teubner, Leipzig 1903–1921, pp. 25–67. Finished in August, 1901, as noted by the author.

TABLE OF CONTENTS

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 3
Gravitation in the Twilight of Classical Physics: Between Mechanics, Field Theory, and Astronomy.

Comprehensive Literature

can be found at the beginning of each section in notes 1, 2, 36, 47, 48, 77, 82, 107.

Introductory Note

In this paper astronomical questions, which are not dealt with in detail until Volume VI, must be mentioned several times. The present paper does not aspire to completeness in this regard, but only draws upon as much astronomical material as is unavoidable for treating the subject.

1. Newton's Law

As is well known, the fundamental gravitational law was first[1] conceived clearly by Newton, and formulated in the third book of his *Philosophiae naturalis principia mathematica*, propositions I–VII. |

It states: If at a certain instant of time two mass elements with masses m_1 and m_2 [27] are at distance r from each other, then at the same instant a force acts on each of the two elements in the direction of the other with a magnitude

$$G\frac{m_1 m_2}{r^2}.$$

In this expression, G is a universal—i.e., only dependent on the system of units—constant, the so-called gravitational constant.

1. DETERMINATION OF THE GRAVITATIONAL CONSTANT[2]

2. Significance of this Determination

The inherent significance of the absolute determination of any physical constant is enhanced in the case of the gravitational constant for two additional reasons:

1. If the gravitational constant is known, the acceleration due to gravity g and the dimensions of the Earth yield the mass and the mean density of the Earth.[3] The latter

---

1 About Newton's forerunners cf. F. Rosenberger, *Isaac Newton und seine physikalischen Prinzipien*, Leipzig 1895. A compilation of nearly all papers (up to 1869) which are in some way related to the *mathematical* implementation of the law of attraction can be found in I. Todhunter, *History of the mathematical theories of attraction and the figure of the Earth*, 2 Vols., London 1873.

2 Principal review literature about absolute determinations: J.H. Poynting, *The Mean Density of the Earth*, London 1894; F. Richarz and O. Krigar-Menzel, *Berl. Abh.* 1898, Appendix; C.V. Boys, *Rapp. congrès internat. phys.* 3, Paris 1900, p. 306–349. Then there are Gehler's *Physikalisches Handwörterbuch*, Leipzig 1825, articles: *Anziehung, Drehwage, Erde, Materie*; S. Günther, *Lehrbuch der Geophysik* 1, 2nd ed., Stuttgart 1879; F. Richarz, Leipzig, *Vierteljahrschr. astr. Ges.* 24 (1887), p. 18–32 and 184–186.

3 If Δ is the mean density of the Earth and R its radius, to first approximation we get

$$g = \frac{4}{3}R\pi\Delta G.$$

Considering the corrections which are caused by flattening, centrifugal force and their differences within various latitudes, then one arrives at the relation explained in detail by F. Richarz and O. Krigar-Menzel[2]

$$9.7800\frac{\mathrm{m}}{\mathrm{sec}^2} = \frac{4}{3}\cdot R_P\pi\Delta G\left(1+\mathfrak{a}-\frac{3}{2}\mathfrak{c}\right),$$

where

R_P = earth radius at the pole = 6356079m,

$\mathfrak{a}$ = flattening = 0.0033416,

$\mathfrak{c}$ = relation between centrifugal force and gravity at the equator = 0.0034672.

was the ultimate aim of most determinations; therefore, they are usually known as *determinations of the mean density of the Earth*.

[28] 2. If the Earth's mass is known, the masses of the | other planets and of the Sun follow, since the proportion of these masses to the Earth's mass is supplied by astronomical observation.[4]

3. Survey of Various Methods

The various methods chosen to gain the value for the gravitational constant G in absolute terms can be divided essentially into three main classes:

1. The force that masses of known magnitudes at known distances exert on each other, was determined directly: determinations with the torsion balance, the double pendulum, and the ordinary balance.[5]

2. Changes in the direction or magnitude of the acceleration due to gravity g caused by masses of known magnitudes were measured: deflection of the plumbline, pendulum observations.

3. It was attempted to calculate the Earth's mean density and thereby the gravitational constant from the density at the surface, based on a more or less hypothetical law about the increase of density towards the center of the Earth.

4. Determinations with the Torsion Balance

a. Static method. The weights attached to the balance beam are attracted by masses *next to* the beam. The resulting rotation of the beam is a measure of the attractive force's magnitude.

This method, which was probably first suggested by Reverend J. Michell,[6] was used by H. Cavendish,[7] F. Reich,[8] F. Baily,[9] A. Cornu and J. Baille,[10] C.V. Boys,[11] and finally by C. Braun.[12]

Reich's advance over Cavendish lies primarily in his use of a mirror arrangement to make measurements. Baily's measurements are particularly valuable because they were extended to a large number of materials, and were varied in other, manifold
[29] ways. Cornu and Baille have shown that the same | accuracy (the same deflection angle) can be achieved despite any reduction of scale, if only a suitable choice in sus-

---

4 But cf. section 11.

5 In the latter method, g enters into the result.

6 Quoted from Cavendish, *Lond. Trans.* 88 (1798).

7 See above note.

8 "Versuche über die mittlere Dichtigkeit der Erde mittelst der Drehwage," Freiberg 1838, and "Neue Versuche mit der Drehwage", Leipzig 1852.

9 *Lond. Astr. Soc. Mem.* 14 (1843).

10 Paris, *C. R.* 76 (1873), p. 954–58.

11 *Lond. Trans.* 186 (1889), p. 1–72.

12 *Wien. Denkschr.* 64 (1897), p. 187–285. Report on this: F. Richarz, *Leipzig Vierteljahrsschr. astr. Ges.* 33 (1898), p. 33–44.

pension provides equal oscillation periods of the torsion balance. Consequently, they used much smaller dimensions for their apparatus and so avoided a number of disturbances. Boys[13] extended this reduction to a smaller scale and made it possible to replace the metal suspension wires with much more favorable quartz fibers. Braun uses a torsion balance in a vacuum to avoid totally the worst disturbance while measuring with the torsion balance, namely the air currents.

Boys partly evaded the deficiency of the extremely small dimensions he used by skillful arrangement of his torsion balance; however, the disadvantage remains that with small dimensions, apart from the strong damping of the oscillations, errors in length determination and deficient homogeneity of the material can easily spoil the result's accuracy.[14] To avoid this deficiency of small dimensions and nevertheless reach high sensitivity, F. R. Burgess[15] suggested that arranging the weights to float on mercury would enable the use of large masses and thin suspension wires. In a pre-experiment with weights of 10×2 kg on both sides, he found a 12° deflection, but has not yet carried out his determinations.

b. Dynamical method. The attracting masses are placed *in line* with the balance beam. Their attraction serves to reinforce the restoring torque of the suspension. The resulting decrease of the oscillation period gives a measure of the attractive force's magnitude.

With this method, C. Braun obtained a value of G that agrees very well with results of his measurements via the static method. R. von Eötvös[16] suggested modifying this | method but has not yet published final results. [30]

5. Determinations with the Double Pendulum

J. Wilsing's vertical double pendulum[17]—a vertical balance beam with a weight at each end, attracted by horizontally displaced masses—does not work by torsion of wires, but uses gravity as the restoring force. The torque is reduced to a minimum by placing the double pendulum's center of gravity only ca. $0,01\,\mathrm{mm}$ beneath the edge. Such an arrangement combines high sensitivity[18] with significant stability, and moreover, in contrast to the torsion balance, has the advantage of being influenced to a lesser degree by air currents.

---

13 At length 2.3 cm of the balance beam, loaded on both sides with 1.3 to 3.98 g and deflected at each side by 7.4 kg, Boys received a deflection of ca. 370 scale sections. For Cavendish the quantities concerned were 196 cm, 730 g, 158 kg; he received a deflection of 6–14 scale sections.

14 Cf. F. Richarz's report cited in note 12.

15 Paris, *C.R.* 129 (1899), p. 407–409. Poynting[2] had already performed a similarly arranged experiment, but abandoned this arrangement due to interfering currents in the fluid.

16 *Ann. Phys. Chem.* 59 (1896), p. 354–400.

17 *Potsdam. Astr.-physik. Obs.* 6 (1887), No. 22 and 23.

18 At 325×0.54 kg, 1 to $10'$ deflection.

6. Determinations with the Ordinary Balance

The principle of this method, seemingly already presented by Descartes,[19] is as follows. Two equal weights m are placed on the scales of a balance. Underneath one of the two scales—possibly simultaneously above the other one—a mass M is brought in. The weight difference now observed provides a measure of the attraction M has on m.

For the purpose of absolute determination of the gravitational constant this method was probably first used by Ph. von Jolly,[20] later by J.H. Poynting,[2] and then by F. Richarz and O. Krigar-Menzel.[2]

Jolly's arrangement, which was already used in Newton's time by Hooke[21] in a quite similar way to determine a decrease of g with height, has the disadvantage that vertical air currents caused by temperature differences can disturb the weighing process by friction on the long suspending wires. Poynting avoided this shortcoming; furthermore, he saw to it that the angle by which the balance beam rotates can be read precisely, and that the attracting weights can be removed or brought closer without having to lock the balance or to expose it to vibration. Richarz's and Krigar-Menzel's
[31] method has the advantage of tolerating extraordinarily large attracting | masses (100,000 kg lead) without excessive difficulties, and moreover of allowing an effective four-fold attraction of this mass. However, the method suffers from the drawback that relieving and locking the balance becomes necessary in the course of determination.

7. Determinations with Plumbline and Pendulum

a. Static method (*plumbline deflection*). Deflecting masses were always mountains, and their deflection of the plumbline was determined by measuring the difference of geographical latitude between two points, if possible taken to the north and south of the deflecting mountain, once astronomically—where the direction of the plumbline enters—and then trigonometrically. The difference between the two determinations is twice the deflection caused by the mountain. The dimensions and the specific weight of the rocks determine the mass of the mountain.

Determinations of this kind were carried out by Bouguer[22] at Chimborazo, by N. Maskelyne and C. Hutton,[23] later by James[24] and Clarke at mountains in the Scottish highlands, by E. Pechmann[25] in the Alps and under particularly favorable conditions by E.D. Preston[26] on the Hawaiian islands.

---

19 Cited in *Observ. Sur la physique*, 2, Paris 1773.
20 *Münchn. Abh.* (2) 14 (1881); *Ann. Phys. Chem.* 14 (1881), p. 331–335.
21 Cited in Rosenberger, note 1.
22 *La figure de la terre*, Paris 1749, sec. VII, chap. IV.
23 *Lond. Trans.* 1775, p. 500–542; 1778, p. 689–788; 1821, p. 276–292.
24 *Phil. Mag.* (4) 12 (1856), p. 314–316; 13 (1856), p. 129–132 and *Lond. Trans.* 1856, p. 591–607.
25 *Wien. Denkschr.* (*math.-naturw. Kl.*) 22 (1864), p. 41–88.
26 Washington, *Bull. Phil. Soc.* 12 (1895), p. 369–395.

It has been suggested to use the sea at low and high tide[27] or a drainable lake[28] instead of a mountain, but a determination never seems to have been performed this way, though it would have significant advantages over using a mountain.

b. Dynamical method (*pendulum observation*). The scheme of such determinations is the following. Either at the foot and the top of a mountain or on the Earth's surface and in the depth of a mine, the oscillation period of the same pendulum is observed. The measured difference in oscillation period and hence in the acceleration due to gravity at the two points | yields the attraction of the mountain or the layer of Earth [32] above the mine, respectively.[29]

Determinations of the first kind are due to Bouguer[22] (Cordills), Carlini[30] (and Plana) (Mont Cenis), under particularly favorable conditions from Mendenhall[31] (Fujiyama, Japan) and E.D. Preston[26] (Hawaiian islands).

Determinations of the second kind were first suggested by Drobisch,[32] later carried out by G.B. Airy[33] and in large number by R.v. Sterneck.[34]

A third method, in principle definitely more favorable, was attempted by A. Berget:[35] artificial alteration of g due to a difference in the water level of a drainable lake. However, his determination was spoiled by an unsuitable measurement of this change in g.

8. Calculation of the Gravitational Constant[36]

1. Laplace[37] based his calculations on the following conditions, as did Clairaut and Legendre:

a. The Earth consists of separate ellipsoidal layers. The density within each layer is constant.

b. The rotation is so slow that the deviation from the spherical shape becomes small, as well as the influence of the centrifugal force on g.

c. The Earth's substance shall be regarded as fluid.

---

27 By Robison 1804 (cited by Richarz and Krigar-Menzel, see note 2), Boscowich 1807 (cited in *Monatl. Korrespondenz z. Beförd. d. Erd- u. Himmelskunde* 21 (1810)), furthermore by von Struve (cited in *Astr. Nachr.* 22 (1845), p. 31 f.)

28 F. Keller, *Linc. Rend.* 3 (1887), p. 493.

29 Cf. for this and the following numbers vol. VI of the *Encyclop., Geophysik*.

30 *Milano Effem.* 1824. Cf. E. Sabine, *Quart. J.* 2 (1827), p. 153 and C.J. Guilio, *Torino Mem.* 2 (1840), p. 379

31 *Amer. J. of Science.* (3) 21 (1881), p. 99–103.

32 *De vera lunae figur*a etc., Lipsiae 1826.

33 *Lond. Trans.* 1856, p. 297–342 and 343–352. For calculation cf. S. Haughton, *Phil. Mag.* (4) 12 (1856), p. 50–51 and F. Folie, *Bruxelles Bull.* (2) 33 (1872), p. 389–409.

34 Wien. *Mitteil. d. milit.-geogr. Inst.* 2–6, 1882–1886 and *Wien. Ber.* 108 (2a), p. 697–766.

35 Paris *C. R.* 116 (1893), p. 1501–1503. Cf. Gouy's objection (Paris *C. R.* 117 (1893), p. 96) that the temperature would have had to be constant at least to 0.2×10^{-6} degrees.

36 Cf. F. Tisserand, *Mécan. cél.* 2, Paris 1891, chap. XIV and XV.

37 *Méc. cél.* 5 (1824), Livr. 11, chap. 5.

Under these circumstances Laplace calculates the conditions of equilibrium, into
[33] which, besides the Earth's elliptical property, the | law expressing the density of an Earth layer as a function of its distance from the center enters. Laplace makes two assumptions for this law:

$$\rho = \rho_0[1 + e(1 - a)], \tag{1}$$

$$\rho = \frac{A}{a}\sin(an), \tag{2}$$

where ρ stands for the density, a for the distance of a layer to the Earth's center (Earth radius = 1), and ρ_0, e as well as A, n are constants. These constants are determined on the one hand by the value of ρ on the Earth's surface, and on the other hand by the derived equilibrium conditions. One then obtains a relation between the mean density of the Earth (and hence of the gravitational constant) and the Earth's surface density ρ_0; that is, from the first assumption regarding increasing density towards the center of the Earth, it follows that

$$\Delta_1 = 1.587 \cdot \rho_0,$$

and from the second assumption, that

$$\Delta_2 = 2.4225 \cdot \rho_0,$$

if the Earth's ellipticity is assumed to equal 0,00326.

2. On essentially the same basis, using Laplace's second assumption regarding increasing density towards the center of the Earth, Clairant's formula for the equilibrium of the rotating Earth, taken to be a fluid, and the value 0.00346 for the ellipticity of the Earth, J. Ivory[38] arrives at the relation:[39]

$$\Delta = 1.901 \cdot \rho_0.$$

3. The recent extensive literature on this question (Lipschitz, Stieltjes, Tisserand, Roche, Maurice Lévy, Saigey, Callandreau, Radau, Poincaré, Tumlirz) can not be discussed here; therefore we refer to the previously cited chapters in Tisserand[40] or to Vol. VI of the Encyclopedia.

9. The Result of the Determinations

Regarding the question of what should be taken as the most probable value for the gravitational constant, the results of the methods discussed in sections 7 and 8 must immediately be excluded. |

38 *Phil. Mag.* 66 (1825), p. 321 f.

39 Taking for the mean density all over the Earth surface S. Haughton's[33] value $\rho_0 = 2.059$, one would obtain according to Laplace: $\Delta_1 = 3.268$, $\Delta_2 = 4.962$, and according to Ivory: $\Delta = 3.914$.

40 Cf. note 36.[1]

Indeed, the terrestrial methods (section 7) actually carried out have the advantage [34] over laboratory methods (sections 4–6) in that the attracting masses and hence the differences to be observed have a relatively significant magnitude. But this advantage is more than outweighed by the fact that dimensions and density of the attracting masses are known only incompletely, and that the inadequately observed mass distribution below the place of observation plays an essential, but entirely uncontrollable, part.[41]

Those terrestrial methods, however, that could have had prospects for success, because not only do they allow for using very large masses, but also because the magnitude of the attracting masses could be determined with sufficient accuracy, and because the influence of the environment—such as change of magnitude or direction of g by a lake or the sea at different levels—would drop out, have not been carried out at all or were carried out only in a flawed manner.

The attempts to calculate the gravitational constant (section 8) can not provide a reasonably reliable result, either. Apart from other uncertainties, the mean surface density of the Earth enters this calculation, and this is far from being known with as much accuracy as the gravitational constant itself when obtained by laboratory determination.

Thus only the results of laboratory determinations remain (sections 4–6). Considering the two most recent determinations from each method only, we get the following compilation:

		Δ	G
Torsion balance	*Boys*	5.527	$6.658 \cdot 10^{-8}\ \mathrm{cm}^3\ \mathrm{s}^{-2}\ \mathrm{g}^{-1}$
	Braun	5.5270[a]	
(Double pendulum	*Wilsing*	5.577	$6.596 \cdot 10^{-8}\ \mathrm{cm}^3\ \mathrm{s}^{-2}\ \mathrm{g}^{-1}$)
Balance	*Poynting*	5.4934	$6.698 \cdot 10^{-8}\ \mathrm{cm}^3\ \mathrm{s}^{-2}\ \mathrm{g}^{-1}$
	Richarz and *Krigar-Menzel*	5.5050	$6.685 \cdot 10^{-8}\ \mathrm{cm}^3\ \mathrm{s}^{-2}\ \mathrm{g}^{-1}$
Mean value of these determinations		5.513[b]	$6.675 \cdot 10^{-8}\ \mathrm{cm}^3\ \mathrm{s}^{-2}\ \mathrm{g}^{-1}$

a. In copies issued later, Braun assumed the most probable result of his observations to be $\Delta = 5{,}52725$ (communicated by Prof. F. Richarz).

b. As we know, Newton (*Principia lib. III, prop. X*) estimated the Earth's mean density to be 5–6. The mean 5.5 thus agrees with the mean value from the most recent measurements to $1/4$ %.

41 Cf. W.S. Jacob, *Phil. Mag.* (4) 13 (1857), p. 525–528. Conversely, such determinations can be significant because they allow for a conclusion about mass distribution close to the place of observation. Cf. R. v. Sterneck's *Messungen*.[34]

[35] | The good agreement[42] between the values obtained by the same method on the one hand, and the relatively significant disagreements among results of different methods[43] on the other, show that these disagreements can only be due to deficiencies in principle of the methods. As long as these have not been cleared up, none of the results can be given more weight than another. It is a pity that Wilsing's method has not yet been employed by a second observer to check Wilsing's result, and that the influence of magnetic permeability of the double pendulum has not yet been examined.[44] Therefore, we did not take Wilsing's result into consideration in the calculation of the mean value above.

2. ASTRONOMICAL AND EXPERIMENTAL TESTS OF NEWTON'S LAW

10. General

Two independent fields insure that Newton's law, even if not absolutely accurate, represents real conditions with a far-reaching accuracy unmatched by hardly any other law.

In the *astronomical*[45] domain, this law yields planetary motions not only to the first approximation (Kepler's laws); but even to the second approximation, the deviations of planetary motion due to pertubations by other planets follow from Newton's law so accurately that the observed perturbations led to the prediction of the orbit and relative mass of a hitherto unknown planet (Neptune).

[36] On the other hand there are a number of astronomical | observations that show deviations compared to calculations based on Newton's law. This deviation amounts to[46]

1. ca. 40″ per century in the perihelion motion of Mercury;
2. 5-fold probable error in the motion of the node of Venus' orbit;
3. 3-fold probable error in the perihelion motion of Mars; and
4. 2-fold probable error (uncertain!) in the eccentricity of Mercury's orbit.

In addition there are:

5. significant anomalies in the motion of Encke's comet; and

---

42 Between torsion balance determinations there is a difference ≤ 0.012 %, between balance determinations there is a difference of ca. 0.2%.

43 The largest difference between balance and double pendulum determinations is ca. 1.5%.

44 According to F. Richarz and O. Krigar-Menzel (*Bemerkungen zu dem ... von Herrn C.V. Boys über die Gravitationskonstante ... erstatteten Bericht*, Greifswald 1901) the deviation of Wilsing's result from others could be caused by such an influence.

45 Discussion of the validity of Newton's law in the astronomical domain in Tisserand, *Méc. cél*. 4 (1896), chap. 29 and S. Newcomb.[48]

46 S. Newcomb, *The elements of the four inner planets* etc., Washington 1895. On page 109 ff. is a discussion of possible explanations of these deviations.

6. small irregularities in the Moon's orbit.

Small corrections of Newton's law are therefore not excluded by the astronomical evidence,[47] even if it is not at all settled—particularly in the cases listed in 5) and 6) where conditions are more complicated and uncertain than for planetary orbits—that the above differences are due to an inaccuracy of the gravitational law.[46]

In the *experimental* area the best determinations of the gravitational constant, which all rest on the assumption of the validity of Newton's law, yield results in rather good agreement.[48] As these determinations were carried out with masses of a great variety of magnitudes, materials and distances, this agreement therefore excludes any considerable inaccuracy of | Newton's law, and allows at most for small [37]
corrections.

11. Dependence on Mass: Astronomical Test

Newton inferred that the force that two bodies exert on each other is proportional to the mass of each body as follows:

a. Observation shows Jupiter conferring *acceleration* to its satellites, the Sun to the planets, Earth to the Moon, and the Sun to Jupiter and its satellites, which is equal at equal distance. Hence, it follows that in these cases the *force* must be proportional to the mass of the *attracted* body.

47 Th. von Oppolzer (*Tagebl. d. 54. Vers. d. Naturf. u. Ärzte*, Salzburg, 1881) even draws quite an apodictic conclusion: "The theory of the Moon makes a conjecture quite probable, the theory of Mercury points at it firmly, Encke's comets lift it up to an irrefutable certainty that the theories built solely on Newton's law of attraction *in present form* are not sufficient for explaining the motion of heavenly bodies."

48 To compare the best terrestrial and laboratory methods:

	Observer	attracting mass	Δ
Laboratory methods	*Boys*	7.4 kg	5.527
	Braun	9.1 "	5.5270
	Poynting	154 "	5.4934
	Wilsing	325 "	5.577
	Richarz and *Krigar-Menzel*	100.000 "	5.5050

	Observer	attracting mass	Δ	
Terrestrial methods	*Mendenhall*	mountain of 3.800 m height	5.77	
	E. D. Preston	" 3.000 "	5.57	5.35
	"	" 4.000 "	5.13	
	von Sterneck	strata of various thickness	5.275	
	(Wien. Ber. 108)	"	5.56	
	"	"	5.3	
	"	"	5.35	

b. The principle of action and reaction then implies that the force must be proportional to the mass of the *attracting* body as well.

M.E. Vicaire[49] raised the following objection against this line of reasoning, which would however require discussion.[50] The examples presented represent a very special case: a very large body attracting a relatively very small body. But then the assumption that at equal distance the attraction can only be a function of the two masses already provides the result that the attractive force must be approximately proportional to the small body's mass.

This is because the function A_{Mm}, which expresses the attraction a large mass M has on small mass m, is certainly homogeneous in M and m. One can hence put: |

[38]

$$\begin{aligned} A_{Mm} &= M^k \cdot f(\frac{m}{M}) \\ &= M^k \Big[\frac{m}{M} \cdot f'(0) + \Big(\frac{m}{M}\Big)^2 \cdot \frac{f''(0)}{1 \cdot 2} + \ldots\Big] \\ &= M^{k-1} \cdot m \cdot f'(0) \qquad \text{approx.}, \end{aligned}$$

so the attraction is to first approximation proportional to m. Hence, from the fact that this proportionality is confirmed by observation, one must not conclude that the attraction is also proportional to the mass of the attracting large body. However, from this it would follow that calculations of planetary masses in relation to the Sun's mass based on Kepler's third law are in principle misguided.

Vicaire also objects to supporting these calculations by the usual calculations from the planetary perturbations. The *secular* perturbations of a planet m by another m', which are primarily observed and drawn upon in these calculations, do not at all result in the relative mass of planet m', but in the proportion $A_{mm'} : A_{Mm}$, which according to the above does not need to be identical to $m' : M$. Only the *periodic* perturbations could provide information about $m' : M$.

12. Dependence on Mass:
Experimental Test for Masses of the Same Material

The G-determinations of Poynting[2] and of Richarz and Krigar-Menzel[2] are of special value in relation to the question of how far the proportionality of the attractive force to the mass is guaranteed for masses of the same material. Both experimenters used unobjectionable laboratory methods carried out with the greatest care. Both

49 Paris, *C.R.* 78 (1874), p. 790–794.

50 This is opposed by the agreement within probable error between the mass of the planets determined from the perturbations which they exert on other planets, and the mass of the same planets determined from the motion of their moons, if they have any. For example, the mass of Mars from Jupiter's perturbations results in $= 1/2.812.526$, from the elongation motions of its moons $= 1/3.093.500$. Cf. as well F.W. Bessel, *Berl. Abh.* 1824 and *Ges. Werke* 1, p. 84.

determinations employed the same material (lead) and the same method of measurement, but masses of very different magnitudes (154 or 100.000 kg). Even though in one case the mass was 650 times greater than the other, the results agree to approximately 0.2%.

13. Dependence on Mass: Experimental Test for Masses of Various Chemical Compositions

Three different methods have been used to examine whether the proportionality of the attractive force to mass is also strictly valid for masses of different chemical compositions.

a. The gravitational constant was determined for masses of different materials.

F. Baily[9] carried out a large number of measurements of this kind. If his results are arranged according to the specific weight | of the mass which was suspended from [39]
the torsion balance,[51] and if we take for each material the mean value from all measurements, the following is revealed. The values of Δ increase—the values of G decrease accordingly—as the specific weight of the mass is decreased.[52] However, there is reason to assume that these disagreements are a matter of a basic error in his arrangement or calculation.[53]

In any case, the fact that the results of Boys[11] and Braun[12] agree to 0.01%, although they refer to different materials, counts against the assumption that these different results are due to a different value of the gravitational constant for different substances. Likewise, with the help of a particularly sensitive torsion balance, v. Eötvös[16] claims to have found that the difference of attraction of glass, antimony, and corkwood from that of brass is less than $1/2 \cdot 10^{-7}$ and of air from that of brass less than $1 \cdot 10^{-5}$ of the total attraction.

b. Pendulums were produced out of various materials to compare their periods of oscillation.

---

51 Same attracting substance everywhere = lead.

52

Substance	specific weight	Δ	
Platinum	21	5.609	
Lead	11.4	5.622	
Brass	8.4	5.638	
Zinc	7	5.691	
Glass	ca. 2.6	5.748	exception
Ivory	1.8	5.745	

53 Cf. also F. Reich in the paper cited in note 8, "*Neue Versuche* etc.", p. 190.

This method, already employed by Newton,[54] has been refined by F.W. Bessel,[55] in particular. While Newton could only conclude from his experiments that the difference of attraction which the Earth exerts on bodies of very different composition is smaller than $1 \cdot 10^{-3}$ of the total attraction, Bessel managed to squeeze this limit down to $1/6 \cdot 10^{-4}$.

c. A sealed vessel which contains two different chemical substances is weighed, then the substances combine, and after completion of the chemical reaction the vessel is weighed again. |

[40] The first experiments of this kind by D. Kreichgauer[56] with mercury and bromine, and with mercury and iodine gives the result "that with the bodies employed, a change of attraction by the Earth due to chemical forces should stay below 1/20,000,000 of the total attraction." But H. Landolt[57] found under conditions as simple as possible—except for reactions, where a change of weight could not be determined with certainty—the following:

1. For reduction of silver sulphate by ferrosulphate in three series of experiments, a weight decrease by 0.167, 0.131 and 0.130 mg.

2. For iodic acid and hydrogen iodide weight decreases in six experimental series, varying between 0.01 and 0.177 mg.

Not only do these decreases in weight exceed probable measurement errors, but some of them also exceed the largest deviation among single measurements. A. Heydweiller[58] resumed these measurements after M. Hänsel[59] established that the deviations observed by Landolt in the first example can not be explained by the influence of magnetic forces. He also obtains decreasing weight in a series of cases and reaches the conclusion: "one may regard a change in weight as ascertained: in the effect iron has on copper sulphate in acid or basic solution ... , regarding the dissolution of acid copper sulphate ... , and in the effect potassium hydroxide has on copper sulphate ... ."

The cases presented above are therefore *well established but for the time being completely unexplained deviations from the proportionality of the action of gravity to mass*.

---

54 *Principia lib. III, propos. VI.*

55 *Astr. Nachr.* 10 (1833), p. 97.

56 *Berl. physik. Ges.* 10 (1891), p. 13–16.

57 *Zeitschr. physik. Chem.* 12 (1894), p. 11. He cites that in the synthesis of iodine and bromide silver J. S. Stas always obtained less than equivalent of the initial quantities. Indeed, the difference amounted on average in five experiments to $1/4 \cdot 10^{-4}$ of the total mass.

58 *Ann. Phys.* 5 (1901), p. 394–420.

59 Diss. Breslau 1899.

14. Dependence on Mass:
Experimental Test for Masses of Various Structures

The conjecture that attraction between two masses could depend on their structure is suggested by several theories explaining gravitation. This was examined experimentally in two directions. |

a. Kreichgauer[56] examined whether a body (acetic sodium) changes weight while crystallizing. He found, however, that any change of weight is below $1/2 \cdot 10^{-7}$ of the total attraction.[60] [41]

b. A. S. Mackenzie[61] as well as J. H. Poynting and P. L. Grey[62] deal with the question of whether the gravitational effect of a crystalline substance varies with different directions. Mackenzie tested calcite against lead, and also calcite against calcite, but he found the difference to be smaller than 1/200 of the total attraction. Poynting and Grey arrive at the result that the attraction of quartz to quartz at parallel and crossed axes differs less than 1/16500 of the total attraction, and that at parallel axes, when one of the crystals is rotated by 180°, the attraction changes by less than 1/2850 of the total.

15. Dependence on Distance: Astronomical Test (cf. Vol. VI)[2]

S. Newcomb[63] discussed the question of the extent to which the $1/r^2$ in Newton's law is fixed by astronomical data. He reaches the following result:

a. The agreement between the observed parallax of the Moon and that calculated from the magnitude of g on the Earth surface shows that for values of r that lie between Earth's radius and the radius of the Moon's orbit, the 2 in r^2 is guaranteed up to 1/5000 of its value.

b. The agreement between the observed perturbation of the Moon by the Sun and the calculation based on Newton's law proves (with about the same accuracy) the validity of r^2 up to distances of the order of magnitude of the Earth's orbit, i.e. approximately up to 24000 times the Earth's radius.

c. The validity of Newton's law up to the limits of the entire planetary system follows from the validity of Kepler's third law; that is, up to distances which amount to 20 times the Earth orbit's radius. Yet for this range the | accuracy with which $1/r^2$ [42] can be established from observation cannot be stated with certainty.

One more touchstone of the same question, as Newton[64] already emphasized, is related to the fact that *perihelion motion* of the planets would result from a deviation in the exponent of the distance from 2. While on the one hand such a deviation can

---

60 Earlier Bessel,[57] and more recently von Eötvös,[16] found no difference between crystalline and amorphous bodies in their experiments.

61 *Phys. Rev.* 2 (1895), p. 321–343.

62 *Lond. Trans.* A 192 (1899), p. 245–256.

63 In the paper cited in note 46.

64 *Principia lib. I, sec. IX.*

not be large, because otherwise this would result in perihelion motions which contradict observation, on the other hand the observed anomalous perihelion motions could be rooted in a minor inaccuracy of the gravitational law. Indeed, M. Hall[65] proved that the law previously examined by G. Green,[66] which replaces $1/r^2$ with $1/r^{2+\lambda}$, where λ stands for a small number, is sufficient to explain the anomalous perihelion motion of Mercury, if $\lambda = 16 \cdot 10^{-8}$. This figure for λ would also give the right result for the observed anomalous perihelion motion of Mars, though for Venus and Earth the consequence would be somewhat too large a perihelion motion.[67] However, Newcomb, after discussing the respective conditions, states that Hall's assumption seems to him "provisionally not inadmissible."

16. Dependence on Distance: Experimental Test

This question was examined directly by Mackenzie,[61] by measuring at various distances the attraction of the same bodies with the torsion balance. He found that the discrepancy between the observed result and that calculated from Newton's law is in any case smaller than 1/500 of the total attraction.

From a theoretical perspective, our confidence in the 2 in the exponent of Newton's law stems essentially from the fact that from the standpoint of field theory (section 34) this law alone is compatible with the assumption of a general, source-free distribution of field strength; i.e., the concept of lines of force of the gravitational field is meaningful only if this law is valid precisely.

17. Influence of the Medium on Gravitation

The analogy of electric and magnetic charges, whose effect depends to a large degree on the medium in which they are contained, makes it seem altogether possible that [43] such an influence is present in gravitation | as well, and that hence the gravitational constant is not universal, as Newton assumed, but rather depends on the medium. Just the relatively good agreement, in spite of the very different *form* of the employed masses, among G-determinations excludes a fairly considerable influence of bodies in the region between the attracting masses.[68] Furthermore, with a torsion balance L.W. Austin and C.B. Thwing[69] directly examined the question of whether a body with a different permeability for gravitation than air exists. Between two bodies attracting each other they inserted plates of various substances whose thickness was 1/3 the distance between the attracting masses. The result was that the difference would have to be smaller than 0.2% of the total attraction.

In another direction, Laplace[70] discussed the question of a possible influence of the medium. He assumes that bodies except air may possess a small absorption coef-

---

65 *Astr. Journ.* 14, p. 45.

66 *Cambr. Trans.* 1835, p. 403.

67 Cf. Newcomb in the paper cited in note 46, p. 109.

ficient α for gravitation so that the gravitational law for two mass elements m_1 and m_2 embedded in such a medium would be:

$$K = G\frac{m_1 \cdot m_2}{r^2} \cdot e^{-\alpha r}.$$

The application of this law to the Sun-Moon-Earth system, however, leads him to the conclusion that the value for Earth (radius R) would have to be:[71]

$$\alpha R < \frac{1}{10^6}.$$

18. Influence of Temperature

Some mechanical theories about the nature of gravitation[72] make it seem quite possible that the gravitational effect is modified by the temperature of | the medium. A [44]
direct examination of this question has not yet been carried out; however, von Jolly points out that in his absolute determinations, the temperature difference was maximally 29.6°, without any difference in the results exceeding the magnitude of experimental error.

19. Dependence on Time: Constancy

The tacit assumption of the gravitational effect's independence on time in Newton's law has been challenged in two respects:

a. Is the gravitational constant also a constant with respect to time, or does it change over the course of time?

b. Does gravitation need time to take effect—does it have a finite speed of propagation, or is the gravitational effect instantaneous?

---

68 Wilsing uses long cylinders, Boys and Braun use spheres, Richarz and Krigar-Menzel use cubes, nevertheless good agreement, namely:

Wilsing	$\Delta = 5.577$	Difference 0.9%
Boys and *Braun*	$\Delta = 5.527$	" 0.4%.
Richarz and *Krigar-Menzel*	$\Delta = 5.505$	

Cf. in particular note 48.

69 *Phys. Rev.* 5 (1897), p. 294–300.

70 *Méc. Cél.* 5, book XVI, chap. IV, §6.

71 Poynting presents an indirect proof against the existence of a specific gravitational permeability: A deflection (refraction) of the gravitational effect has never been observed. However this question appears not to have been carefully examined to date.

72 Cf. part V of this article.

R. Pictet[73] discussed the first question based on the idea that gravitation is caused by impacts of aether particles.[72] His reasoning is the following: The total energy of the solar system consists of two parts: 1) the *vis viva* of planets and Sun; 2) the *vis viva* of aether particles. The *vis viva* of planets varies strongly depending on their momentary position with respect to the Sun. If the total energy of the solar system remains constant, then it follows that the *vis viva* of aether atoms and thereby the gravitational constant have to change with the course of time.

Experiments to prove such a temporal change of the gravitational constant would have a chance to succeed according to R. Pictet and P. Cellérier,[74] since the difference in the *vis viva* of the planets—the decisive ones are Jupiter and Saturn—e.g., between the minimum of year 1898–99 to the maximum of 1916–17, comes to about 18%.

20. Dependence on Time: Finite Speed of Propagation[75]

The second question, whether gravitation operates instantaneously or has a finite speed of propagation, was examined recently in terms of planetary motion by R. Lehmann-Filhès[76] and J. v. Hepperger.[77] |

[45] Both works introduce a finite propagation velocity in the same way. At the moment when the planet (mass m) is at distance r from the Sun (mass M), the force $G \cdot M \cdot m/r^2$ as per Newton's law is propagated from the Sun with a finite velocity. This force then takes effect on the planet at a time when its distance from the Sun is different from r in direction as well as in magnitude. The same holds for the force that the planet exerts on the Sun.

There is some difference between Lehmann-Filhès and von Hepperger in their equations of motion, as the former takes the Sun's velocity, the latter the velocity of the Sun's and the planet's center of gravity, to be constant.

Both arrive at the result that the most influential change of planetary motion would be a secular change of the mean radius. From this it follows: first, that the introduction of a finite speed propagation *while retaining Newton's law* does not contribute anything to remove the difficulties regarding planetary orbits as presented on p. 86 [p. 36]; and second that the hypothetical propagation velocity would have to be much larger than the velocity of light, because otherwise a secular change of the mean radius would result, by an amount that would contradict observation. If the velocity of the Sun's proper motion lies between 1 and 5 km/sec,[78] then the propaga-

---

73 *Genève Bibl.* (6 sér., 3 période) 7 (1882), p. 513–521.

74 *Genève Bibl.* (6 sér., 3 période) 7 (1882), p. 522–535.

75 Lecture on this question: S. Oppenheim, *Jahresber. kais. kgl. akad. Gymn. Wien* 1894–1895, p. 3–28; F. Tisserand, *Méc. cél.* 4 (1896), chap. 28; F. Drude, *Ann. Phys. Chem.* 62 (1897).

76 *Astr. Nachr.* 110 (1885), p. 208.

77 *Wien. Ber.* 97 (1888), p. 337–362.

78 According to recent examinations, however, this should be appr. 15 km/sec. (cf. H. C. Vogel, *Astr. Nachr.* 132 (1893), p. 80 f.

tion velocity of gravitation would have to be at least 500 times larger than the velocity of light, according to v. Hepperger.

A stricter test of the assumption of a temporal propagation velocity is provided by its application to the Moon's motion, as carried out by R. Lehmann-Filhès.[79] He draws the conclusion that in order to keep the perturbations of the Moon's radius below an acceptable amount while retaining Newton's law, the propagation velocity of gravitation would have to be given an enormous value, perhaps a million times the velocity of light. Also the sign of the perturbation does not correspond to the discrepancy found between observation and theory for the Moon.

Th. v. Oppolzer[47] comes across similar difficulties when | applying the assumed [46]
finite propagation velocity to calculate orbits of comets.

3. EXTENSION OF NEWTON'S LAW TO MOVING BODIES[80]

21. Transferring Fundamental Electrodynamic Laws to Gravitation

The result of the attempts to introduce a finite propagation of gravitation while retaining Newton's law for moving bodies as well, and thereby to remove the existing disagreements between observation and calculation, must be characterized as rather unsatisfactory. It is therefore small wonder that attempts were made to question the validity of Newton's law for moving bodies, to regard it merely as a special case for bodies at rest, and to replace it with an extended law for moving bodies.

Above all it was examined whether the previously known electrodynamic fundamental laws were sufficient for this purpose.

C. Seegers[81] and G. Holzmüller[82] applied Weber's fundamental law, according to which the potential for two mass elements m_1 and m_2 at a distance r is

$$P = \frac{G \cdot m_1 \cdot m_2}{r^2}\left[1 - \frac{1}{c^2} \cdot \left(\frac{dr}{dt}\right)^2\right] \qquad (c = \text{velocity of light}),$$

which as is well known Zöllner thought to be the fundamental law of all action-at-a-distance forces, to planetary motion in general, and the planetary motions were calculated numerically by F. Tisserand[83] and H. Servus.[84] For Mercury, the application of Weber's law results in an anomalous secular perihelion motion of ca. 14''.

---

79 *Münch. Ber.* 25 (1896), p. 371.

80 Reviews of a part of the work in this area in S. Oppenheim,[77] P. Drude,[77] and F. Tisserand.[77]

81 Diss. Göttingen 1804.

82 *Zeitschr. Math. Phys.* 1870, p. 69–91.

83 Paris, *C. R.* 75 (1872), p. 760 and 110 (1890), p. 313.

84 Diss. Halle 1885. F. Zöllner cites (based on correspondence) that W. Schreibner calculated a secular perihelion motion of 6.7″ for Mercury based on Weber's law. The reason for this figure's two-fold deviation from the figures cited in the text is that Schreibner equates the constant c in Weber's law with $\sqrt{2}$ times the velocity of light.

Transferring Gauss'[85] fundamental electrodynamic law to gravitation, in the sense that one introduces an attractive force K between two mass elements with co-ordinates x_1, y_1, z_1 and x_2, y_2, z_2, given by: |

[47]
$$K = \frac{G \cdot m_1 m_2}{r^2}\left\{1 + \frac{2}{c^2}\left[\left(\frac{d(x_1 - x_2)}{dt}\right)^2 + \left(\frac{d(y_1 - y_2)}{dt}\right)^2 + \left(\frac{d(z_1 - z_2)}{dt}\right)^2 - \frac{3}{2}\left(\frac{dr}{dt}\right)^2\right]\right\},$$

gives a secular perihelion motion of Mercury of only 28", according to F. Tisserand's[86] calculation.

Riemann's[87] fundamental law

$$P = \frac{G \cdot m_1 \cdot m_2}{r}\left\{1 - \frac{1}{c^2}\left[\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2\right]\right\}$$ (x, y, z are co-ordinates of m_1 relative to m_2),

would imply, according to M. Lévy,[88] twice the perihelion motion of Mercury that follows from Weber's law.

Therefore, Levy suggested a combination of Riemann's and Weber's laws in the form:

$$\begin{aligned} P &= P_{Weber} + \alpha(P_{Riemann} - P_{Weber}) \\ &= \frac{G \cdot m_1 \cdot m_2}{r}\left\{1 - \frac{1}{c^2}\left[(1-\alpha)\left(\frac{dr}{dt}\right)^2 + \alpha\left(\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2\right)\right]\right\} \end{aligned}$$

where α was then to be determined from the observed secular perihelion motion of Mercury. Assuming the perihelion motion of $38''$ as observed and $14.4''$ [89] as given by Weber's law, one finds $\alpha = 1.64 =$ approx. $5/3$. [88] On the basis of a perihelion motion of Mercury of $41.25''$, as given by other observers, and a motion given by Weber's law of $13.65''$,[89] α becomes $= 2.02$.

The law one obtains in this way has the decisive advantage of matching the achievement of Riemann's and Weber's laws in electrodynamics, and moreover it represents an extension of Newton's law to moving bodies that eliminates the worst disagreement between observation and calculation that has persisted until now.

---

85 *Ges. Werke* 5, p. 616 f., Nachlass.

86 Paris, *C.R.* 110 (1890), p. 313.

87 *Schwere, Elektrizität und Magnetismus*, ed. Hattendorf, Hannover 1896, p. 313 ff.

88 Paris, *C. R.* 110 (1890), p. 545–551. For motion of two masses the law was earlier discussed in general by O. Limann (Diss. Halle 1886).

89 Tisserand[85] (Paris, *C. R.* 75), and Servus.[86]

22. Transferring Lorentz's Fundamental Electromagnetic Equations to Gravitation

H.A. Lorentz[90] has attempted to use Maxwell's equations,[91] as extended by him to moving bodies, for gravitation. His conception | of the constitution of gravitating [48] molecules is essentially in agreement with that of F. Zöllner, though in slightly modernized form. The foundation of Lorentz's approach is covered in section 36.

The additional forces Lorentz obtains, apart from the ones given by Newton's law, have a factor of either $(p/c)^2$ or $(p \cdot w)/c^2$, where p is the velocity of the central body taken to be constant, w is the velocity of the planet relative to the central body, and c is the velocity of light. These additional forces are so small that they will probably be beyond observation in all cases; in the case of Mercury they are certainly below what is observable, as shown by Lorentz's calculation. It follows that Lorentz's equations, combined with Zöllner's conception of the nature of gravitating molecules, can be applied to gravitation,[92] but they do not contribute to removing existing disagreements between observation and calculation.

23. Laplace's Assumption

Previously Laplace[93] envisaged an extension of Newton's law for moving bodies in quite a different way. He seems to imagine the force coming from an attracting body m_1 as a sort of wave, which exerts an attractive force on each body m_2 it encounters of magnitude $G \cdot m_1 m_2 / r^2$ in the direction in which it propagates. The effect such a wave has on a moving body m_2 depends only on the relative motion of wave and body. One can thus imagine body m_2 at rest in space, if one ascribes to the wave another velocity component apart from its velocity in the r-direction, equal and opposite to the velocity of m_2. If $v =$ velocity of m_2 and c indicates the propagation velocity of gravitation, then the body m_2 receives a force component opposite to its orbit's direction and of value $(m_1 m_2 / r^2) \cdot (v/c)$,[94] rather than receiving only a force component in the r-direction. |

Following through with this point of view gives little satisfaction with respect to [49] the planets: it does not result in a perihelion motion at all, but in a secular change of the mean radius; e.g., this change for the Moon has a value such that the lowest limit for c would have to be about 100,000,000 times the velocity of light. It is, however, not uninteresting that Laplace's conception achieves the same effect as a *resistance of the medium* proportional to the velocity of the planet.

---

90 *Amsterdam Versl.*, April 1900.

91 Harlem, *Arch. Néerl.* 25 (1892), p. 363.

92 This includes the possibility that *the propagation velocity of gravitation is equal to the velocity of light*.

93 *Méc. cél.* 4, book X, chap. VII, §19 and 22.

94 These conditions would therefore correspond completely with those for the aberration of light.

According to Encke[95] and v. Oppolzer,[96] a resistance of the medium—however, proportional to the square of the velocity—could perhaps explain the irregularities of Encke's comet presented on p. 36. The anomalies of Winnecke's comet, presupposed by Oppolzer and explained the very same way, have since been shown to be non-existent by E. v. Haerdtl's[97] calculations.

24. Gerber's Assumption

P. Gerber's[98] two premises are:

a. The potential P transmitted from a mass μ to a second one m is $\frac{\mu}{r}$, where r is the distance from μ to m *at the moment of transmitting* the potential. This potential propagates with finite velocity c.

b. A certain time is necessary for the potential "to reach m, to impart itself to the mass; i.e., to evoke in m the state of motion corresponding to the potential." "If the masses are at rest, the motion of the potential passes m with its own velocity; then its value transmitted to m is in inverse proportion to the distance. If the masses speed towards each other, the time of transmission as well as the transmitted potential decrease proportionally to the ratio of the characteristic velocity of the potential to the sum of its and the masses' velocity, as the potential has this total velocity relative to m."

Gerber arrives at the value that the potential must have under these assumptions in the following manner:

"The potential moves with the velocity of the attracting mass in addition to its
[50] own velocity c. The space $| r - \Delta r$[99] traversed in time Δt by the two motions, one of the potential and the other one of the attracted mass, is thus

$$\Delta t\left(c - \frac{\Delta r}{\Delta t}\right),$$

while $r = c\Delta t$. So for the distance where the potential starts developing and to which it is in inverse proportion, one obtains

$$r - \Delta r = r\left(1 - \frac{1}{c}\frac{\Delta r}{\Delta t}\right).$$

Since, moreover, the velocity with which the motions pass each other has the value

$$c - \frac{\Delta r}{\Delta t},$$

95 Cited by von Oppolzer.
96 *Astr. Nachr.* 97, p. 150–154 and 228–235.
97 *Wien. Denkschr.* 56 (1889), p. 179 f.
98 *Zeitschr. Math. Phys.* 43 (1898), p. 93–104.
99 $\Delta r > 0$ for increasing r.

the potential turns out also to be proportional to

$$\frac{c}{c - \frac{\Delta r}{\Delta t}}.$$

due to the time consumed to impart itself to m. Thus one finds:

$$P = \frac{\mu}{r\left(1 - \frac{1}{c}\frac{\Delta r}{\Delta t}\right)^2}.$$

As long as the distance Δr is short and therefore $\Delta r / \Delta t$ small compared to c, one may replace the latter by dr/dt. So it becomes

$$P = \frac{\mu}{r\left(1 - \frac{1}{c}\frac{dr}{dt}\right)^2},$$

from which it follows with help of the binomial law to the second power:

$$P = \frac{\mu}{r}\left[1 + \frac{2}{c}\frac{dr}{dt} + \frac{3}{c^2}\left(\frac{dr}{dt}\right)^2\right]."$$

The application of this equation to planetary motions yields the following remarkable result: If the propagation velocity c is determined from Mercury's observed perihelion motion, then one finds $c = 305.500$ km/sec, which is the velocity of light with surprising accuracy. In other words, *if in Gerber's equation the velocity of light replaces the propagation velocity of gravitation, then this equation yields exactly the observed anomalous perihelion motion of Mercury.*

No difficulties for the other planets follow from Gerber's assumption, | except for [51]
Venus, where Gerber's approach gives a slightly too large secular perihelion motion of $8''$.

Gerber's assumption thus shows, as does Lévy's, that a propagation velocity of gravitation of the same magnitude as the velocity of light is not only possible, but can also serve to eliminate the worst disagreement that has existed between astronomical observation and calculation so far. To be sure, this was achieved only by confining the validity of Newton's Law to bodies at rest and postulating an extended law for moving bodies.

4. EXTENSION OF NEWTON'S LAW TO INFINITELY LARGE MASSES

25. Difficulty with Newton's Law for Infinitely Large Masses

Doubts have been expressed concerning the universal validity of Newton's law leading in quite a different direction, and the necessity of an extension has been considered.

In case the universe contains infinitely many masses, to obtain the force acting at any point one would strictly speaking have to solve the problem: to specify the effect of infinitely many masses of finite size at one particular point.

C. Neumann[100] was probably the first to point out that in this case the forces resulting from Newton's law may become indefinite. H. Seeliger[101] examined this question in a more general way, and showed that for infinite masses Newton's law can produce infinitely large forces as well as leaving them completely indefinite.

26. Elimination of the Difficulty by Altering the Law of Attraction

Seeliger suggests a slight modification of Newton's law in order to eliminate this difficulty, and he discusses various possibilities.

The form already discussed by Laplace

$$K = \frac{G \cdot m_1 \cdot m_2}{r^2} \cdot e^{-\alpha r}$$

[52] | is physically expected to suffice for the above purpose, as it corresponds to the assumption of absorption by the medium. In fact, it does suffice, and would moreover have the advantage of giving planetary perihelion motion. Yet the value of $\alpha = 0.00000038$ taken from Mercury's observed perihelion motion gives perihelion motions for the other planets which are difficult to reconcile with observations.[102]

The laws discussed by C. Neumann, according to which the potential P takes the form

$$P = G \cdot m_1 m_2 \left(\frac{Ae^{-\alpha r}}{r} + \frac{Be^{-\beta r}}{r} + \ldots \right),$$

serve the same purpose, but the resulting perihelion motions of the planets stand in severe contradiction to observation.

In contrast, Green-Hall's law discussed earlier,

---

100 *Leipz. Abh.* 1874.

101 *Astr. Nachr.* 137 (1895), p. 129–136; *Münchn. Ber.* 26 (1896), p. 373–400. Controversy between J. Wilsing and H. Seeliger about this issue, *Astr. Nachr.* 137 and 138.

102 *Münchn. Ber.* 26 (1896), p. 388.

$$P = \frac{G \cdot m_1 m_2}{r^{1+\lambda}},$$

which would be suitable to account for the perihelion motions of the planets, retains the same problems as Newton's law with regards to infinite masses.

27. Elimination of the Difficulty by Introducing Negative Masses

A. Föppl[103] introduced the idea that the difficulty of Newton's law emphasized by Neumann and Seeliger is to be eliminated by introducing "negative masses" and maintaining the law, rather than by altering the form of the law. As with the gravitational force lines emitted by the familiar positive masses, there would be force lines flowing into negative masses. If the sum of the negative masses is taken to equal that of the positive ones, the total would be 0; as in the electric and the magnetic domain there would be the same number of sources and sinks.

With this assumption, the expression for field energy cannot be based upon the usual one,

$$\frac{1}{2} a |\mathfrak{R}|^2 dS,$$

where a is a constant of the medium, $\mathfrak{R}$ the vector of field strength defining the gravitational field, $|\mathfrak{R}|$ its | absolute value, and dS a volume element. Rather, as Maxwell [53]
already pointed out, one must replace this expression by

$$\left(C - \frac{1}{2} a |\mathfrak{R}|^2\right) dS \qquad (C = \text{constant})$$

to obtain an attraction between masses of the same sign. Cf. section 34 regarding the significance of the constant C.

Prior to Föppl, C. Pearson[122] had already suggested the mere introduction of negative masses of the same magnitude as the familiar positive masses. This suggestion is actually a consequence of his theory, which attempts to derive electrical, optical, chemical and gravitational phenomena from suitably chosen aether motions.

The introduction of negative masses hardly causes any problems. For the fact that repulsion between two masses has never been observed, i.e. a negative mass has never been noticed, points to the possibility—though not to the necessity—that such masses were driven to spaces no longer accessible to observation due to the repulsion from positive masses in our system. On the other hand, according to A. Schuster,[104] who had the same thought (though merely in a "holiday dream"), the introduction of negative masses could perhaps serve to shed completely new light upon several phenomena, such as comet tails.

---

103 *Münchn. Ber.* 27 (1897), p. 93–99.
104 *Nature* 58 (1898), p. 367 and 618.

5. ATTEMPTS TO EXPLAIN GRAVITATION THROUGH MECHANICS[105]

28. Pressure Differences and Currents in the Aether[106]

The conjecture that gravitation could be caused by *pressure differences* in the supposedly homogeneous aether surrounding gravitating | masses stems from Newton[107] himself. According to him the aether would become denser the further it is from masses. Since each body has the tendency—later on he speaks of an elastic force of the medium—to go from the denser parts of the medium to the less dense ones, each of the two bodies must move in the direction of the other.
[54]

Similar ideas have been worked out by Ph. Villemot,[108] L. Euler,[109] J. Herapath[110] and in a slightly different way by J. Odstrčil.[110]

A consequence of the assumption of pressure differences in the aether, combined with the idea that aether behaves like a fluid or a gas, is that *aether currents* must flow into the atoms.[3] According to J. Bernoulli,[108] B. Riemann,[111] and J. Yarkovski,[112] it is these aether currents which carry the body along and hence cause gravitation. G. Helm,[150] as well as C. Pearson,[113] arrived at a similar conception while trying "to explain gravitation with energy transfer in the aether."

Yarkovski pondered the question of the *cause* of the aether currents, but produced an explanation that is physically not tenable.

Among the many objections raised against these theories, there is also the question of what happens to the aether that flows into the atoms. There are only two possible answers: either the aether accumulates or it disappears inside them. Bernoulli, Helm, Yarkowski have decided for the former possibility; Riemann for the latter, who allows matter in ponderable bodies constantly to make a transition "from the physical world into the spiritual world."

---

105 Review articles: W.B. Taylor, *Smithson. Inst. Rep. for 1876* (1877), p. 205–282: Detailed discussion of papers up to 1873. C. Isenkrahe, a) *Isaac Newton und die Gegner seiner Gravitationstheorie* etc., *Progr. Gymn. Crefeld*, 1877–1878. b) *das Rätsel von der Schwerkraft*, Braunschweig 1879. c) *Zeitschr. Math. Phys.* 37, Suppl. (1892), p. 161–204; P. Drude[77]; partly also H. Gellenthin, "Bemerkungen über neuere Versuche, die Gravitation zu erklären etc.", *Progr. Realgymn. Stettin* 1884 and Gehler,[2] Articles: *Anziehung, Materie*.

106 The term "aether" is not always used with the same meaning in the following text; also in the original papers it is not always sufficiently defined. What is meant roughly in each case, is given by the context.

107 According to W. B. Taylor,[107] Newton expressed this view in a letter and repeated it in his *Optice*.

108 Cf. Taylor.[107]

109 Cf. Taylor[107] and especially Isenkrahe.[107]

110 *Wien. Ber.* 89 (1884), p. 485–491.

111 *Ges. Werke*, 2nd ed. 1853, p. 529.

112 *Hypothèse cinétique de la gravitation universelle* etc. Moscou 1888.

113 *Amer. J. of math.* 13 (1898), p. 419.

29. Aether Vibrations

The idea that aether vibrations in the form of longitudinal waves may cause not only the phenomena of light and heat, but also gravitation, has been developed in two directions. |

1. According to the first view the attracting body, or its atoms, are supposed to [55]
vibrate themselves; these vibrations pass on to the aether, propagate to the attracted body and cause its approach.

Hooke,[114] Newton's inventive rival, already expressed this conception, which was taken up again by J. Guyot and F. Guthrie. The latter two seem to have arrived at this through the observation that light objects close to a vibrating body are pushed towards it. However, the fact that the approach takes place only under very particular conditions, and that under different conditions one observes an apparent repulsion—such was also included by F.A.E. and E. Keller[115] for explaining gravitation—proves that the assumption of an elastic aether and vibrating atoms does not suffice to explain gravitation. There must be at least one more assumption which produces conditions that guarantee an *attraction* under all circumstances.

To find these conditions, J. Callis[116] examined the following question analytically and in detail: What effect do longitudinal waves in a fluid whose pressure changes proportionally to changes in density, have on small inelastic, smooth spheres embedded in the elastic fluid medium? He reaches the conclusion that if the wave length is large compared to the spheres' radius, the spheres are then pushed towards the center of the spherical wave. For explaining gravitation, one would thus need to assume that there are vibrations whose wave lengths in the aether are large compared to the dimensions of the gravitating atoms.

A deficiency of this treatment is the requirement that only the attracting body emits waves. Such a principled differentiation between attracting and attracted body is incompatible with the nature of gravitation. The question must not be what effect do spherical waves have on bodies at rest, but rather what effect do they have on a body which is itself vibrating.

This complete problem was probably first handled mathematically by C.A. |
Bjerknes,[117] for the case of an incompressible aether and pure pulsation of the [56]
spheres (atoms). He proved that two pulsating spheres, whose radius is small compared to their separation, show an apparent attraction, and that this attraction is proportional to the intensity of pulsation and inversely proportional to the square of distance, if their pulsations agree in frequency and phase. If gravitation is to be attributed to pulsating atoms and molecules,[2] then at least the following additional assumptions are needed:

---

114 Cf. W.B. Taylor[105] and F. Rosenberger.[1]

115 Paris, *C. R.* 56 (1863), p. 530–533; also cf. Taylor.[105]

116 E.g. *Phil. Mag.* (4) 18 (1859), p. 321–334 and 442–451, cf. Taylor[105] about other work by Callis.

117 Cf. the compilation in V. Bjerknes, "Vorlesungen über hydrodynamische Fernkräfte nach C.A. Bjerknes' Theorie", Leipzig 1900.

a. The pulsations of all atoms or molecules must agree in frequency and phase.

b. The intensity of pulsation must be proportional to the mass.

There is one more thing. A.H. Lealy[118] pointed out that for the case of a *compressible* fluid, the effect of two spheres pulsating with equal phase and frequency reverses its sign if the distance between them exceeds half a wave length. So if one wants to use Bjerknes' results for gravitation, one would have to suppose either that the aether is *completely* incompressible (Bjerknes) or that it has such low compressibility that half the wave length of aether vibrations is larger than the distances for which observations have established the validity of Newton's law (A. Korn[119]). Only then is attraction always guaranteed, in agreement with observation.

Bjerknes' conception received further development by C. Pearson[120] and in the work by A. Korn just mentioned. The latter extended these ideas mainly to electromagnetic phenomena, the former to phenomena of optics and molecular physics, assuming complicated modes of vibration of the atoms. In his last paper, Pearson abandoned the assumption of oscillations for gravitation and only retained this
[57] assumption for optics and molecular physics, while replacing the | pulsating atoms by places in the incompressible aether at which aether continuously flows in and out in an oscillatory manner ("aether squirts"). For gravitation, he assumes that there is a constant flow in addition to the oscillating one at the locations concerned. With this requirement, the assumed incompressibility of the aether leads directly to the conclusion that apart from places of emission (source points, ordinary masses) there must be just as many places of absorption (sink points, "negative masses").[121]

Using Bjerknes' results for explaining gravitation suffers from the obvious deficiency that assumptions are required which would first have to be explained themselves. Attempts to supply real reasons have been made for only *one* of these assumptions, the synchronous pulsation of atoms. J.H. Weber[122] points out that in the attempt to demonstrate Bjerknes' results the synchronization of the two spheres happens quickly "on its own"; i.e., due to the forces which are caused by the vibrations in the fluid, even if the pulsations were not synchronous at first. From this he concludes that if atoms pulsate at all, the pulsations should become synchronous "on their own" (as specified above).

According to Korn, the assumption of synchronous pulsation can be replaced by another one, which is that the whole solar system is exposed to a periodic pressure. This assumption may be preferred due to its simplicity, but this is the only advantage over Bjerknes's assumption.

---

118 *Cambr. Trans.* 14 (1) (1885), p. 45, 188.

119 "Eine Theorie der Gravitation und der elektrischen Erscheinungen auf Grundlage der Hydrodynamik", 2nd ed., Berlin 1898.

120 *Quart. J.* 20 (1883), p. 60, 184; *Cambr. Trans.* 14 (1889), p. 71 ff.; *Lond. math. Proc.* 20 (1888–1889), p. 38–63; *Amer. J. of math.* 13 (1898).

121 See section 27 of this article.

122 *Prometheus* 9 (1898), p. 241-244, 257–262.

2. The second class of attempts to base an explanation of gravitation on aether vibrations assumes that the atoms are not themselves vibrating, but that their activity consists only in a kind of shielding or absorption of aether vibrations.

Representatives of this view include F. and E. Keller,[115] Lecoq de Boisbaudran,[123] and, in a slightly different way, N. von Dellinghausen.[124]

30. Aether Impacts: The Original Ideas of Le Sage

The starting point of all aether impact theories is an idea which | Le Sage[125] developed in a particularly clear and skillful way. According to him the gravitational aether surrounding the atoms of a body consists of discrete particles—"corpuscules ultramondains"—which zoom about in all directions with the same extraordinarily high velocity. No continuous motion is imparted to a single atom embedded in this aether due to the impacts of these aether particles, since the effect of aether impacts from all directions cancels out. But if two atoms A_1 and A_2 are brought into this aether, the conditions change in two respects: [58]

a. A_2 shields A_1 from a part of the aether atoms: The side of A_1 turned towards A_2 is hit by fewer aether particles than its side turned away from A_2. The consequence would have to be that A_1 is driven towards A_2 by the action of the aether impacts, and conversely A_2 is driven towards A_1.

If the atoms are assumed to be very large in comparison to the aether particles, it follows directly that this *shielding effect* of one atom of a body on another decreases with the square of distance. To make the shielding proportional to mass, Le Sage introduces the assumption that the gravitating masses are extraordinarily porous[126] to the aether particles so that the efficacy of the whole body becomes proportional to the number of atoms it contains.[127]

b. Due to the *reflection* of aether particles on A_2, a number of aether particles also hit the atom A_1 that would not have hit A_1 without the presence of A_2.[128] If these reflected aether atoms had the same velocity as those hitting A_1 directly, then they would cancel out the approach of A_1 towards A_2 caused by the shielding effect of A_2; thus, gravitation would not be produced.

---

123 Cf. Paris, *C. R.* 69 (1869), p. 703–705; cf. Taylor.[107]

124 "Die Schwere oder das Wirksamwerden der potentiellen Energie," *Kosmos* 1, Stuttgart 1884. Cf. C. Isenkrahe.[107]

125 *Berlin Mém*. 1782 and in P. Prévost, *Deux traîtés de Physique mécanique*, Paris 1818. In the last paper it is quoted that similar theories have been established before (by Nicolas Fatio and F. A. Redecker).

126 Strangely enough, Le Sage extends the assumption of very high porosity to every single atom of a body and therefore arrives at the conception of the peculiar "box atoms" [*Kastenatome*].

127 But cf. section 32, c).

128 In P. Drude[77] we find the note that Le Sage simply ignores reflection and thus his observation lacks rigor. This is probably a mistake: Le Sage devotes chapter IV to reflection in P. Prévost.

This is why Le Sage assumes furthermore that the aether particles are *absolutely*
[59] *inelastic*—"privé de toute élasticité"—and | states that under this assumption the average velocity of the reflected atoms = $2/3$ of the non-reflected ones.[129]

Thus the difference of effects a) and b) still results in both atoms approaching each other.

31. Aether Impacts: Further Development of Le Sage's Theory

Recently, Le Sage's theory was primarily defended by C. Isenkrahe, who emphasized in particular the assumption that collisions between the aether particles and atoms are subject to the laws of *inelastic* collision. Isenkrahe's progress beyond Le Sage consists of the following points:

a. He ascribes to gravitational aether the properties of a gas in the sense of kinetic gas theory. Hence he is giving up the assumption of *equal* velocity[130] of the aether atoms.

b. He does not explain the porosity of the body to aether particles by porosity of the atoms themselves, but by assuming that the distance between atoms[131] is large compared to their dimensions.

c. To achieve the proportionality between attraction and mass, which was guaranteed by Le Sage's assumption only for bodies of the same composition, he assumes that "the final components of matter are all of equal size; they may be the aether atoms themselves."

A. Rysáneck's[132] assumptions are very similar. His achievement consists of the precise implementation of the ideas of kinetic gas theory.[133] In his calculations he actually takes into consideration that the velocities of aether atoms are distributed according to Maxwell's law, whereas e.g. Isenkrahe does assume different velocities of aether atoms, but replaces them by *one* average velocity in all his derivations.

[60] Prior to Isenkrahe S.T. Preston[134] | already pointed out that Le Sage's conceptions could be suitably replaced by ideas of kinetic gas theory, if the mean free path of aether atoms is assumed to be of the order of planetary distances. He developed this idea in several papers, though without going into details as carefully as Isenkrahe and Rysáneck.

---

129 About justification and validity of this information cf. C. Isenkrahe[107] in paper b., p. 155 ff.

130 Which Le Sage also chose merely for simplicity, as he explicitly points out the *different* velocities in the reflections of aether particles and atoms.

131 Which are assigned a spherical form for simplicity.

132 *Repert. Exp.-Phys.* 24 (1887), p. 90–115.

133 But cf. section 33.

134 *Phil. Mag.* (5) 4 (1877); *Wien Ber.* 87 (1882); *Phil. Mag.* (5) 11 (1894); *Diss. München* 1894.

32. Aether Impacts: Difficulties of these Theories

a. A necessary condition for a gravitational effect is that aether atoms lose translational velocity upon collision with atoms, which is achieved most easily by assuming inelastic collisions.

However, this assumption leads to the problem of where the energy lost at impact goes. P. Leray[135] and later P.A. Secchi,[136] W. Thomson,[137] S.T. Preston,[134] then A. Vaschy,[138] Isenkrahe himself and Rysáneck tried to avoid this problem in many different ways. None of these attempts, however, is itself unobjectionable.[139]

b. J. Croll[140] turns against the assumption made in most aether impact theories, which is that the distance between two molecules is very large compared to their dimensions, or rather compared to their spheres of action. He notes that this assumption grossly contradicts W. Thomson's estimates regarding the size of the molecules and their number per unit volume.

c. Objections can be raised from a different angle against the assumption of high porosity of the body for the aether atoms. If the porosity is presumed to be so large that the aether atoms that passed one layer of a body hit the next layer with completely undiminished velocity, then the proportionality between attraction and mass would be strictly preserved. At the same time, this requirement excludes any attraction at all. Hence, one must assume that the aether atoms forfeit a noticeable amount of their energy when passing a body layer. A.M. Bock[141] has shown that this assumption | is not incompatible with the required strict proportionality between attraction [61]
and mass.

d. Bock pointed out one more problem. If a third mass comes between two masses, then the attraction of the two masses is considerably modified, such that the third mass seems to have a larger permeability, as shown by a mathematical examination of this case based on aether impact theories. Because this case is not rare, e.g., for the Moon, Earth and Sun, there would have to be observable perturbations over the course of time. But in fact no perturbations of this type have ever been observed.

e. Le Sage already discussed another objection against aether impact theories. If any body, e.g. a planet, moves in an aether with the assumed properties, then it must experience resistance. But none has been observed for the planets.

The last question was examined more precisely by Rysáneck, Bock and W. Browne[142] on the basis of astronomical data.[143] Since the secular changes of

---

135 Paris, *C.R.* 69 (1869), p. 615–621; also cf. Taylor.
136 Cited in Isenkrahe.[107b]
137 *Phil. Mag.* (4) 45 (1871), p. 321–332.
138 *J. de Phys.* (2) 5 (1886), p. 165–172.
139 Cf. C. Isenkrahe[107b]; Maxwell, *Encycl. Brit.*, 9th ed. Article: *Atom und Scient. Pap.* 2, p. 445, Cambridge 1890.
140 *Phil. Mag.* (5) 5 (1877), p. 45–46.
141 Diss. München 1891. Isenkrahe[107b] has already examined this question, though not fully.
142 *Phil. Mag.* (5) 10 (1894), p. 437–445.
143 Cf. also section 23.

planetary orbits give an upper limit on this hypothetical resistance, the aether impact theories yield a lower limit on the velocity of the aether atoms, if their density is assumed to be known. By using a density of the same order of magnitude as has been estimated for the optical aether, one obtains enormous numbers as the lower limit on the mean velocity. Rysáneck, e.g. based on calculations on Neptune's orbit, obtained the number $5 \cdot 10^{19}$cm/sec.

f. Of all the objections that P. du Bois-Reymond[144] raised against the aether impact theories, one is particularly noteworthy.

Think of a ponderable truncated cone (cross-section $ABCD$) with a molecule α close to the top. According to the aether impact theories, the acceleration which α receives towards the cone is the difference between the effect the aether atoms with solid angle ω_1 and the effect the aether atoms with solid angle ω_2 have on the molecule. The first effect remains unaltered, but the second decreases if R, the distance
[62] between base CD and cone top O, increases. | Therefore, the total effect always remains smaller than the effect of the aether atoms of solid angle ω_1.

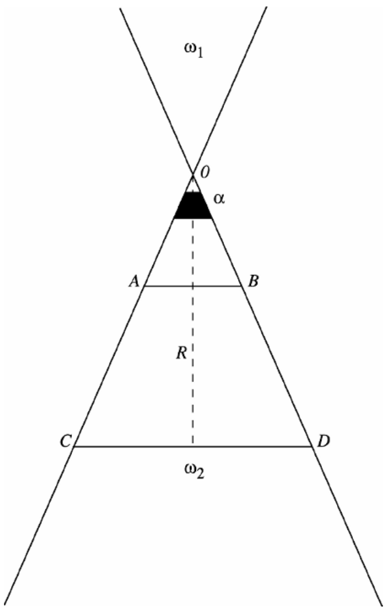


Figure 1

On the other hand, since according to Newton's law the attraction of the cone on α increases with R and exceeds any specifiable number, if the same is assumed about R, there are then only two possibilities: either to presume that the effect of

144 *Naturw. Rundschau* 3 (1888), p. 169–178.

aether atoms in space ω_1 on molecule α is infinitely large, or to assume that Newton's law no longer holds for infinitely extended masses.[145]

Isenkrahe[146] countered P. du Bois-Reymond's objection with the latter assumption. However, the difficulty persists, in that one must ascribe an enormous, if not infinite, magnitude to the aether atom's effect, which leads to shortcomings in other areas.[147]

33. Aether Impacts: Jarolimek's Objections and Theory

A. Jarolimek[148] emphasized a deficiency of all aether impact theories that suppose aether to be a gas in the sense of kinetic gas theory. The derivation of the law of gravity in these theories is based on a simple calculation with a certain mean free path of the aether atoms, which fails to take the variety of paths into account.

Regarding this point, Jarolimek notes that *in order to produce mutual attraction of two molecules, only those aether atoms whose actual path length is larger than the distance between the two molecules can be effective*. Thus it depends precisely upon the *absolute* |—*and not the average*—path length. But if one takes the variety of [63] absolute path lengths into account, under the usual conditions of aether impact theory one does not obtain Newton's law at all.

Referring to Isenkrahe's[149] assumption that the atoms of a body may themselves be an aggregate of the extremely fine aether atoms, Jarolimek points out one more difficulty: this assumption contradicts a shielding effect of two body elements that decreases with the square of the distance. If these elements are identical to aether atoms, then an element could shield another only from those aether atoms whose center lies exactly on the line connecting the two elements; the shielding effect would therefore not depend on the distance at all, if the distance is so large compared to the element's radius that the element could be regarded as devoid of size.

Jarolimek establishes the following theory based on such considerations. He keeps Isenkrahe's assumption—the ultimate elements of atoms are identical to gravitational aether atoms. This practically fully frees him from a shielding effect. He arrives at a decrease of the gravitational effect with the square of the distance in the following way: "One has to think of the infinite number of *swarming* aether atoms as uniformly distributed in space at every instant, and one has to imagine that atoms *bouncing off* from one point fly off in straight lines in all directions. Considering a cone bundle, whose vertex is the point of origin and whose cross-section accordingly increases in quadratic proportion with distance to the vertex and which therefore contains at increasing distance *more uniformly distributed aether atoms* in quadratic proportion, one must realize that for atoms that bounced off (of which a *definite* number

---

145 Cf. section 4.

146 In the book: *Über die Fernkraft und das durch P. du Bois-Reymond aufgestellte* etc., Leipzig 1889.

147 The field representation leads to a similar problem (see section 34).

148 Wien. Ber. 88^2 (1883), p. 897–911.

149 Cf. section 31.

passes the observed cone bundle from the vertex), the probability to hit another atom in space must increase in quadratic proportion to the distance between the two.

From this it follows directly that the number of freely and linearly moving atoms decreases with increasing distance in quadratic proportion, or in other words: *that the aether contains* n^2 *times as many atoms with free path* r *as with free path* nr." Con-
[64] sequently, "*the simplest explanation for the law of gravity | is provided by the inequality of the path lengths of the aether molecules*."

6. REDUCTION OF GRAVITATION TO ELECTROMAGNETIC PHENOMENA

34. Gravitation as a Field Effect

Before reporting on explanatory attempts based on electromagnetism, the empirical facts contained in Newton's law will be mathematically formulated by describing the "gravitational field" without reference to any particular conception of its nature.[150]

One is used to regarding Newton's law as the most distinguished example of an action at a distance. On the other hand, it must be emphasized that its content can be formulated by the following statement, which corresponds to the field theory standpoint: "*The field strength of gravitation is irrotational, and source-free in those regions of space where there are no masses. Where there are masses, the divergence of the field strength is proportional to the local mass density* ρ."

We understand by field strength the attractive force exerted on a *unit mass*; the force exerted on *mass* m_1 is m_1-times as large as the field strength. The proportionality factor for the divergence of the field strength is identical to $4\pi G$. In formulas,[151] the expression of our description of the gravitational field takes the following form, if $\mathfrak{R}$ stands for the vector of field strength:

$$\mathrm{curl}\,\mathfrak{R} = 0, \qquad \mathrm{div}\,\mathfrak{R} = 0 \quad \text{or} \quad = -4\pi G\rho.$$

This formulation and the classical formulation given in section 1 are mathematically exactly equivalent; in particular, it follows from the above differential equations according to the laws of potential theory that the field strength due to a single mass m_2 at distance r is calculated to be

$$\mathfrak{R} = \mathrm{grad}\frac{m_2 G}{r}.$$

This yields the following value of the field strength (or its magnitude in the r-direction) in agreement with Newton's law:

---

150 Field representations of particular kinds are given by G. Helm, *Ann. Phys. Chem.* 14 (1881), p. 149; O. Heaviside, *Electrician* 31 (1893), p. 281 and 359.

151 Because of the significance of vector symbols rot, div, grad, cf. the beginning of the 2nd semi-volume V of the encyclopedia.

$$|\mathfrak{R}| = \frac{d}{dr}\frac{m_2 G}{r}.$$

| So far the field interpretation of gravitation offers neither advantages nor disadvantages over the action-at-a-distance interpretation. An advantage of the former would arise if a finite propagation velocity of gravitational effects could be proven with certainty, especially if it turned out to be equal to the velocity of light. Then the above differential equations for stationary gravitational effects would have to be extended to the case of a time-varying gravitational effect, which could easily be patterned after the example of the electromagnetic equations. On the other hand, the field interpretation involves a serious problem which Maxwell[152] pointed out. Enquiring about the gravitational energy contained in a volume element dS of the field, one must, in order to get an *attraction* of masses of the same sign, assume the form [65]

$$\left(C - \frac{1}{2}a|\mathfrak{R}|^2\right)dS$$

for this energy, where the constant a is identical to $1/4\pi G$. The constant C would have to be larger than $(a/2)|\mathfrak{R}'|^2$, so that the gravitational energy has a positive value throughout, where $|\mathfrak{R}'|$ stands for the largest value of field strength at any point in space. However, it would follow from this that at points of vanishing field strength, e.g. between the Earth and Sun at the point where the Sun's and the Earth's attraction compensate each other, the energy content of space would have to have the enormous quantity C per unit volume. Maxwell adds that he can not possibly imagine a medium with such a property.

35. Electromagnetic Vibrations

The hypothesis that gravitation could be caused by aether vibrations, already discussed in section 29, was examined by H.A. Lorentz[92] under the following assumptions:

a. The gravitating molecules consist of ions possessing an electric charge.

b. The aether vibrations are electromagnetic vibrations whose wave length is small compared to all those distances over which Newton's law is still valid.

Lorentz arrives at this result: an attraction is possible under these conditions only if | electromagnetic energy flows continuously into the volume elements that contain [66]
gravitating molecules. If the assumptions are changed so that such a disappearance of electromagnetic energy is avoided, then no attractive forces are obtained. This is why Lorentz himself dismisses this theory and in the further course of his discussion joins Mossotti-Zöllner's conception (see below).

---

152 *Lond. Trans.* 155 (1865), p. 492 = *Scient. Papers* 1, p. 570, Cambridge 1890.

36. Mossotti's Assumption and its Modern Development

In a completely different direction, O.F. Mossotti[153] tried to reduce gravitation to electrical forces, apparently following Aepinus. He assumes that a repulsion takes place between two molecules and likewise between two "aether atoms," but that there is an attractive force between a molecule and an aether atom which exceeds the repulsion between two molecules or two aether atoms. This assumption provides an attraction between two molecules embedded in the aether, as required by Newton's law.

This idea was simplified by F. Zöllner.[154] He imagines that each gravitating molecule or atom consists of one negatively and one positively charged particle, and assumes that the repulsion between two equal charges is smaller than the attraction between two unequal ones of the same size.

Zöllner's assumption was examined mathematically by W. Weber[155] based on his electrodynamic fundamental law. This was applied to moving bodies only recently by H.A. Lorentz,[90] using his generalized Maxwell equations (cf. section 22). Lorentz's approach is continued in a paper by W. Wien.[156] Cf. the end of article 14 of this volume concerning this most recent phase of the gravitational problem.

As attractive as the explanatory attempts based on electromagnetism, in particular, may appear today, it behooves one to wait, since little of the subject has been worked
[67] out, to see if tangible advantages arise from it for understanding the | gravitational effect and for release from persistent problems. According to section 22, it seems that not much can be gained in this direction by the electromagnetic conception.

For the time being, one will have to summarize the above considerations as follows: all attempts to connect gravitation with other phenomena in a satisfying way are to be regarded as unsuccessful or as not yet adequately established. With this, however, one has, at the beginning of the 20th century, returned to the view of the 18th century, to the view that takes gravitation to be a fundamental property of all matter.

EDITORIAL NOTES

[1] In the original, Zenneck mistakenly refers to note 46 (44 according to our numbering), rather than 36.

[2] This reference is to volume 6 of the *Encyklopädie der mathematischen Wissenschaften*, which covers astronomy.

[3] The terms "Körperatom" and "Körpermolekül" have been translated as "atom" and "molecule" throughout this article, whereas "Ätheratom" and "Äthermolekül" have been translated as "aether atom" and "aether molecule."

---

153 Sur les forces qui régissent la constitution intérieure des corps, Turin 1836.

154 *Erklärung der universellen Gravitation aus den statischen Wirkungen der Elektrizität*, Leipzig 1882.

155 Cf. F. Zöllner.[156]

156 Über die Möglichkeit einer elektromagnetischen Begründung der Mechanik, *Arch. Néerl.* 1900.

HENDRIK A. LORENTZ

CONSIDERATIONS ON GRAVITATION

Originally published in Proceedings Royal Academy Amsterdam 2 (1900), pp. 559–574. A Dutch version appeared under the title "Beschouwingen over de zwaartekracht" in Verslag Koninklijke Akademie van Wetenschapen te Amsterdam 8 (1900), pp. 603–620. A French version appeared under the title "Considérations sur la Pesanteur" in Archives néerlandaises 7 (1902), pp. 325–338.

§1. After all we have learned in the last twenty or thirty years about the mechanism of electric and magnetic phenomena, it is natural to examine in how far it is possible to account for the force of gravitation by ascribing it to a certain state of the aether. A theory of universal attraction, founded on such an assumption, would take the simplest form if new hypotheses about the aether could be avoided, i. e. if the two states which exist in an electric and a magnetic field, and whose mutual connection is expressed by the well known electromagnetic equations were found sufficient for the purpose.

If further it be taken for granted that only electrically *charged* particles or ions, are directly acted on by the aether, one is led to the idea that every particle of ponderable matter might consist of two ions with equal opposite charges—or at least might contain two such ions—and that gravitation might be the result of the forces experienced by these ions. Now that so many phenomena have been explained by a theory of ions, this idea seems to be more admissible than it was ever before.

As to the electromagnetic disturbances in the aether which might possibly be the cause of gravitation, they must at all events be of such a nature, that they are capable of penetrating all ponderable bodies without appreciably diminishing in intensity. Now, electric vibrations of extremely small wave-length possess this property; hence the question arises what action there would be between two ions if the aether were traversed in all directions by trains of electric waves of small wave-length.

The above ideas are not new. Every physicist knows Le Sage's theory in which innumerable small corpuscula are supposed to move with great velocities, producing gravitation by their impact against the coarser particles of ordinary ponderable matter. I shall not here discuss this theory which is not in harmony with modern physical views. But, when it had been found that a pressure against a body may be produced as well by trains of electric waves, by rays of light e. g., as by moving projectiles and when the Röntgen-rays with their remarkable penetrating power had been discovered,

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 3
Gravitation in the Twilight of Classical Physics: Between Mechanics, Field Theory, and Astronomy.

it was natural to replace Le Sage's corpuscula by vibratory motions. Why should there not exist radiations, far more penetrating than even the X-rays, and which might therefore serve to account for a force which as far as we know, is independent of all intervening ponderable matter?

[560] I have deemed it worthwhile to put this idea to the test. In | what follows, before passing to considerations of a different order (§5), I shall explain the reasons for which this theory of rapid vibrations as a cause of gravitation can not be accepted.

§2. Let an ion carrying a charge e and having a certain mass, be situated at the point $P(x,y,z)$; it may be subject or not to an elastic force, proportional to the displacement and driving it back to P, as soon as it has left this position. Next, let the aether be traversed by electromagnetic vibrations, the dielectric displacement being denoted by $\mathfrak{d}$, and the magnetic force by $\mathfrak{H}$, then the ion will be acted on by a force

$$4\pi V^2 e\mathfrak{d},$$

whose direction changes continually, and whose components are

$$X = 4\pi V^2 e\mathfrak{d}_x, \qquad Y = 4\pi V^2 e\mathfrak{d}_y, \qquad Z = 4\pi V^2 e\mathfrak{d}_z. \tag{1}$$

In these formulae V means the velocity of light.

By the action of the force (1) the ion will be made to vibrate about its original position P, the displacement (x, y, z) being determined by well known differential equations.

For the sake of simplicity we shall confine ourselves to simple harmonic vibrations with frequency n. All our formulae will then contain the factor, $\cos nt$ or $\sin nt$ and the forced vibrations of the ion may be represented by expressions of the form

$$\begin{aligned} x &= ae\mathfrak{d}_x - be\dot{\mathfrak{d}}_x, \\ x &= ae\mathfrak{d}_y - be\dot{\mathfrak{d}}_y, \\ x &= ae\mathfrak{d}_z - be\dot{\mathfrak{d}}_z, \end{aligned} \tag{2}$$

with certain constant coefficients a and b. The terms with $\dot{\mathfrak{d}}_x$, $\dot{\mathfrak{d}}_y$ and $\dot{\mathfrak{d}}_z$, have been introduced in order to indicate that the phase of the forced vibration differs from that of the force (X, Y, Z); this will be the case as soon as there is a resistance, proportional to the velocity, and the coefficient b may then be shown to be positive. One cause of a resistance lies in the reaction of the aether, called forth by the radiation of which the vibrating ion itself becomes the center, a reaction which determines at the same time an apparent increase of the mass of the particle. We shall suppose however that we have kept in view this reaction in establishing the equations of motion, and in
[561] assigning their values to the coefficients a and b. | Then, in what follows, we need only consider the forces due to the state of the aether, in so far as it not directly produced by the ion itself.

Since the formulae (2) contain e as a factor, the coefficients a and b will be independent of the charge; their sign will be the same for a negative ion and for a positive one.

Now, as soon as the ion has shifted from its position of equilibrium, new forces come into play. In the first place, the force $4\pi V^2 e\mathfrak{d}$ will have changed a little, because, for the new position, $\mathfrak{d}$ will be somewhat different from what it was at the point P. We may express this by saying that, in addition to the force (1), there will be a new one with the components

$$4\pi V^2 e\left(x\frac{\partial \mathfrak{d}_x}{\partial x} + y\frac{\partial \mathfrak{d}_x}{\partial y} + z\frac{\partial \mathfrak{d}_x}{\partial z}\right), \text{etc.} \tag{3}$$

In the second place, in consequence of the velocity of vibration, there will be an electromagnetic force with the components

$$e(\dot{y}\mathfrak{H}_z - \dot{z}\mathfrak{H}_y), \text{etc.} \tag{4}$$

If, as we shall suppose, the displacement of the ion be very small, compared with the wave-length, the forces (3) and (4) are much smaller than the force (1); since they are periodic—with the frequency $2n$, —they will give rise to now vibrations of the particle. We shall however omit the consideration of these slight vibrations, and examine only the mean values of the forces (3) and (4), calculated for a rather long lapse of time, or, what amounts to the same thing, for a full period $2\pi/n$.

§3. It is immediately clear that this mean force will be 0 if the ion is *alone* in a field in which the propagation of waves takes place equally in all directions. It will be otherwise, as soon as a second ion Q has been placed in the neighborhood of P; then, in consequence of the vibrations emitted by Q after it has been itself put in motion, there may be a force on P, of course in the direction of the line QP. In computing the value of this force, one finds a great number of terms, which depend in different ways on the distance r. We shall retain those which are inversely proportional to r or r^2, but we shall neglect all terms varying inversely as the higher powers of r; indeed, the influence of these, compared with that of the first mentioned terms will be of the order λ/r if λ is the | wave-length, and we shall suppose this to be a very [562] small fraction.

We shall also omit all terms containing such factors as $\cos 2\pi kr/\lambda$ or $\sin 2\pi kr/\lambda$ (k a moderate number). These reverse their signs by a very small change in r; they will therefore disappear from the resultant force, as soon as, instead of *single* particles P and Q, we come to consider systems of particles with dimensions many times greater than the wave-length.

From what has been said, we may deduce in the first place that, in applying the above formulae to the ion P, it is sufficient, to take for $\mathfrak{d}$ and $\mathfrak{H}$ the vectors that would exist if P were removed from the field. In each of these vectors two parts are to be distinguished. We shall denote by $\mathfrak{d}_1$, and $\mathfrak{H}_1$, the parts existing independently of Q and by $\mathfrak{d}_2$ and $\mathfrak{H}_2$ the parts due to the vibrations of this ion.

Let Q be taken as origin of coordinates, QP as axis of x, and let us begin with the terms in (2) having the coefficient a.

To these corresponds a force on P, whose first component is

$$4\pi V^2 e^2 a\left(\mathfrak{d}_x \frac{\partial \mathfrak{d}_x}{\partial x} + \mathfrak{d}_y \frac{\partial \mathfrak{d}_x}{\partial y} + \mathfrak{d}_z \frac{\partial \mathfrak{d}_x}{\partial z}\right) + e^2 a(\dot{\mathfrak{d}}_y \mathfrak{H}_z - \dot{\mathfrak{d}}_z \mathfrak{H}_y). \tag{5}$$

Since we have only to deal with the mean values for a full period, we may write for the last term

$$-e^2 a(\mathfrak{d}_y \dot{\mathfrak{H}}_z - \mathfrak{d}_z \dot{\mathfrak{H}}_y)$$

and if, in this expression, $\dot{\mathfrak{H}}_y$ and $\dot{\mathfrak{H}}_z$ be replaced by

$$4\pi V^2\left(\frac{\partial \mathfrak{d}_z}{\partial x} - \frac{\partial \mathfrak{d}_x}{\partial z}\right) \text{ and } 4\pi V^2\left(\frac{\partial \mathfrak{d}_x}{\partial y} - \frac{\partial \mathfrak{d}_y}{\partial x}\right)$$

becomes

$$2\pi V^2 e^2 a \frac{\partial(\mathfrak{d}^2)}{\partial x}, \tag{6}$$

where $\mathfrak{d}$ is the numerical value of the dielectric displacement.

Now, $\mathfrak{d}^2$ will consist of three parts, the first being $\mathfrak{d}_1^2$, the second $\mathfrak{d}_2^2$ and the third depending on the combination of $\mathfrak{d}_1$ and $\mathfrak{d}_2$.

Evidently, the value of (6), corresponding to the first part, will be 0.

As to the second part, it is to be remarked that the dielectric displacement, produced by Q, is a periodic function of the time. At distant points the amplitude takes
[563] the form c/r where c is independent | of r. The mean value of $\mathfrak{d}^2$ for a full period is $c^2/2r^2$ and by differentiating this with regard to x or to r, we should get r^3 in the denominator.

The terms in (6) which correspond to the part

$$2(\mathfrak{d}_{1x}\mathfrak{d}_{2x} + \mathfrak{d}_{1y}\mathfrak{d}_{2y} + \mathfrak{d}_{1z}\mathfrak{d}_{2z})$$

in $\mathfrak{d}^2$, may likewise be neglected. Indeed, if these terms contain no factors such as to $\cos 2\pi kr/\lambda$ or $\sin 2\pi kr/\lambda$ there must be between $\mathfrak{d}_1$ and $\mathfrak{d}_2$, either no phase-difference at all, or a difference which is independent of r. This condition can only be fulfilled, if a system of waves, proceeding in the direction of QP, is combined with the vibrations excited by Q, in so far as this ion is put in motion by that system itself. Then, the two vectors $\mathfrak{d}_1$ and $\mathfrak{d}_2$ will have a common direction perpendicular to QP, say that of the axis of y, and they will be of the form

$$\mathfrak{d}_{1y} = q\cos n\left(t - \frac{x}{V} + \varepsilon_1\right)$$

$$\mathfrak{d}_{2y} = \frac{c}{r}\cos n\left(t - \frac{x}{V} + \varepsilon_2\right).$$

The mean value of $\mathfrak{d}_{1y}\mathfrak{d}_{2y}$ is

$$\frac{1}{2}\frac{qc}{r}\cos n(\varepsilon_1 - \varepsilon_2),$$

and its differential coefficient with regard to x has r^2 in the denominator. It ought therefore to be retained, were it not for the extremely small intensity of the. systems of waves which give rise to such a result. In fact, by the restriction imposed on them as to their direction, these waves form no more then a very minute part of the whole motion.

§4. So, it is only the terms in (2), with the coefficient b, with which we are concerned.The corresponding forces are

$$-4\pi V^2 e^2 b\left(\dot{\mathfrak{d}}_x \frac{\partial \mathfrak{d}_x}{\partial x} + \dot{\mathfrak{d}}_y \frac{\partial \mathfrak{d}_x}{\partial y} + \dot{\mathfrak{d}}_z \frac{\partial \mathfrak{d}_x}{\partial z}\right) \tag{7}$$

and

$$-e^2 b(\ddot{\mathfrak{d}}_y \mathfrak{H}_z - \ddot{\mathfrak{d}}_z \mathfrak{H}_y). \tag{8}$$

| If Q were removed, these forces together would be 0, as has already been remarked. On the other hand, the force (8), taken by itself, would then likewise be 0. Indeed, its value is [564]

$$n^2 e^2 b(\mathfrak{d}_y \mathfrak{H}_z - \mathfrak{d}_z \mathfrak{H}_y) \tag{9}$$

or, by Poynting's theorem $(n^2 e^2 b S_x / V^2)$, if S_x be the flow of energy in a direction parallel to the axis of x. Now, it is clear that, in the absence of Q, any plane must be traversed in the two directions by equal amounts of energy.

In this way we come to the conclusion that the force (7), in so far as it depends on the part $(\mathfrak{d}_1)$, is 0, and from this it follows that the total value of (7) will vanish, because the part arising from the combination of $(\mathfrak{d}_1)$ and $(\mathfrak{d}_2)$, as well as that which is solely due to the vibrations of Q, are 0. As to the first part, this may be shown by a reasoning similar to that used at the end of the preceding section. For the second part, the proof is as follows.

The vibrations excited by Q in any point A of the surrounding aether are represented by expressions of the form

$$\frac{1}{r}\vartheta \cos n\left(t - \frac{r}{V} + \varepsilon\right),$$

where ϑ depends on the direction of the line QA, and r denotes the length of this line. If, in differentiating such expressions, we wish to avoid in the denominator powers of r, higher than the first—and this is necessary, in order that (7) may remain free from powers higher than the second– $1/r$ and ϑ have to be treated as constants. Moreover, the factors ϑ are such, that the vibrations are perpendicular to the line QA. If, now, A coincides with P, and QA with the axis of x, in the expression for d_x we shall have $\vartheta = 0$, and since this factor is not to be differentiated, all terms in (7) will vanish.

Thus, the question reduces itself to (8) or (9). If, in this last expression, we take for d and H their real values, modified as they are by the motion of Q, we may again write for the force

$$\frac{n^2e^2b}{V^2}S_x;$$

this time, however wo have to understand by S_x the flow of energy as it is in the actual case. |

[565] Now, it is clear that, by our assumptions, the flow of energy must be symmetrical all around Q; hence, if an amount E of energy traverses, in the outward direction, a spherical surface described around Q as center with radius r, we shall have

$$S_x = \frac{E}{4\pi r^2}$$

and the force on P will be

$$K = \frac{n^2e^2bE}{4\pi V^2r^2}.$$

It will have the direction of QP prolonged.

In the space surrounding Q the state of the aether will be stationary; hence, two spherical surfaces enclosing this particle must be traversed by equal quantities of energy. The quantity E will be independent of r, and the force K inversely proportional to the square of the distance.

If the vibrations of Q were opposed by no other resistance but that which results from radiation, the total amount of electromagnetic energy enclosed by a surface surrounding Q would remain constant; E and K would then both be 0. If, on the contrary, in addition to the just mentioned resistance, there were a resistance of a different kind, the vibrations of Q would be accompanied by a continual loss of electromagnetic energy; less energy would leave the space within one of the spherical surfaces than would enter that space. E would be negative, and, since b is positive, there would be attraction. It would be independent of the signs of the charges of P and Q.

The circumstance however, that this attraction could only exist, if in some way or other electromagnetic energy were continually disappearing, is so serious a difficulty, that what has been said cannot be considered as furnishing an explanation of gravitation. Nor is this the only objection that can be raised. If the mechanism of gravitation consisted in vibrations which cross the aether with the velocity of light, the attraction ought to be modified by the motion of the celestial bodies to a much larger extent than astronomical observations make it possible to admit.

§5. Though the states of the aether, the existence and the laws of which have been deduced from electromagnetic phenomena, are found insufficient to account for universal attraction, yet one may try to establish a theory which is not wholly different from that of | electricity, but has some features in common with it. In order to obtain a [566] theory of this kind, I shall start from an idea that has been suggested long ago by Mossotti and has been afterwards accepted by Wilhelm Weber and Zöllner.

According to these physicists, every particle of ponderable matter consists of two oppositely electrified particles. Thus, between two particles of matter, there will be four electric forces, two attractions between the charges of different, and two repulsions between those of equal signs. Mossotti supposes the attractions to be somewhat greater than the repulsions, the difference between the two being precisely what we call gravitation. It is easily seen that such a difference might exist in cases where an action of a specific electric nature is not exerted.

Now, if the form of this theory is to be brought into harmony with the present state of electrical science, we must regard the four forces of Mossotti as the effect of certain states in the aether which are called forth by the positive and negative ions.

A positive ion, as well as a negative one, is the center of a dielectric displacement, and, in treating of electrical phenomena, these two displacements are considered as being of the same nature, so that, if in opposite directions and of equal magnitude, they wholly destroy each other.

If gravitation is to be included in the theory, this view must be modified. Indeed, if the actions exerted by positive and negative ions depended on vector-quantities of the same kind, in such a way that all phenomena in the neighborhood of a pair of ions with opposite charges were determined by the resulting vector, then electric actions could only be absent, if this resulting vector were 0, but, if such were the case, no other actions could exist; a gravitation, i.e. a force in the absence of an electric field, would be impossible.

I shall therefore suppose that the two disturbances in the aether, produced by positive and negative ions, are of a somewhat different nature, so that, even if they are represented in a diagram by equal and opposite vectors, the state of the aether is not the natural one. This corresponds in a sense to Mossotti's idea that positive and negative charges differ from each other to a larger extent, than may be expressed by the signs $+$ and $-$.

After having attributed to each of the two states an independent and separate, existence, we may assume that, though both able to act on positive and negative ions, the one has more power over the positive particles and the other over the negative ones. This difference | will lead us to the same result that Mossotti attained by means [567] of the supposed inequality of the attractive and the repulsive forces.

§6. I shall suppose that each of the two disturbances of the aether is propagated with the velocity of light, and, taken by itself, obeys the ordinary laws of the electromagnetic field. These laws are expressed in the simplest form if, besides the dielectric displacement $\mathfrak{d}$, we consider the magnetic force $\mathfrak{H}$, both together determining, as we shall now say, *one* state of the aether or one field. In accordance with this, I shall introduce two pairs of vectors, the one $\mathfrak{d}$, $\mathfrak{H}$ belonging to the field that is produced by the positive ions, whereas the other pair $\mathfrak{d}'$, $\mathfrak{H}'$ serve to indicate the state of the aether which is called into existence by the negative ions. I shall write down two sets of equations, one for $\mathfrak{d}$, $\mathfrak{H}$, the other for $\mathfrak{d}'$, $\mathfrak{H}'$, and having the form which I have used in former papers[1] for the equations of the electromagnetic field, and which is founded on the assumption that the ions are perfectly permeable to the aether and that they can be displaced without dragging the aether along with them.

I shall immediately take this general case of moving particles.

Let us further suppose the charges to be distributed with finite volume-density, and let the units in which these are expressed be chosen in such a way that, in a body which exerts no electrical actions, the total amount of the positive charges has the same numerical value as that of the negative charges.

Let ρ be the density of the positive, and ρ' that of the negative charges, the first number being positive and the second negative.

Let $\mathfrak{v}$ (or $\mathfrak{v}'$) be the velocity of an ion.

Then the equations for the state $(\mathfrak{d}, \mathfrak{H})$ are[2]

$$\left.\begin{aligned} \mathrm{Div}\,\mathfrak{d} &= \rho \\ \mathrm{Div}\,\mathfrak{H} &= 0 \\ \mathrm{Curl}\,\mathfrak{H} &= 4\pi\rho\mathfrak{v} + 4\pi\dot{\mathfrak{d}} \\ 4\pi V^2\mathrm{Curl}\,\mathfrak{d} &= -\dot{\mathfrak{H}} \end{aligned}\right\} \qquad \text{(I)}$$

[568] | and those for the state (d',H')

$$\left.\begin{aligned} \mathrm{div}\,\mathfrak{d}' &= \rho' \\ \mathrm{div}\,\mathfrak{H}' &= 0 \\ \mathrm{Curl}\,\mathfrak{H}' &= 4\pi\rho'\mathfrak{v}' + 4\pi\dot{\mathfrak{d}}' \\ 4\pi V^2\mathrm{Curl}\,\mathfrak{d}' &= -\dot{\mathfrak{H}}' \end{aligned}\right\} \qquad \text{(II)}$$

1 Lorentz, La théorie électromagnétique de Maxwell at son application aux corps mouvants, *Arch. Néerl. XXV*, p. 363; Versuch einer Theorie der electrischen und optischen Erscheinungen in bewegten Körpern.

2 $\mathrm{Div}\,\mathfrak{d} = \frac{\partial\mathfrak{d}_x}{\partial x} + \frac{\partial\mathfrak{d}_y}{\partial y} + \frac{\partial\mathfrak{d}_z}{\partial z}$. $\mathrm{Curl}\,\mathfrak{d}$ is a vector, whose components are: $\frac{\partial\mathfrak{d}_z}{\partial y} - \frac{\partial\mathfrak{d}_y}{\partial z}$, etc.

In the ordinary theory of electromagnetism, the force acting on a particle, moving with velocity $\mathfrak{v}$, is

$$4\pi V^2\mathfrak{d} + [\mathfrak{v}.\mathfrak{H}],$$

per unit charge.[3]

In the modified theory: we shall suppose that a positively electrified particle with charge e experiences a force

$$k_1 = \alpha\{4\pi V^2\mathfrak{d} + [\mathfrak{v}.\mathfrak{H}]\}e \tag{10}$$

on account of the field $(\mathfrak{d}, \mathfrak{H})$, and a force

$$k_2 = \beta\{4\pi V^2\mathfrak{d}' + [\mathfrak{v}.\mathfrak{H}']\}e \tag{11}$$

on account of the field $(\mathfrak{d}', \mathfrak{H}')$, the positive coefficients α and β having slightly different values.

For the forces, exerted on a negatively charged particle I shall write,

$$k_3 = \beta\{4\pi V^2\mathfrak{d} + [\mathfrak{v}'.\mathfrak{H}]\}e' \tag{12}$$

and

$$k_4 = \alpha\{4\pi V^2\mathfrak{d}' + [\mathfrak{v}'.\mathfrak{H}']\}e', \tag{13}$$

expressing by these formulae that e is acted on by $(\mathfrak{d}, \mathfrak{H})$ in the same way as e' by $(\mathfrak{d}', \mathfrak{H}')$, and vice versa.

§7. Let us next consider the actions exerted by a *pair* of oppositely charged ions, placed close to each other, and remaining so during their motion, For convenience of mathematical treatment, we may even reason as if the two charges penetrated each other, so that, if they are equal, $\rho' = -\rho$. |

On the other hand $\mathfrak{v}' = \mathfrak{v}$, hence, by (I) and (11), [569]

$$\mathfrak{d}' = -\mathfrak{d} \qquad \text{and} \qquad \mathfrak{H}' = -\mathfrak{H}.$$

Now let us put in the field, produced by the pair of ions, a similar pair with charges e and $e' = -e$, and moving with the common velocity $\mathfrak{v}$. Then, by (10)–(13),

$$k_2 = -\frac{\beta}{\alpha}k_1, \qquad k_3 = -\frac{\beta}{\alpha}k_1, \qquad k_4 = k_1.$$

The total force on the positive particle will be

3 $[\mathfrak{v}\,.\,\mathfrak{H}]$ is the vector-product of $\mathfrak{v}$ and $\mathfrak{H}$.

$$k_1 + k_2 = k_1\left(1 - \frac{\beta}{\alpha}\right)$$

and that on the negative ion

$$k_3 + k_4 = k_1\left(1 - \frac{\beta}{\alpha}\right).$$

These forces being equal and having the same direction, there is no force tending to *separate* the two ions, as would be the case in an *electric field*. Nevertheless, the pair is acted on by a resultant force

$$2k_1\left(1 - \frac{\beta}{\alpha}\right).$$

If now β be somewhat larger than α the factor $2(1 - \beta/\alpha)$ will have a certain negative value $-\varepsilon$, and our result may be expressed as follows:

If we wish to determine the action between two ponderable bodies, we may first consider the forces existing between the positive ions in the one and the positive ions in the other. We then have to reverse the direction of these forces, and to multiply them by the factor ε. Of course, we are led in this way to Newton's law of gravitation.

The assumption that all ponderable matter is composed of positive and negative ions is no essential part of the above theory. We might have confined ourselves to the supposition that the state of the aether which is the cause of gravitation is propagated in a similar way as that which exists in the electromagnetic field. |

[570] Instead of introducing two pairs of vectors $(\mathfrak{d}, \mathfrak{H})$ and $(\mathfrak{d}', \mathfrak{H}')$, both of which come into play in the electromagnetic actions, as well as in the phenomenon of gravitation, we might have assumed one pair for the electromagnetic field and one for universal attraction.

For these latter vectors, say $\mathfrak{d}$, $\mathfrak{H}$, we should then have established the equations (I), ρ being the density of ponderable matter, and for the force acting on unit mass, we should have put

$$-\eta\{4\pi V^2\mathfrak{d} + [\mathfrak{v}.\mathfrak{H}]\},$$

where η is a certain positive coefficient,

§8. Every theory of gravitation has to deal with the problem of the influence, exerted on this force by the motion of the heavenly bodies. The solution is easily deduced from our equations; it takes the same form as the corresponding solution for the electromagnetic actions between charged particles.[4]

I shall only treat the case of a body A, revolving around a central body M, this latter having a given constant velocity p. Let r be the line MA, taken in the direc-

4 See the second of the obeys cited papers.

tion from M towards A, x,y,z the relative coordinates of A with respect to M, to the velocity of A's motion relatively to M, ϑ the angle between w to and p, finally p_r, the component of p in the direction of r.

Then, besides the attraction

$$\frac{k}{r^2}, \tag{14}$$

which would exist if the bodies were both at rest, A will be subject to the following actions.

1st. A force

$$k \cdot \frac{p^2}{2V^2} \cdot \frac{1}{r^2} \tag{15}$$

in the direction of r.

2nd. A force whose components are

$$-\frac{k}{2V^2}\frac{\partial}{\partial x}\left(\frac{p_r^2}{r}\right), -\frac{k}{2V^2}\frac{\partial}{\partial y}\left(\frac{p_r^2}{r}\right), -\frac{k}{2V^2}\frac{\partial}{\partial z}\left(\frac{p_r^2}{r}\right). \tag{16}$$

| 3rd. A force [571]

$$-\frac{k}{V^2}p \cdot \frac{1}{r^2}\frac{dr}{dt} \tag{17}$$

parallel to the velocity p.

4th. A force

$$\frac{k}{V^2}\frac{1}{r^2}pw\cos\vartheta \tag{18}$$

in the direction of r.

Of these, (15) and (16) depend only on the common velocity p, (17) and (18) on the contrary, on p and w to conjointly.

It is further to be remarked that the additional forces (15)-(18) are all of the second order with respect to the small quantities $\frac{p}{V}$ and $\frac{w}{V}$.

In so far, the law expressed by the above formulae presents a certain analogy with the laws proposed by Weber, Riemann and Clausius for the electromagnetic actions, and applied by some astronomers to the motions of the planets. Like the formulae of Clausius, our equations contain the absolute velocities, i. e. the velocities, relatively to the aether.

There is no doubt but that, in the present state of science, if we wish to try for gravitation a similar law as for electromagnetic forces, the law contained in (15)-(18) is to be preferred to the three other just mentioned laws.

§9. The forces (15)-(18) will give rise to small inequalities in the elements of a planetary orbit; in computing these, we have to take for p the velocity of the Sun's motion through space. I have calculated the *secular* variations, using the formulae communicated by Tisserand in his *Mécanique céleste*.

Let a be the mean distance to the Sun,
e the eccentricity,
φ the inclination to the ecliptic,
θ the longitude of the ascending node,
$\tilde{\omega}$ the longitude of perihelion,
[572] κ' the mean anomaly at time, $t = 0$, in this sense that, if n be the mean motion, as determined by a, the mean anomaly at time t is given by

$$\kappa' + \int_0^t n dt.$$

Further, let λ, μ and ν, be the direction-cosines of the velocity p with respect to: 1st. the radius vector of the perihelion, 2nd. a direction which is got by giving to that radius vector a rotation of $90°$, in the direction of the planet's revolution, 3rd. the normal to the plane of the orbit, drawn towards the side whence the planet is seen to revolve in the same direction as the hands of a watch.

Put $\omega = \tilde{\omega} - \theta$, $p/V = \delta$ and $na/V = \delta'$ (na is the velocity in a circular orbit of radius a).

Then I find for the variations *during one revolution*

$$\left.\begin{aligned}
\Delta a &= 0 \\
\Delta e &= 2\pi\sqrt{(1-e^2)}\left\{\lambda\mu\delta^2\frac{(2-e^2)-2\sqrt{(1-e^2)}}{e^3} - \lambda\delta\delta'\frac{1-\sqrt{(1-e^2)}}{e^2}\right\} \\
\Delta\varphi &= \frac{2\pi}{\sqrt{(1-e^2)}}\nu\left\{[-\lambda\delta^2\cos\omega + \delta(e\delta'-\mu\delta)\sin\omega]\frac{1-\sqrt{(1-e^2)}}{e^2} + \mu\delta^2\sin\omega\right\} \\
\Delta\theta &= -\frac{2\pi}{\sqrt{(1-e^2)}\sin\varphi}\nu\left\{[\lambda\delta^2\sin\omega + \delta(e\delta'-\mu\delta)\cos\omega]\frac{1-\sqrt{(1-e^2)}}{e^2} + \mu\delta^2\cos\omega\right\}
\end{aligned}\right\} \quad (19)$$

$$\Delta\tilde{\omega} = \pi(\mu^2 - \lambda^2)\delta^2\frac{(2-e^2) - 2\sqrt{(1-e^2)}}{e^4} + 2\pi\mu\delta\delta'\frac{\sqrt{(1-e^2)} - 1}{e^3}$$

$$-\frac{2\pi tg\frac{1}{2}\varphi}{\sqrt{(1-e^2)}}\nu\left\{[\lambda\delta^2\sin\omega + \delta(e\delta' - \mu\delta)\cos\omega]\frac{1-\sqrt{(1-e^2)}}{e^2} + \mu\delta^2\cos\omega\right\}$$

$$\Delta\kappa' = \pi(\lambda^2 - \mu^2)\delta^2\frac{(2+e^2)\sqrt{(1-e^2)} - 2}{e^4} - 2\pi\delta^2 - 2\pi\mu^2\delta^2$$

$$-2\pi\mu\delta\delta'\frac{(1-e^2) - \sqrt{(1-e^2)}}{e^3}$$

§10. I have worked out the case of the planet Mercury, taking $276°$ and $+34°$ for the right ascension and declination of the apex of the Sun's motion. I have got the following results: |

$$\Delta a = 0$$ [573]

$$\Delta e = 0.018\delta^2 + 1.38\delta\delta'$$

$$\Delta\varphi = 0.95\delta^2 + 0.28\delta\delta'$$

$$\Delta\theta = 7.60\delta^2 - 4.26\delta\delta'$$

$$\Delta\tilde{\omega} = -0.09\delta^2 + 1.95\delta\delta'$$

$$\Delta\kappa' = -6.82\delta^2 - 1.93\delta\delta'$$

Now, $\delta' = 1.6 \times 10^{-4}$ and, if we put $\delta = 5.3 \times 10^{-5}$, we get

$$\Delta e = 117 \times 10^{-10}$$

$$\Delta\varphi = 51 \times 10^{-10}$$

$$\Delta\theta = -137 \times 10^{-10}$$

$$\Delta\tilde{\omega} = 162 \times 10^{-10}$$

$$\Delta\kappa' = -355 \times 10^{-10}.$$

The changes that take place in a century are found from these numbers, if we multiply them by 415, and, if the variations of φ, θ, $\tilde{\omega}$ and κ' are to be expressed in seconds, we have to introduce the factor 2.06×10^5. The result is, that the changes in φ, θ, $\tilde{\omega}$ and κ' amount to a few seconds, and that in e to 0.000005.

Hence we conclude that our modification of Newton's law cannot account for the observed inequality in the longitude of the perihelion—as Weber's law can to some extent do—but that, if we do not pretend to explain this inequality by an alteration of the law of attraction, there is nothing against the proposed formulae. Of course it will be necessary to apply them to other heavenly bodies, though it seems scarcely probable that there will be found any case in which the additional terms have an appreciable influence.

The special form of these terms may perhaps be modified. Yet, what has been said is sufficient to show that gravitation may be attributed to actions which are propagated with no greater velocity than that of light.

As is well known, Laplace has been the first to discuss this question of the velocity of propagation of universal attraction, and later astronomers have often treated the same problem. Let a body B be attracted by a body A, moving with the. velocity p. Then, if the action is propagated with a finite velocity V, the influence which reaches B at time t, will have been emitted by A at an anterior moment, say $t-\tau$. Let A_1 be the position of the acting body at this moment, A_2 that at time t. It is an easy mat-
[574] ter to calculate the distance between those positions. Now, if the action at time | t is calculated, as if A had continued to occupy the position A_1, one is led to an influence on the astronomical motions of the order p/V if V were equal to the velocity of light, this influence would be much greater than observations permit us to suppose. If, on the contrary, the terms with p/V are to have admissible values, V ought to be many millions of times as great as the velocity of light.

From the considerations in this paper, it appears that this conclusion can be avoided. Changes of state in the aether, satisfying equations of the form (I), are propagated with the velocity V; yet, no quantities of the first order p/V or w/V (§8), but only terms containing p^2/V^2 and pw/V^2 appear in the results. This is brought about by the peculiar way—determined by the equations—in which moving matter changes the state of the aether; in the above mentioned case the condition of the aether will *not* be what it would have been, if the acting body were at rest in the position A_1.

BENEDICT AND IMMANUEL FRIEDLAENDER

ABSOLUTE OR RELATIVE MOTION?

Originally published in German as a pamphlet "Absolute oder relative Bewegung?" by Verlag von Leonhard Simion, Berlin 1896. 1: Immanuel Friedlaender; Berlin, Spring 1896. 2: Dr. Benedict Friedlaender; Berlin, January 1896.

1. The Question of the Reality of Absolute Motion and a Means for its *Experimental* Resolution.
2. On the Problem of Motion and the Invertibility of Centrifugal Phenomena on the *Basis of Relative Inertia*.

PREFACE

The question treated in the following is old; in addition to the cited writings there exists a rather extensive literature on it, about which we reserve comment to a later time. Our work was conceived without knowledge of that literature. It once more illuminates the uncertainty of the basis of our mechanics and of all exact science in general—wherein it must necessarily deal with long-known material—and points the way toward an attempt to solve the question, without any claim of already presenting the solution. If experiments now in progress lead to a result, they may silence the frequently raised objection that the limitations on our experience do not permit a decision. But if the experiments do not succeed, we may perhaps lead others to successful work in this direction by pointing out that the invertibility [*Umkehrbarkeit*] of the centrifugal force, considered already by Newton, urgently needs an experimental treatment as well as a theoretical investigation by reduction to a law of relative inertia. | [3]

1. THE QUESTION OF THE REALITY OF ABSOLUTE MOTION [5]

A body whose position in space changes is said to be in motion. This is one of two possible but fundamentally different definitions of motion. It presupposes that the process of change in position of a *single* body in space, quite independent of actual or possible relations to other bodies, has *content*, that is, that it differs from the state of rest by *recognizable effects*. When we apply this concept to any mechanical problems we have to refer observed motions to a coordinate system fixed in space in order to represent them properly; to do this we must know the true motion of our Earth—from

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 3
Gravitation in the Twilight of Classical Physics: Between Mechanics, Field Theory, and Astronomy.

which all measurements originate—so that we can transform from its co-moving coordinates, which we have fixed in the Earth, to a system of coordinates at absolute rest in the universe.

This definition with its consequences forms the basis of our present-day mechanics, and the well-founded high regard for the achievements of this science often prevents critics from daring to examine the justification of this definition. Nevertheless many have already stated the suspicion that this definition of motion is wrong,
[6] because only the relative motions of two or several bodies have *reality*, and | because the motion of one body, apart from its relations to other bodies, does *not* differ from a state of rest. This notion corresponds to our laws of thought and powers of imagination, but, as we shall see, not to the world of phenomena as understood according to the principles of mechanics.

Two groups of phenomena are relevant:

1. The appearance of the centrifugal force,
2. The stability of the free axes and of the plane of a Foucault pendulum.

The question of whether an absolute motion possesses reality, or whether there can be only relative motion, can be answered in different ways; to be precise there are *three* views that we want to consider here.[1] We will totally neglect those who deny the reality of absolute motion at the beginning of textbooks, give only the definition of relative motion, and then as they go along introduce a coordinate system fixed in space without allowing it to be noticed (without noticing it themselves?), particularly for the phenomena of rotation. Others, Kirchhoff for example, introduce *ab initio* coordinates at absolute rest, and in that way at least avoid self-contradiction. Only rarely is the question posed and discussed before it is answered.

Newton was the first to encounter the difficulty; he opines that it should be solved by experiment, and on the basis of an admittedly totally inadequate experiment[2] he
[7] arrives at | the view that absolute motion is real (see Mach, *Geschichte der Mechanik*, p. 317).

Kant grasped the question in its full significance with perfect clarity. He arrives at a solution by construing a difference between purely mathematical, "phoronomical" motion and physical, "phenomenological" motion; about the first he states the following (*Metaphysische Anfangsgründe der Naturwissenschaft*, 1786, Phoronomie, Grundsatz 1):[1]

> Every motion, as object of a possible experience, can be viewed arbitrarily as motion of the body in a space at rest, or else as rest of the body, and, instead, as motion of the space in the opposite direction with the same speed.

By way of contrast, propositions 1 and 2 of the Phenomenology state:

---

1 In the following only the three main points of view will be briefly sketched, as seems necessary for understanding the experiment, without entering into the theoretical treatment of the question by numerous authors.

2 Cf. below, [p. 20 in the original].

> 1. The rectilinear motion of a matter with respect to an empirical space, as distinct from the opposite motion of the space, is a merely *possible* predicate. The same when thought in no relation at all to a matter external to it, that is *as absolute motion*, is *impossible*. [...]
>
> 2. The circular motion of a matter, as distinct from the opposite motion of the space, is an *actual* predicate of this matter; by contrast, the opposite motion of a relative space, assumed instead of the motion of the body, is no actual motion of the latter, but, if taken to be such, is mere semblance.

What Kant states here can be found in Budde's general mechanics[2] in a different form, which more nearly corresponds to the modern point of view and will therefore be more easily understood. He does admit (p. 6) that the determination of position is relative in nature, but after making the transition from *phoronomics to kinetics* he observes—presupposing the principle of inertia—that we are forced by the facts to assume an absolute coordinate system,

> a | fundamental system within which the center of the Sun lies, with an angular velocity [8] $-w$ relative to the Earth. The principle of inertia holds in this system F, and in every other system that is at rest or in uniform translation relative to it; but this principle is no longer valid in a different system that rotates with respect to F, as shown by our experience on Earth.

We can properly state this claim of Kant and Budde as follows: in the purely geometric phenomenon of motion, or in mathematical space, there are neither fixed directions nor a fixed system of coordinates, all motion here is only relative; but in the dynamical phenomena of motion, or in physical space, there is such a coordinate system, which however can be translated parallel to itself without resulting in any real change. In the following I will call this a *fixed system of directions*, and for purposes of visualization I wish to compare physical space with the interior of a crystal imagined to be infinite in extent, in which to a certain extent each rotation, but no translational motion, would have a physical meaning. Indeed, Budde concludes that space is filled with a *medium*, and that the principle of inertia is a property of matter relative to this medium.[3]

Finally, the third possible conception is taken by Mach (p. 322); namely, that only relative motion is real, that our present-day mechanics is incomplete, and that *"the mechanical principles can probably be put in such a form that centrifugal forces result also for relative rotation."*

In short, the three conceptions are:[4] |

1. There is absolute motion, translation as well as rotation; physical space, as [9] opposed to purely geometrical space, possesses a *fixed coordinate system*.

---

3 Earlier authors already expressed similar views.

4 The first point of view is taken by Newton, Euler, Laplace, Lagrange, Poinsot, Poisson, Narr, the second, however partially in a quite different formulation, by Maxwell, Thomson-Tait, Streintz, Lange and others; the third especially by Mach. Some of these authors vacillate in their view or try, like Lange, to unite views 2 and 3.

2. There is no absolute translation, but there is absolute rotation. Real space, as opposed to that corresponding to our conceptual ability [*Denkvermögen*], has an *absolute system of directions*.

3. There is nothing but relative motion; *physical* space does *not differ* from *mathematical* space; but our present-day mechanics explains the phenomena of rotation incorrectly, or at least incompletely.

In the following we wish to treat the special case of centrifugal force first, in order to test the validity of the three views

Imagine a rigid system, consisting of a weightless rigid rod of length $2r$, which has two equal spheres, each of mass m, attached to its ends.

If we allow this system to rotate about an axis perpendicular to the rod at its midpoint—motions being referred to the Earth or to the fixed stars, because in the following we want to assume that the motion of the Earth relative to the fixed stars can be neglected as rather minor compared to that of our system—then there occurs a tension in the rod that equals mv^2/r, according to the well-known formula, where v is the velocity of the mass m. Now we treat this system, which we take to be located on the Earth, based on the three possible assumptions concerning motion. If we first pro-
[10] ceed on the assumption that there is only relative motion, and that | our coordinate system can be fixed completely arbitrarily, then whenever we take the rod itself as a coordinate, for example, or choose any coordinate system rigidly connected to our system of bodies, there is no accounting for the occurrence of the centrifugal force, that is, the measurable stress in the rod. For the rotational motion would then be represented as a rotation of the Earth and the whole firmament of fixed stars about our rod with the two spheres.

However, the rotation of these external bodies cannot explain the existence of the stress in the rod according to any law of mechanics known to date.

Secondly, let us assume that space is constituted somewhat like an infinitely large crystal, that there is no fixed system of axes in it, but that it possesses a system of directions. In this case only the angular velocity and the direction of the rotational axis would be uniquely determined, and we would therefore be entitled to assume that our system is rotating about a fixed axis, perpendicular to the rod and at a distance $a \leq r$ from its midpoint. We then obtain on one side of the axis the stress

$$F_1 = \frac{mv^2(r-a)^2}{(r-a)r^2},$$

and on the other side,

$$F_2 = \frac{mv^2(r+a)^2}{(r+a)r^2},$$

where v denotes the *same* speed as in our earlier equation. Since the rotational axis just assumed—which is only imaginary—is actually free, the rod will take on a uniform stress along its entire extent, namely

$$F = \frac{F_1 m + F_2 m}{2m},$$

but this stress equals

$$\frac{F_1 + F_2}{2} = \frac{mv^2}{2r^2}(r - a + r + a) = \frac{mv^2}{r}.^5$$

| What we have shown in this special example is valid in general according to Budde's treatment: A system of directions in space, which suffices only to define the angular velocity of a system without yielding information on its translational motion, is enough to determine uniquely the centrifugal forces. [11]

In our case just mentioned the motion would consist not only in rotation of one sphere about the other, which is simultaneously rotating about its axis with the same angular velocity and in the same sense, but at the same time the Earth together with the system of the world considered as a whole would move in a circle of radius r about our axis of rotation. This rotation would take place with the same angular velocity as that of our system and in the same sense, but such that the Earth and system of the world would always be only parallely displaced, and every line placed through the Earth and the firmament of fixed stars would maintain a constant direction.

Let us finally go on to the third point of view, the one which admittedly appeals least to us. Let us assume that there is absolute motion—even translation—that is, a space with a fixed coordinate system. Then an absolute rotation can be determined by using systems of masses such as our rod with the two spheres. Thereby we are enabled to refer our motion to a coordinate system of which we can claim that relative to space—relative to an absolute coordinate system—it does not rotate, but it may execute an arbitrary translational motion, for whose determination we have no basis. The whole edifice of our mechanics is based on this view, and by taking this view no contradiction with experience has yet turned up, as far as I know. But until one is in a position to exhibit an absolute translation, the known facts | agree equally [12]
well with a space in which only a system of directions possesses reality. Experiments in this direction have been performed, but we do not want to go into that here.

So the phenomena of centrifugal force teach us—on the basis of the usual mechanical views, namely the formula mv^2/r —that there is an absolute space, or at least a system of directions, and that the motion of a rigid system in this space in certain cases *really* differs from rest in a perceptible way, that is, by the existence of the measurable centrifugal force. But this assumption means that there is a real and in certain circumstances mechanically effective space that significantly differs from the space of mathematics. The only attempt known to us to make this idea moderately

5 The tension in a thread at whose two ends the mass m_1 is being pulled under the influence of an acceleration g_1 by a force F_1, and m_2 with the acceleration g_2 and the force F_2, is given by

$$\frac{m_1 m_2 g_1 + m_1 m_2 g_2}{m_1 + m_2} = \frac{F_1 m_2 + F_2 m_1}{m_1 + m_2}.$$

understandable is Budde's above-mentioned attempt[6] to regard space as occupied, and to regard inertia as a consequence of relative motion with respect to the occupying medium.

The second group of phenomena mentioned above, namely the fact that free axes and the plane of the Foucault pendulum (at the pole) remain fixed with respect to the sky of fixed stars, can be treated more briefly here, since they behave in exactly the same way. Suppose we take the view that there is only relative motion. Then we cannot explain the phenomena mentioned above if we treat, say, the Earth as fixed. Why the Sun, Moon and stars should drag along the free axes or the plane of the pendulum upon their daily revolution about the Earth could not be explained by mechanics. If
[13] we take the second view—that space | has a system of directions—then the phenomena mentioned above prove that the Earth possesses its own absolute rotation. They are explained and calculable under this assumption.

Of course, the same follows on the basis of a space in which not only absolute rotations, but also absolute translational motions are real.

We briefly summarize again what we have stated so far: our way of thinking is in accordance with a space in which one position as such does not differ from another position, nor one direction from another direction. That a body or a rigid system moves in this space, or that it rotates about an interior or exterior point or stays at rest, would then be only different ways of stating the same set of facts, depending on whether it pleases us to fix our system of coordinates by an external body, with respect to which the corresponding relative motion takes place, or by the system in question itself.

But from this point of view we cannot—as shown above—explain the two groups of phenomena of centrifugal force and preservation of axes on the basis of an arbitrary coordinate system.

We can then ascribe to space a system of directions, or assume that space is fixed in a sense and that we can recognize absolute rotation by the appearance of centrifugal phenomena—on the Earth, for example, by a decrease in g toward the equator, etc.; whereas, at least for the time being, we lack evidence for the recognition of any absolute translational motion of the Earth.

To repeat, our intuition opposes this result; the proper reason for this opposition is
[14] that we see ourselves forced to admit a factual difference between mathematical | space and actual space, or differently put, between the space that corresponds to our conceptual ability and the space of phenomena perceived by our senses.

We can try to explain this result the way Budde did. Similar ideas are found already in Kant, who also tried to explain that the content of *vis viva* [*lebendige Kraft*] in a moving mass equals $mv^2/2$ by filling space with a medium.[7] But if we are not content with this, we must contest the result by disputing the premises; that is,

6 This idea was already formulated by Kant in the 1747 treatise regarding the true assessment of *vis viva* [*lebendige Kraft*] (which is in other respects full of obscurities), and it was later expressed by many others, in different forms.[3]

we must question the basis of the usual mechanical explanation of the rotational phenomena under discussion (namely, the law of inertia and the formula $mv^2/2$ for the *vis viva*).

Without knowing that Mach had already done this, I have for many years had doubts about the completeness of these foundations of mechanics, and in particular I have become convinced that the phenomenon of centrifugal force, properly mechanically understood, should also be explicable solely in terms of the relative motions of the system concerned without taking refuge in absolute motion. But I was well aware that the mere statement of these doubts does not amount to much, and that one must find either a new formulation of the expression for *vis viva* of a moving mass and thereby an improved version of the law of inertia, | or one must prove the inadequacy [15]
of the prevailing view experimentally. The phenomena of the centrifugal force in particular seemed to me suited for such an experimental resolution of the question: if the centrifugal force that appears in a flywheel can be explained solely from its relative motion, then it must be possible to derive it also under the assumption that the flywheel is at rest, but that the Earth is turning with the same angular speed about the flywheel axis in the opposite sense. Now, just as the centrifugal force appeared in the resting wheel as a consequence of the rotation of the massive Earth together with the universe, so there should appear, I reasoned, in correspondingly smaller measure an action of centrifugal force in resting bodies in the vicinity of massive moving flywheels. If this phenomenon was verifiable, this would be the incentive for a reformulation of mechanics, and at the same time further insight would have been gained into the nature of gravity, since these phenomena must be due to the actions at a distance[8] of masses, and here in particular to the dependence of these actions at a distance on relative rotations.

In light of the smallness of the masses available for our experiment, I had little hope of an experimental solution, until I thought of a promising experimental arrangement in the fall of 1894. This arrangement consists of putting the most sensitive of all physical instruments, a torsion balance, in the extension of the axis of a heavy mass that rotates as rapidly as possible, namely a large flywheel, for example in a rolling-mill. If the beam of the balance, bearing two small spheres at its ends, is not parallel to the (vertical) plane of the flywheel when the latter is at rest but inclined by an angle of about $45°$ to it, then according to our | theory tensile forces that tend [16]
to separate the spheres from the extension of the axis must appear, so as to make the balance parallel. However, a sensitive balance is a delicate instrument, and a rolling-

7 See footnote 6. Even if I think of space as filled, because otherwise action at a distance would remain inexplicable, I nevertheless strongly doubt that the phenomenon of inertia (respectively, of *vis viva*) derives from motion relative to the nearly massless aether, but I am of the opinion that motion relative to the aether as influenced by nearby gravitating masses, or equivalently motion relative to the gravitating masses themselves causes the inertia (respectively, the *vis viva*) of a moving mass through mediation by the aether.

8 Whether these actions are, as the writers believe, transmitted through a medium or not makes no difference to the matter.

mill is probably not the most comfortable and optimal location for precision measurements, and so due to the many sources of error my experiments, which I started already in November of 1894 at the rolling-mill in Peine—with the kindest support on the part of the management and engineers—have brought to light no incontestable results that I would want to transmit to the public, even though a deviation in the expected sense was established at the beginning and the end of the motion. But since the balance that was used also reacted to other influences, in particular being deviated by a burning candle at a distance of 4 meters in a room (the closer sphere being "attracted"), and since at the rolling-mill at Peine certain furnace doors are opened and closed in the same time intervals as the starting and stopping of the machines we used, I do not yet consider my results unobjectionable. Experiments with a torsion balance inside a double-walled box of copper with a layer of water about 14 mm thick between the two walls, have shown that even this precaution does not suffice to make the needle completely independent of external heat sources. Whether the heat perturbations are to be explained simply by the circulation of air in the interior of the box—which is always sealed airtight to the exterior—or whether effects as with a radiometer are to be considered, I do not dare to decide for the time being. A new instrument that is being prepared will, we hope, exclude all previous sources of error. Although reliable results are not yet available, the constant occupation with the matter and the frequent discussions with my brother Dr. Benedict Friedlaender have led
[17] us to the conclusion that the matter is of sufficient importance | to publish our thoughts already. My brother called my attention to Mach right at the start and in joint work we have drawn many of the consequences that would result from our view. The results of these considerations, as well as several opinions that I cannot fully share, were put together by my brother in the second part of this treatise, and there he has also attempted to state the law of inertia differently, so that one can derive from it the relativity, hence also the invertibility, of centrifugal forces. But it seems to me that the correct formulation of the law of inertia will be found only when the relative inertia, as an effect of masses on each other, and gravity, which is after all also an effect of masses on each other, are reduced to a unified law.[9] The challenge to theoreticians and calculators to attempt this will only be truly successful when the invertibility of centrifugal force has been successfully demonstrated. |

[18] 2. ON THE PROBLEM OF MOTION AND THE INVERTIBILITY OF CENTRIFUGAL PHENOMENA

We are accustomed to regard mechanics as the most perfect of all natural sciences, and efforts are made to reduce all other sciences to mechanics or, so to speak, to resolve

9 For this it would be very desirable to resolve the question of whether Weber's law applies to gravity, as well as the question concerning gravity's speed of propagation. Regarding the latter question, one might use an instrument that makes it possible to measure statically the diurnal variations of the Earth's gravity as it depends upon the position of the celestial bodies.

them in mechanics. Mechanics is the most concrete and nevertheless also the mathematically (that is, for quantitative calculations) finest, clearest and most exactly developed science. We have no cause to examine to what extent this evaluation of mechanics is justified, and we recalled the above statements only in order to indicate that all considerations or matters of fact related to the foundations of mechanics may claim more than the usual interest and importance. Our subject at present is such a consideration, which truly concerns the foundations of mechanics and thereby those of our whole scientific worldview; however, this is no new subject, but surely a nearly forgotten and at any rate not always sufficiently respected problem, *the problem of motion*. This problem is probably connected with another one, which has been discussed far more frequently in former and more recent times, but which is apparently still a long way from an even moderately satisfactory solution, *the problem of gravitation*. |

Concerning gravity we have nothing more than the knowledge of superficially [19] perceptible facts, along with a purely *mathematical* theory, which is in no way physical. No mind thinking scientifically could ever have permanently and seriously believed in unmediated action at a distance; the apparent force at a distance can be nothing other than the result of the effects of forces that are transmitted in some way by the medium being situated between two gravitating bodies. But our presentations refer primarily only to the problem indicated first, that of motion, as we may briefly denote it, and not directly to the latter problem. For the science of motion with all its derivations and consequences contains an image, or rather a formulation that can *not* be imagined, which must be a hint of a present flaw to all who are convinced *that something that cannot be imagined also cannot be actual, i.e. acting*. *Vis viva*, for example, is defined to be proportional to the velocity squared $(mv^2/2)$; and the velocity is defined as the measure of distance divided by that of time (l/t). Centrifugal phenomena are explained by the conflict between a constrained, curvilinear motion and inertia, which tends to maintain rectilinear motion. In both cases, and more generally, one thinks of motion—or let us say, rather, one defines motion—as *absolute* motion, as motion of a mass from one "absolute position" to another "absolute position"; for *thinking* about absolute motion is just what we can *not* do, and precisely this motivated the following considerations. The root of these considerations, as far as the author is concerned, lies in some difficulties and obscurities of none less than Newton, which | I came to know in the early 90's from quotes of Mach in his [20] *Geschichte der Mechanik*.

In themselves the arguments are as simple as they are unusual, which can easily lead to obscurity if one does not demonstrate and think through the matter step by step from the simplest case.

What we perceive of motion is always only relative motion, changes in position of masses *with respect to other masses*. Our hand moves against the rest of the body, considered as relatively motionless; we move on the deck of a ship, the ship moves on the surface of the Earth, it changes its distance from the continent, thought of as fixed. Our planet moves in the universe, namely with respect to a coordinate system that is considered fixed somewhere (in the Sun, for example), and so forth.

Now think (if you can) of a progressive motion of a *single* body in a universal space that is otherwise imagined to be totally empty; how can the motion be detected, i.e. distinguished from rest? By *nothing* we should think; indeed, the whole idea of such an absolute, progressive motion is meaningless. And still, in one case, namely when in motion, our absolutely moving body, thought of as isolated, is supposed to possess an amount of energy that is proportional to (half) the square of its (meaningless) velocity!

That this cannot be imagined is no new discovery; it is so apparent, and in connection with our astronomical knowledge the reasoning suggests itself so naturally, that many should have encountered it. (Cf., e.g., Wundt, *System der Philosophie* etc.)

More tenacious than the absurd idea of an absolute, so-called *translational* motion is that of the rotation of a sphere, for example, about an axis taken in its abso-
[21] lute sense. | To make the picture more concrete and impressive, let us immediately consider one of those well-known apparatuses used in schools for illustrating centrifugal effects. An approximately spherical framework of elastic brass blades can be rapidly rotated about its axis, so that all the blades are bent in such a way that this originally spherical frame suffers a polar flattening and an equatorial thickening. Here too, according to the usual interpretation, one thinks of the rotational motion as absolute and explains the whole phenomenon in the well-known way. Connected with this, or at least with an entirely equivalent example, is the difficulty that Newton already felt strongly, and which seduced that researcher into statements that seem to us more than merely risky. Newton had suspended a glass of water on two strings so that after twisting the strings about each other they put the glass into rapid rotation as they unwound. Since the friction between the glass and the water is rather small, and the inertia of the mass of water in the glass is quite considerable, at the beginning the glass turns nearly alone, whereas the water remains behind; only slowly does the mass of water take part in the rotation. One observes, as is well known, that no centrifugal effects whatever occur as long as the glass rotates alone (or almost alone); the surface of the water remains flat. But as the water takes part in the rotation more and more, it increasingly rises at the rim and falls at the center. Newton concludes from this that the centrifugal effects are a consequence of absolute—but not of relative—rotation; for in the beginning the glass turns compared to the objects "at rest" in the room, including also the mass of water in the glass; but afterwards, according to N., the water rotates "absolutely," and this absolute rotation brings centrifugal effects
[22] into play. It is not difficult to establish the | untenability or even the incorrectness of this Newtonian idea, as Mach did. In the beginning the glass turns with respect to the objects at rest in the room or with respect to the Earth, or more correctly the universe; whereas later the mass of water rotates with respect to the universe. In the first case there was relative rotation between *mass of water* (+ universe) and *glass*, in the second between *mass of water* (+ glass) and the *universe*; and only the latter, but not the former, produced the centrifugal effects. Mach justifiably points out that "absolute rotation" is a fiction, or more accurately, an unthinkable absurdity. The naive mind will immediately object that it is just not possible to hold the mass of water fixed and

now “let the universe rotate” about the same rotation axis; but the more acute mind will quickly agree that *both ideas are plainly identical, namely indistinguishable in any logical or practical way.* The true fact of the matter is just this, that the rotation of the glass with respect to the mass of water releases no centrifugal forces, but that the rotation of the universe with respect to the glass (or equivalently the rotation of the mass of water with respect to the universe) does do so. In the first case it is the very small mass of the glass’s wall that rotates with respect to the mass of water, in the second case it is the universe; we should not be surprised that the vanishingly small mass of the glass does not call forth any noticeable centrifugal forces, but no one could know, as Mach aptly remarked, how the experiment would turn out if the thickness of the glass were significantly increased and its walls ultimately became several leagues thick. Of course, this experiment cannot be executed in this form. But there is still the question of whether an experimental arrangement is possible in practice that would amount to the same thing, and allow us to establish the *invertibility* or *relativity* of centrifugal effects. |

Now it was my brother Immanuel who devised and tried to test an experimental [23]
arrangement of this type. Technical difficulties, some of them quite unexpected, have so far prevented the realization of a reliable result.

It is well known that the torsion balance is the most sensitive of all instruments. The large flywheels in rolling mills and other large factories are probably the largest rotating masses with which we can *experiment*. The centrifugal forces express themselves in a push causing recession from the axis of rotation. Thus, if we place a torsion balance at not too great a distance from a large flywheel, so that the point of suspension of the part of the torsion balance that can turn (the “needle”) lies exactly or approximately on the extension of the flywheel’s axis, then if the needle was not originally parallel to the plane of the flywheel it should tend to approach that position and show a corresponding deflection. Namely, the centrifugal force acts on each element of mass that is not on the rotation axis in a direction tending to move it away from the axis. It is immediately apparent that the greatest possible separation is reached when the needle becomes parallel. So far the difficulties opposing the experiment were that a sensitive torsion balance is also put into motion by perturbing influences —particularly effects of *heat*— as if, by the way, the warmer parts of the apparatus would have an *attractive* effect.

Now if we assume that the experiment were to work flawlessly, we would have thereby discovered a new mechanical-physical phenomenon, whose consequences would be extraordinarily far-reaching. Certainly the phenomenon would be explicable and predictable, as shall be shown immediately, if one were to recast motion and
all concepts connected with it, including | *inertia* in particular, in such a way that *rel-* [24]
ative motion would replace the present tacitly assumed concept of *absolute* motion. However, the predictability on the basis of this suggested recasting would be no objection to the claim that we would be dealing with a fundamentally new phenomenon; for precisely this recasting has not been carried out and tested by anyone. The law of inertia in the usual manner of representation can be transparently described by

saying that every body opposes any change of its velocity (conceived as absolute) with a resistance proportional to its mass in the corresponding direction. Here, the remaining bodies of the universe are completely ignored; in fact, a point that must be especially emphasized is that the concept of mass is, except for its derivation from gravity (mg), derived precisely from the facts of inertia. Every change in velocity, i.e., every acceleration (in the simplest case, for example, the imparting of motion to a body previously at rest until it reaches a certain velocity) is held to be opposed by a *resistance*, the overcoming of which requires the quantity of energy that is afterwards, when the body is in motion, supposed to be contained in that motion as "kinetic energy." It is here to be noted once more that translational motion of a single body in space otherwise regarded as empty is an absurdity, namely, it does not differ from its opposite, rest. Thus, the creation of such a chimera should not require any energy; therefore, if in contrast the actual world does agree with our prerequisites of thought [*Denknotwendigkeiten*], it should surely make a difference with respect to which other bodies motion is to be created, in a word, what *relative* motion of previously nonmoving bodies is to be created. Accordingly, inertia is to be grasped rela-
[25] tively; one could formulate the law of relative inertia as follows: All | masses strive to maintain their *mutual* state of motion with respect to speed and direction; every change requires positive or negative energy expenditure, that is, work is either required—in the case of an increase in velocity—or is released—in the case of a decrease in velocity. The resistance to changes in velocity would then, as soon as we regard all motions as relative, be expressed not only in the one body that, as we are accustomed to say, we "set in motion" (that is, set in motion relative to the Earth) but also in all the others that we regard as being at rest in accordance with the usual conception. According to the usual conception, inertia occurs on a railway train that is to be brought from rest to a certain velocity, but not on the Earth; according to our view it occurs also on the Earth, even though this is not noticeable due to the extraordinarily much larger mass. But when we put very large masses into motion, to the extent permitted by our technology, and we can observe the behavior of very easily moved masses in the vicinity, it is possible that the relativity of inertia can be shown directly. Precisely that experiment of my brother, namely to find the needle in parallel orientation, which has been unsuccessful so far, would be, as is easily demonstrated, not only the proof of the invertibility of the centrifugal force, but also of the relativity of inertia. In our view, both are the same.

The application of the thought indicated here is very simple but unusual to a high degree. For if we consider the resistance to acceleration that some body exhibits, we do not have the slightest thought of other masses that are nearby! But if we do so and hold firmly to the guiding thought that the masses strive to maintain their *relative*
[26] velocity, it turns out that (for motion on a | straight line of body A relative to body B as the simplest case):

$$\left.\begin{array}{l}\text{accelerated approach and}\\ \text{decelerated withdrawal}\end{array}\right\}\text{must have a repulsive effect}$$

$$\left.\begin{array}{l}\text{accelerated withdrawal and}\\ \text{decelerated approach}\end{array}\right\}\text{must have an attractive effect.}$$

In the first two cases a recession, and conversely in the last two cases an approach of the second body B would satisfy the striving to maintain the relative velocity, that is, the inertia considered as relative. But since the second body B has to be treated as inert with respect to the Earth as well, the motion induced in it in this way by body A, having changing velocities, will not annul the relative velocity, but only reduce it; and no matter how large we may choose the mass of body A, it will always be extremely small compared to the Earth's mass, so the motion of body B with respect to the Earth can only be very small, and if it be detectable at all, then only by a sensitive apparatus. As further illustration one can say that due to its inertia with respect to A, body B strives to set itself in motion with respect to the Earth, but because of its inertia with respect to the Earth it strives to move relative to A rather than relative to the Earth.

Let us now apply these considerations to our flywheel and the torsion balance placed before it.

Let the circle $AFCBDF'A$ [in Fig. 1] represent the rim of the flywheel and P a readily movable body or mass point within the rim of the flywheel, as close as possible to its plane, namely a part of the mass of the arm of the torsion balance. For simplicity, let us assume that the point P actually lies within the plane of the flywheel, which of course cannot be realized with strict accuracy | for common wheels whose [27]
spokes and rim lie in one plane.

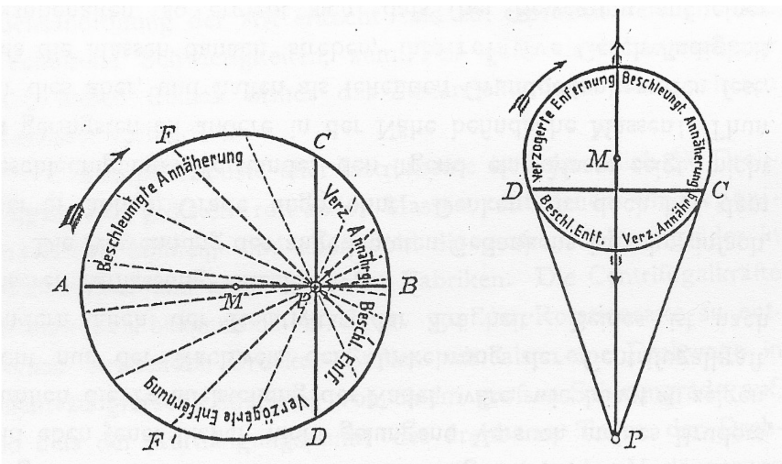


Fig. 1. Fig. 2.

[Beschleunigte Annäherung–accelerated approach; Verz. Annäherg–decelerated approach; Beschl. Entf.–accelerated withdrawal; Verzögerte Entfernung–decelerated withdrawal.]

Let us now join point P to the center of the flywheel M, rotating in the direction of the arrow, and extend this line until it meets the rim at A on the left, at B on the right and also erect the perpendicular on AB at P, cutting the rim above at C, below at D. Then it is clear that every mass point of the rim on its way from A over C to B approaches the point P and then on the way from B over D to A recedes from it. However, the approach on the semicircle AB is accelerated up to C and then decelerated to B; similarly, the withdrawal on the semicircle BA is accelerated to D and decelerated from D to A. In view of the simplicity of the situation we can dispense with an analytic proof. But since in accordance with what we have said accelerated approach and decelerated withdrawal act in the same sense, namely both repulsively, while decelerated approach and accelerated withdrawal both act attractively, we see
[28] that we can divide the rim into two parts that differ | in their effect, namely, the part left of CD, which repels, and that part right of CD, which attracts the point P. But it is also easily understood that only the force components acting along the line AB are effective. Namely, opposite to every point F on the rim there is another, F', whose force components along the line AB reinforce each other, whereas the perpendicular components cancel, being equal and opposite. All forces that push from the left of CD and all those that pull from the right of CD act together along the line AB in the direction from A to B. Therefore, on the basis of the conception of the relativity of inertia an acceleration *away from the axis* is imparted to the point P, as our conception of the invertibility of centrifugal force requires. The relative rotation between the wall of Newton's bucket and the water contained in it would indeed generate appreciable centrifugal forces in the water if the wall were sufficiently massive to be no longer practically non-existent compared with the mass of the Earth.[10]

If the ideas sketched here are correct, many consequences will follow, some of which will admittedly seem very strange. The same amount of gunpowder, acting on the same cannon ball in the same cannon, would impart to the projectile a greater velocity on, for example, the Moon than on the Earth. Naturally, however, the greater velocity would represent, rather than a greater amount, the same amount of energy as the smaller velocity that the projectile receives on our more massive planet. This
[29] would reveal itself in the fact that despite the greater velocity, | the penetration capacity on the Moon would not be greater than on the Earth. For the $mv^2/2$ as measure of the so-called kinetic energy would not be the complete formula, as it fails to take into account the surroundings with regards to mass and distance, that is, the specification of the masses for which the velocity "v" holds.

One should not be overly hasty in rejecting our ideas as obviously incorrect because of their actual or apparent consequences. For example, Foucault's pendulum should be explicable according to our ideas as well. According to the conception of relativity of all motion the fact that the plane of the pendulum is carried along not by the Earth, assumed to be rotating, but rather by the universe when it is assumed to be

10 We originally said "universe" but now the "Earth." It is to be assumed that the Earth will probably play a much larger role than the more distant masses of the universe.

rotating; the plane notoriously follows the latter, and *not* the Earth. To resolve the difficulty one only has to assume that the forces that turn the plane have their origin in the one-sided action of the masses of Sun and Moon, whereas the uniformly distributed mass of the Earth has no effect.

The phenomena of the tides would also have to have a treatment different from the usual one; namely, in our figure we merely need to take our point P outside the circle and draw tangents from it to C and D [Fig. 2]; the circle then represents the Earth, the point P the Moon or the Sun, and PC and PD the axial section of a cone tangent to the Earth from the Moon, treated as a point. One will then see that the Earth will be divided by the approximately circular plane CD, which appears in the figure as a line, into two parts that, on account of the distance, are very nearly equal; of these, the half below CD, i.e., the part turned toward the Moon, will be attracted, while the part above CD, away from the Moon, must be repelled; the mobile water follows these attractive and repulsive forces and excites both the tidal waves that circle the Earth in the time between two culminations | of the Moon. Incidentally, this [30]
explanation differs from the standard one essentially, rather than only in the point of view. Were it possible to connect the two objects rigidly, so that rotation but no approach or recession were possible, then a difference would result; according to the usual explanation the wave away from the Moon could *not* be realized; according to our explanation the occurrence of the tides would not be materially changed.

It is not our intention to dwell on the many other consequences, and likewise we forgo a more detailed interpretation of the analogies that suggest themselves. Still, let us mention the following parallels only in order to indicate the extent to which the problem of motion that we have raised and hypothetically solved here is related to that of the nature of gravity and at the same time comes rather close to the known effects of electric forces: a body that approaches a second one or moves away from it would be without influence on the latter as long as the velocity of approach (to be taken either with a positive or a negative sign) remains unchanged; any change of this velocity on the other hand would entail the above-demonstrated effect.

As is well known, the presence of a current in a conductor is not sufficient for the generation of induction effects, either the magnitude of the current or the distance must vary; in our case the *change* of distance, i.e. the motion, would not suffice for the generation of the attractive or repulsive effects, but rather the velocity itself has to change.

If we think of the effects in question as originating from some as yet unspecified waves, for example from longitudinal pressure waves (although most people would rather think first of transverse waves, due to the prevailing opinions!), then | the [31]
breaking of the waves with equal speed would have no effect, whereas acceleration of the rhythm would induce a repulsive force as long as it lasts, and deceleration an attractive force. It should be noted that these last considerations are only hints and not completed developments that could be understood by attentive reading only. They are also hypothetical to a high degree, as are our main statements, which of course can be regarded as facts only if the experiments described above (or equivalent experiments)

are successful. But let us finally emphasize that the success of the experiments would prove the presence of the type of action in question (even if not our interpretation), but its failure would not disprove this action. It remains questionable whether the effect is of an order of magnitude that would be reliably observable in the face of the experimental error sources. But if we deny the existence of the "inverse centrifugal force," for short, there would be consequences that would really be totally untenable. The incomprehensible would be deemed a fact, the logical absurdity of absolute motion would have to be regarded as having an effect, hence also as actual or real. To imagine and grasp this again in a concrete way, let the reader imagine being on a seat that is fixed on the axis of a rotational apparatus, freely rotating in otherwise empty space, in such a way that he must take part in the "rotation" of the apparatus. The "rotation" of the apparatus would then be accompanied by no change in position with respect to other heavenly bodies, it would not only be imperceptible *as such*, it would be totally unthinkable. It would be a logical monster. Nevertheless, according to the prevailing view that unthinkable "rotation," which cannot be differentiated from rest, is supposed to generate centrifugal forces so that the reader sitting on that seat can
[32] observe the phantom "rotation" by the | equatorial bulge of his little speck of matter, and can even *measure* the motion, not to say the ghost of the motion, by the amount of the deformation.

The world as a whole, we dare say, is not made in a way that would be in conflict with our prerequisites of thought. And therefore the idea of the relativity of all motion and also the origin of the centrifugal effects in *relative* acceleration resistance may have a priori probability on its side, and not against it, in spite of all its unaccustomed and seemingly alarming consequences.

On the basis of our conception it is naturally also necessary to modify the interpretation of the astronomical facts. The Ptolomaic and the Copernican system are both equally "*correct*" as far as they both describe the actual motions of the celestial bodies truthfully; but this description takes a much *simpler* form if one puts the coordinate system at the larger so-called central body, rather than at the Earth. In accordance with the conception of the relativity of all motions, including therefore central motions, a revolution of the Earth can be completely replaced by an axial rotation of the Sun *insofar as only these two bodies come into consideration*. The circumstance that the Earth, despite the "attraction," does not plunge into the Sun, or the Moon into the Earth, is of course explained on the basis of the usual conception by the motion of revolution of the smaller celestial body; while, for example, the axial rotation of the Sun with respect to the universe is taken to be negligible, and plays no role at all. If our conception is correct, the so-called axial rotations are not irrelevant to the equilibrium of the world systems but must be equally taken into account like all other factors. Incidentally, the assumption of an attraction of the Earth by the Sun is not a felicitous interpretation of the factual situation insofar as the so-called *attractive* |
[33] forces can only be adduced from the reduction of distance; naturally, this is not to say that the Sun would not attract the Earth if the relative motions of the two bodies were other than they actually are. However, as the facts stand, that true attraction does *not*

obtain; in accordance with everything we know, it would indeed occur in the case of relative rest of the bodies and bring about the fall of the Earth into the Sun. The attraction is compensated by the existing relative motions, and this would correspond to the usual conception if it would take into account the relative motions instead of operating with the phantom of absolute rotation and inertia treated correspondingly as absolute.

It is also apparent that according to our conception the motions of the bodies of the solar system could be seen as pure *inertial motions*, whereas according to the usual view the inertial motion, or rather its permanent gravitationally modified tendency, would strive to produce a rectilinear-tangential motion.

The central point upon which our view differs from the conventional one can also be expressed precisely as follows. The prevailing view refers all locations, hence all motions and derived concepts such as accelerations and inertia in particular, to a coordinate system considered *absolutely fixed in space*; the absolutely fixed point in space would accordingly be not only an idea, but would have a most real meaning; it would be actual because it could act. It would be such although no criterion can be given for its being fixed. A sympathetic devotee of Kant objected in a private communication that my law of inertia is | not well defined, whereas the usual is definite and moreover [34]
is merely an application of the law of causality to mechanics, with its (allegedly) *a priori* character. "No body changes its motion without cause." By no means do we acknowledge the *a priori* character of the law of causality; if so, the work of Galileo, to the extent that he discovered inertia, would then have been labor in vain and would have resulted only in trivialities, so to speak, which could have been obtained far more simply by deductively applying theorems that were certain *a priori*. To repeat, this we deem incorrect; yet our law of inertia may be given quite an analogous formulation, such as the statement that "no bodies change their *relative* motions without cause"; wherein the "old" law, you see, is only completed by the emphasized word "*relative*." That Kantian objection is surprising, since it, in particular, further supports the objectively real meaning of spatial relations. This is also the basis for the interest that our treatment may perhaps claim, even in the case that it would for some reasons turn out to be untenable. Namely, in that case we would have shown that it is *possible* to proceed from the relativity of all motions, that one *can* explain inertia and centrifugal motions on the basis of the relativity hypothesis; but that upon following this chain of thought further one hits upon factual contradictions, which make the assumption of absolute motion necessary and therefore make manifest the real significance of a coordinate system taken to be absolutely fixed in space, and thereby with even greater probability make manifest the reality of the spatial relations.

From private objections I gather incidentally that if *incorrectly*, that is incompletely, applied, our hypothesis seems to include a violation of the principle of conservation of energy; when considered more exactly, however,| namely when our point [35]
of view is completely implemented, this apparent contradiction disappears. It is true, as we already emphasized, that the formulas for the kinetic energy and everything

depending on it are in need of an appropriate completion by respecting the other masses in the vicinity.

EDITORIAL NOTES

[1] Translation from: Henry Allison, Peter Heath (eds.): *Theoretical Philosophy after 1781*. (*The Cambridge Edition of the Works of Immanuel Kant*.) Cambridge University Press: Cambridge, 2002, 200, 261, 262.

[2] This reference is to Budde, E. *Allgemeine Mechanik der Punkte und starren Systeme: Ein Lehrbuch für Hochschulen*. First edition. Reimer: Berlin, 1890.

[3] This reference is to Kant, Immanuel: *Gedanken von der wahren Schätzung der lebendigen Kräfte und der Beurtheilung der Beweise derer sich Herr von Leibnitz und andere Mechaniker in dieser Streitsache bedienet haben, nebst einigen vorhergehenden Betrachtungen welche die Kraft der Körper überhaupt betreffen*. (1746). Published in: *Kants gesammelte Schriften*, vol. 1, Berlin: Königlich Preußische Akademie der Wissenschaften, 1910, 1–182.

AUGUST FÖPPL

ON ABSOLUTE AND RELATIVE MOTION

Originally published as "Über absolute und relative Bewegung" in Sitzungsberichte der Bayerischen Akademie der Wissenschaften, mathematisch-physikalische Klasse (1904) 34: 383–395 (submitted November 5, 1904). Excerpts already translated by Julian Barbour have been used. (Julian Barbour and Herbert Pfister (eds.)"Mach's Principle: From Newton's Bucket to Quantum Gravity.")

The most acute observations on the physical significance of the law of inertia and the related concept of absolute motion are due to Mach. According to him, in mechanics, just as in geometry, the assumption of an absolute space and, with it, an absolute motion in the strict sense is not permitted. Every motion is only comprehensible as a relative motion, and what one normally calls absolute motion is only motion relative to a reference system, a so-called inertial system, which is required by the law of inertia and has its orientation determined in accordance with some law by the masses of the universe.

Most authors are today in essential agreement with this point of view, as expressed most recently by Voss[1] and Poincaré[2] in particular. A different standpoint is adopted by Boltzmann,[3] who does not believe he can simply completely deny an absolute space and, with it, an absolute motion. Here, however, I shall proceed from Mach's view and attempt to add some further considerations to it. |

Mach summarizes his considerations in the following sentence:[4] "The natural [384] standpoint for the natural scientist is still that of regarding the law of inertia provisionally as an adequate approximation, relating it in the spatial part to the heaven of fixed stars and in the time part to the rotation of the Earth, and to await a correction or refinement of our knowledge from extended experience." Now it seems to me not entirely impossible that just such an extended experience could now be at hand. In a recent publication of K. R. Koch[5] on the variation in time of the strength of gravity we read: "Accordingly, the assumption of a genuine variation of gravity, or, more pre-

1 A. Voss, "Die Prinzipien der rationellen Mechanik" *Enzyklop. d. math. Wissensch.*, Band IV, 1, p. 39, 1901.

2 H. Poincaré, *Wissenschaft und Hypothese*. German translation by F. und L. Lindemann, Leipzig 1904.

3 L. Boltzmann, *Prinzipe der Mechanik*, II, p. 330, Leipzig 1904.

4 E. Mach, *Mechanik*, 4th ed. p. 252, Leipzig 1901.

5 K. R. Koch, Drude's *Annalen der Physik*, Band 15, p. 146, 1904.

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 3
Gravitation in the Twilight of Classical Physics: Between Mechanics, Field Theory, and Astronomy.

cisely, its difference between Stuttgart and Karlsruhe, seems to me appropriate." We shall naturally have to wait and see if this assertion stands up to further testing; at the least, we must now reckon with the real possibility that it is correct.

An explanation of such a phenomenon, if it is correct, would be very difficult on the basis of known causes. This circumstance encourages me to come forward now with a consideration that I have already developed earlier and long ago led me to the assumption that small periodic variations of gravity of measurable magnitude should be considered as a possibility.

Experience teaches us first that the inertial system required by the law of inertia can be taken to coincide with the heaven of the fixed stars to an accuracy adequate for practical purposes. It is also possible to choose a reference system differently, for example, fixed relative to the Earth, in order to describe the phenomena of motion. However, it is then necessary to apply to every material point the additional Coriolis
[385] forces of relative motion if one is to predict the motions correctly. One can therefore | say that the inertial system is distinguished from any other reference system by the fact that in it one can dispense with the adoption of the additional forces. Rectilinear uniform translation of the chosen reference system can be left out of consideration here as unimportant.

However, it is obvious that the fixing of the inertial system relative to the heaven of the fixed stars cannot be regarded as fortuitous. Rather, one must ascribe it to the influence, expressed in some manner, of the masses out of which it is composed. We can therefore pose the question of the law in accordance with which the orientation of the inertial system is determined when the instantaneous form and relative motion of the complete system of masses, i.e., the values of the individual masses, their separations, and the differential quotients of these separations with respect to the time, are regarded as given.

The logical need for such a formulation of the problem if one wishes to avoid the assumption of an absolute space was also felt by Boltzmann when he referred in passing to the possibility[6] that the three principal axes of inertia of the complete universe could provide the required orientation. If this rather natural supposition could be maintained, the conceptual difficulties would indeed be overcome. However, I believe that this supposition is not admissible. Let us imagine, for example, a universe that is otherwise arranged like ours but with the only difference that there are no forces at all between the individual bodies in the universe. Then for the inertial system valid for this universe, all the bodies in it would move along straight lines. However, a calculation that is readily made shows us that under this assumption the principal axes of inertia of the complete system would in general execute rotations relative to the inertial system. It is therefore necessary to look for a different condition that can enable us to understand the fixing of the inertial system. |

[386] If first we assume that all the bodies of the universe are at rest relative to each other except for a single mass point that I suppose is used to test the law of inertia,

6 Loc. cit., p. 333.

and which I will call the "test point," [*Aufpunkt*] then in accordance with the experiences we already have one could not doubt that the test point would, when no forces act on it, describe a straight path relative to a reference system rigidly fixed to the masses. In this case, the inertial system would be immediately fixed in space.

We can now imagine the case in which the bodies of the universe consist of two groups, one of which is "overwhelmingly" large compared to a smaller group and in which the masses within each group do not change their relative separations, whereas the smaller group, regarded in its totality, does carry out at the considered time a motion, say a rotation, relative to the larger group. If only one of the two groups were present, the inertial system would have to be fixed relative to it. Since the two work together, and one of the groups has been assumed to be much more "powerful" than the other, the inertial system will now be indeed very nearly at rest relative to the first group, but it will still execute a small motion relative to that group, which, of course, will be the consequence of the influence of the second, smaller group.

Given such a situation, what would be the most expedient way to proceed? I believe that one cannot be in doubt. One would fix the reference system exclusively using the first, overwhelming group and calculate as if this were the inertial system but take into account the influence of the second group by applying in this case to every test point the very weak additional forces of the relative motion that the chosen reference system executes relative to the true inertial system. If one makes such a decision, then these Coriolis forces no longer appear as mere computational quantities that arise from a coordinate transformation but as physically existing forces that are exerted by the masses of the smaller group | on every test point and arise because [387]
these masses have a motion relative to the chosen reference system.

To develop this idea further, one could start by investigating the case in which the second, smaller group that I just mentioned is represented by a single body. One then has the task of determining the magnitude and direction of the force, which will depend on the velocities of the single body and the test point relative to the reference system determined by the remaining bodies of the universe and on the separation between the single body and the test point. If we suppose that this problem has been solved for a single body, then, using the superposition law, we can also obtain the influence of a whole group of moving bodies.

The securely established observational results that are currently available are certainly not adequate to solve this fundamental problem; however, one does not therefore need to doubt that on the basis of further observations we could arrive at a solution.

After these preliminary considerations, I now turn to the case that corresponds to reality. Using the circumstance that the constellation of the fixed stars changes little in the course of several years or centuries, we can suppose that a reference system that more or less coincides with the inertial system is fixed relative to three suitably chosen stars. However, in order to take into account the small deviations that still remain, one must suppose that to each test point there are applied Coriolis forces, which, as we have just described, are to be interpreted as forces that depend on the velocities of the individual bodies in the universe and the velocity of the test point.

We are now in the position—and on this I put considerable value—to specify a condition meeting our requirement for causality that must be satisfied by the true inertial system required by the law of inertia. Namely, the true inertial system is the
[388] reference system for which all the | velocity-dependent forces that arise from the individual bodies of the universe are in balance at the test point. Even if in practice it is clear that we have not gained very much through this statement, it does appear to me that we have thereby obtained a very suitable basis for forming a clear concept of what is known as absolute motion in mechanics. There is at the least a prospect opened up of a way of determining the inertial system once the law that establishes the velocity-dependent forces has been found. In other words, it will be possible to construct the absolute space that appears in the law of inertia without having to sacrifice the notion that ultimately all motions are merely relative.

Besides, in all these considerations my main aim is to make it at least plausible that if one is to find a satisfactory solution to the questions that relate to the law of inertia it will be necessary to assume the existence of forces between the bodies in the universe that depend on their velocities relative to the inertial system. If this is accepted, then there follows the task of looking for possible phenomena whose relation to the expected general law of nature could be such that the law governing the velocity-dependent forces could be inferred. These forces, which for brevity I shall in what follows simply call "velocity forces," have nothing to do with gravitational forces, which arise concurrently with them, and specifically they can—and probably will— follow a quite different law than the gravitational forces with regard to distance dependence.

At this point I should like to make a remark in order to divide this communication into two quite separate sections. I believe that I can defend with complete definiteness and confidence what I have said up to now. However, I regard what follows as merely
[389] an attempt that could very | well fail; nevertheless, it is an attempt that at the least has a prospect of success and therefore must be brought forward at some time.

It seems to me that the most promising way of proving the existence of the postulated velocity forces and finding the law in accordance with which they act is to observe with the greatest possible accuracy phenomena associated with motions near the Earth that occur with great velocity. Just as the discovery of gravitation had as its starting point the observation of free fall, here too the first step to the solution of the puzzle could be obtained through observations of terrestrial motions and their correct interpretation. The immediate vicinity of the Earth's mass opens up some prospect of proving the existence of velocity forces more accurately than would be possible with the finest astronomical observations, which, as experience teaches, are certainly only very weak under normal circumstances.

This thought led me some time ago to make the gyroscope experiments that I reported to the Academy very nearly a year ago.[7] I expected then, as I explicitly said, to establish a behavior of the gyroscope that did not agree with the usual theory in the

hope that the observed deviation could be attributed to the velocity forces I seek and that these would therefore be made accessible to experimental research. Now certain indications of a deviation were indeed discernible, but as a careful and conscientious experimentalist I could not put any weight on them and I was forced, as I did, to declare a negative result of the experiment as regards the direction that it was intended to follow in the first place. In the meanwhile, I have made some further experiments with the same apparatus, though admittedly few, since they are very laborious and time consuming. However, the result could do nothing but | strengthen [390]
me in the view that the accuracy that can be achieved with this experimental arrangement is not sufficient to prove the existence of the velocity forces if they exist at all.

More promise of success probably lies in a further continuation of the free fall experiments, whose results to date can already be described as rather encouraging, after all. The ordinary theory, which does not take velocity forces into account, leads one to expect in the northern hemisphere, in addition to an easterly deviations of free fall motion from the vertical, a southerly deviation of such an extraordinarily small amount only that its experimental confirmation would be entirely out of the question. Nevertheless observers have time and again found southerly deviations of measurable magnitude, which are of a totally different order of magnitude (several hundred times larger and more) than those expected from theory. The newest observations in this area, due to the well-known American physicist E. H. Hall,[8] famous as an experimentalist and discoverer of "Hall's Phenomenon," have confirmed these experiences again. Indeed Hall considers further experiments on a larger scale (for greater heights of fall) necessary, and he holds out a prospect of such experiments. One may expect very valuable insights from them. Perhaps it will also serve to further the continued performance of such experiments if the hope for a positive result is strengthened by the theoretical considerations as I have offered here. For it takes indeed no small measure of courage to undertake painstaking and lengthy experiments, if the unanimous opinion of all theoreticians comes to this, that they cannot possibly lead to the expected result. This consideration has been for me the main motive to come forward with my views, although | I must admit that so far they are too deficient in an ade- [391]
quate experimental basis to be likely to meet with much approval.

Now I come to the admittedly most doubtful conjecture that I formed in connection with the above, and which is connected with the observation of Koch mentioned in the beginning. One can understand immediately that I must also expect velocity forces that arise from the motion of the Earth with respect to the Sun. The Sun is a fixed star like others and it contributes its part to the determination of the inertial system; or, in other words, it exerts velocity forces, if we account for the motion relative to a system of reference that is established without regard to the Sun. Even if nothing is known about the dependence of these forces on distance we may nonetheless

7 *Sitzungsberichte* 1904, p. 5.

8 Edwin H. Hall, *Physical Review,* 17, p. 179 and p. 245, 1903; further *Proceedings of the American Acad.* 39, No. 15, p. 399, 1904.

regard it as probable that the influence of a closer body is larger than that of a much more distant one. Therefore nothing is more natural than the assumption of velocity forces of such a kind that could cause a small periodic change of gravity with a diurnal as well as an annual period.

One difficulty, a very serious and possibly insurmountable one, arises only when one assumes that these velocity forces could be of such a magnitude that they would be measurable on the surface of the Earth and that the observations of Koch could be evaluated in this sense. Then one necessarily encounters the astronomers' objections, who have noticed nothing of the occurrence of such forces in spite of the great accuracy with which they can predict the phenomena of motion in the solar system.

This objection is so lucid that one would almost abandon the hope of being able to [392] silence it. To be sure, as long as nothing is conclusive about the | laws of action of the velocity forces in other ways, one could retreat to the view that a remote possibility exists that this contradiction could be cleared up later. And in this hope one could first quietly wait and see what consequences derive from such observational results as those found by Koch, temporarily neglecting the contradiction. If for example the original observation of Koch were not only confirmed, but if also the daily period of gravity fluctuations were really found as are expected on the basis of the views proposed here, then one could regard this as a certain confirmation of the suggested theory in spite of all objections.

However, I understand that such a position would be untenable. If it is not possible even now to make it reasonably credible that the interpretation of Koch's observations that I regard as possible does not necessarily have to be in contradiction with astronomical experience, then no one will pay heed to my interpretation, and the danger could arise that the same fate awaits Koch's observations as so far has befallen the southerly deviation of falling bodies, that is, that one does not take it seriously and immediately tends to assign it to errors in observation, because it does not fit with the accepted theory.

Only with this intention and by no means in order to represent the several possibilities I am about to discuss as somehow particularly probable, I still mention the following.

Consider a planet that circles its central body in agreement with the first two laws of Kepler. Let the law of the velocity forces be of the form that the planet is subject to an attraction by its sun that is proportional to the velocity component orthogonal to the radius vector and inversely proportional to the first power of the distance. One immediately recognizes that under these circumstances one would not need any grav- [393] itational force in addition to the velocity force in order to | explain the motion of the planet that is given by the observations. The astronomers of a solar system with only a single planet would have indeed no means to decide whether Newton's gravitational force or the velocity force adopted in the indicated manner were correct if they wished to restrict themselves to observation of the orbit alone. However, the difference would immediately be apparent when they took into account observations on their planet.

In accordance with Newton's gravitational law as well, there is, as is well known, a daily period of variation of the gravity force that gives rise to the contribution of the Sun to the motion of the tides but is too weak to be established by pendulum observations. However, if the astronomers of that solar system were to make the attempt to replace Newton's law of gravitation by the law of the velocity forces that we have mentioned, they would have to expect a much greater daily period, which, for the same relations between our Earth and the Sun, would be about 180 times greater than would be expected in the other case.

It should also be remarked here that the velocity law, which was chosen at random, is in fact only one of infinitely many that would all achieve the same, namely, the explanation of the motion of a single planet around its sun in agreement with Kepler's first two laws without having to invoke in addition Newton's gravitational force. All one needs to do is to allow the velocity component in the direction of the radius vector, which was hitherto assumed to be without influence, to participate as well in accordance with some arbitrary law and then arrange the law according to which the orthogonal velocity component acts on the force of attraction in such a way that the required motion results. There is also no need to make a restriction to the first power of the velocity; one could also consider the second or other powers. |

When a solar system has more than one planet, it is naturally much more difficult [394]
to explain all the planetary orbits merely with the help of velocity forces, since it is now necessary to satisfy Kepler's third law as well. So far as I can see, one would then be forced to make quite artificial assumptions. Even if one could achieve success in a simpler manner than it now appears to me, it would still be questionable if one could also explain the disturbances of the planetary orbits, the motions of the moons, etc.

However, one should not forget the aim of this discussion. It is in no way my intention to replace Newton's law by a law of velocity forces. I only want to make it plausible that under certain circumstances the velocity forces by themselves could have effects very similar to those of the gravitational forces. If this is then granted, it immediately follows that in such an event it would be very difficult to separate out from the astronomical observations the part due, on the one hand, to gravitational forces and, on the other, to the velocity forces.

On the basis of this consideration, I believe it is best not to be deflected by the admittedly very weighty objections of the astronomers from seeking phenomena that could be related to velocity forces. If it does prove possible, following this entirely independent research approach, to derive a law of the velocity forces, it will still be possible to make, as the best test of the admissibility of the result, an accurate comparison with the astronomical observations, taking into account the error limits that are relevant.

Naturally, I would not recommend such a procedure if I did not have great confidence in the very existence of the velocity forces, even though I must leave it as an
open question whether they have a magnitude such | that they are measurable in [395]
motions accessible to our perception. If one will admit an absolute space, then, of course, every reason for the assumption of velocity forces disappears. However, in

this point at least—that I do not recognize an absolute space—I am in agreement with the majority of natural scientists, and I therefore hope that I shall receive recognition among them, at least for the conclusions drawn in the first part of this communication.

AN ASTRONOMICAL ROAD TO
A NEW THEORY OF GRAVITATION

MATTHIAS SCHEMMEL

THE CONTINUITY BETWEEN CLASSICAL AND RELATIVISTIC COSMOLOGY IN THE WORK OF KARL SCHWARZSCHILD

1. KARL SCHWARZSCHILD: PIONEER OF RELATIVISTIC ASTRONOMY

Only a few weeks after Einstein had presented the successful calculation of Mercury's perihelion advance on the basis of his new theory of general relativity in late 1915, the German astronomer Karl Schwarzschild (1873–1916) published the first non-trivial exact solution of Einstein's field equations (Schwarzschild 1916a). The solution describes the spherically symmetric gravitational field in a vacuum and holds a central place in gravitation theory, comparable to that of the Coulomb potential in electrodynamics. It was not only an important point of departure for further theoretical research but also, up to recent times, the basis for all empirical tests of general relativity that proved not only the principle of equivalence but also the field equations themselves. Schwarzschild made a further substantial contribution to the theory when he found another exact solution describing the interior gravitational field of a sphere of fluid with uniform energy density (Schwarzschild 1916b). In this communication an important quantity makes its first appearance. It is the quantity that is later known as the *Schwarzschild radius*, which plays an important role in the theory of black holes many decades later.[1] But even long before the final theory of general relativity was established, Schwarzschild had already occupied himself with possible implications of its predecessors for astronomy; in 1913 he carried out observations of the solar spectrum in order to clarify if the gravitational redshift predicted by Einstein on the basis of the equivalence principle was detectable (Schwarzschild 1914).

In view of the fundamental role played by general relativity in astronomy, astrophysics, and cosmology today, it appears quite natural that an astronomer would engage in the study of this theory. Astronomical objects of all scales ranging from supermassive stars via galaxy nuclei and quasars to the universe as a whole are described on its basis. However, at the time when Schwarzschild made his contributions, the situation was quite different. None of the spectacular objects nowadays so successfully described by general relativity were in the focus of research, most of

1 For a thorough analysis of the early history of the interpretation of Schwarzschild's solutions and the Schwarzschild radius in particular, see (Eisenstaedt 1982; 1987; 1989). See also (Israel 1987), in particular sec. 7.7 on the *Schwarzschild 'Singularity'*.

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 3
Gravitation in the Twilight of Classical Physics: Between Mechanics, Field Theory, and Astronomy.

them not even known at all. Rather, the deviations from Newtonian theory that general relativity predicted were so small that in most cases they lay on the verge of detectability, even on astronomical scales. General relativity could thus easily be considered a physical theory—it was developed in the attempt to solve problems in physics such as the incompatibility of Newtonian gravitation theory and special relativity—with little implications on astronomy. And even as a physical theory it was still controversial, as is strikingly illustrated by the case of the physicist Max von Laue who as late as 1917 preferred Nordström's theory of gravitation to Einstein's.[2] Accordingly, at the time, astronomers showed little interest in general relativity. Einstein's plea to put the theory to an empirical test went unheard by most of them. In his attempts to provide empirical evidence for the theory, Erwin Freundlich, an outsider to the astronomical community, even met with hostility among Germany's most prominent astronomers.[3] Why was Schwarzschild an exception to this? What put him in the position to recognize so early the significance of general relativity?

The clue for answering these questions lies in the study of work Schwarzschild had done long before the rise of general relativity. In the course of the late 19th century, foundational questions surfaced in classical physics that had implicit consequences for astronomy: consequences that were often of a cosmological dimension. Mach's critique of Newton's absolute space, for example, immediately led to the question of an influence of distant stars on terrestrial physics. The deviation of the geometry of physical space from Euclidean geometry, to give another example, had become a possibility with the work of Gauss and Riemann and could be imagined to be measurable on cosmological scales. A further example is provided by the various attempts to modify Newton's law of gravitation. Such a modification would have consequences not only for planetary motion but also touches upon questions concerning the stability of the whole universe and the large-scale distribution of matter therein.[4] These foundational questions were, despite their astronomical implications, not on the agenda of contemporary astronomical research. Nevertheless, they were studied by a few individual scientists, among them Karl Schwarzschild.

In this paper it is argued that a continuity exists between Schwarzschild's prerelativistic work on foundational problems on the borderline of physics and astronomy and his occupation with general relativity. After a brief biographical introduction (sec. 2), Schwarzschild's prerelativistic considerations on the relativity of rotation (sec. 3) and on the non-Euclidean nature of physical space (sec. 4) are presented as they are documented in his publications as well as in his unpublished notes. On this background, Schwarzschild's reception of general relativity will then be shown to have been shaped to a large extent by his earlier experiences. In fact, what at first sight may appear to be a rather technical contribution to a physical theory—Schwarz-

2 See (Laue 1917).

3 For an account on Freundlich's work on empirical tests of general relativity and the astronomers' reaction to it, see (Hentschel 1997).

4 For a discussions of fundamental problems arising in Newtonian cosmology, see (Norton 1999).

schild's derivation of an exact solution of Einstein's field equations—turns out to have been motivated by Schwarzschild's concern for a consolidation of the connection between astronomy and the foundations of physics as established by Einstein's successful calculation of Mercury's perihelion motion (sec. 5). What is more, Schwarzschild was reexamining his prerelativistic cosmological considerations in the framework of the new theory of relativity as hitherto neglected manuscript evidence reveals for the case of the problem of rotation (sec. 6, a manuscript page from Schwarzschild's Nachlass is reproduced with annotations in the Appendix). Furthermore it turns out that, prepared by his earlier cosmological considerations, Schwarzschild was the first to consider a closed universe as a solution to Einstein's field equations (sec. 7). Summing up, Schwarzschild's road to general relativity may be called an astronomical one. Concluding this paper it will be argued that it was no coincidence that Schwarzschild of all astronomers took this road, but that this was the natural outcome of his interdisciplinary approach to the foundations of the exact sciences (sec. 8).

2. KARL SCHWARZSCHILD: ASTRONOMER, PHYSICIST AND ASTROPHYSICIST

Schwarzschild was born on October 9, 1873 in Frankfurt am Main, the eldest of seven children of a Jewish businessman.[5] He studied astronomy in Strasbourg and in Munich, where he obtained his doctoral degree in 1896 under Hugo von Seeliger (1849–1924), one of the most prominent German astronomers at the time. After having worked for three years at the Kuffner Observatory in Ottakring near Vienna, Schwarzschild obtained his post-doctoral degree (*Habilitation*) in Munich in 1899. On this occasion, Schwarzschild had to defend five theses, mostly concerned with foundational questions, that inspired him, as we will see, to much of the work relevant to our discussion. It is therefore interesting to question the extent to which Schwarzschild's teacher, von Seeliger, was involved in formulating these theses. While it may well be the case that Schwarzschild himself played some role in their creation, their exact wording makes it plausible that they were formulated by von Seeliger (see the discussion below). Thus this sheds some light on von Seeliger's ambivalent role in the early history of relativity. On one hand he was known to be very sceptical of relativity theory. For example, he severely criticized Erwin Freundlich's attempts to provide empirical evidence supporting general relativity. On the other hand he was interested in foundational questions of theoretical astronomy and apparently inspired Schwarzschild to much of the work discussed here.

In 1901, Schwarzschild was appointed professor of astronomy and director of the observatory of Göttingen University. He became closely associated with the circle of mathematicians and natural scientists around Felix Klein and furthered the integration of Göttingen astronomy with general scientific life.[6] Schwarzschild left Göttin-

5 For a short biographical account on Schwarzschild, see (Schwarzschild 1992, 1–25).

gen in 1909 and became director of the *Astrophysikalisches Institut* in Potsdam, but, for the short remainder of his life, he maintained the personal and scientific relationships established in Göttingen. Thus it was his Göttingen colleagues and acquaintances who, on several occasions, wrote him about the latest developments of Einstein's theory and pointed out the importance of its astronomical verification.[7] On May 11, 1916, Schwarzschild died an untimely death from a skin disease he contracted while serving at the Russian front.

Schwarzschild's scientific work is characterized by its rare breadth. The range of topics from physics and astronomy covered by his more than one hundred publications is hardly surpassed by any other single scientist of the twentieth century. Schwarzschild is further known to be one of the founders of astrophysics in Germany and was its most prominent exponent at the time. While disciplinary astrophysics itself was a rather specialized enterprise—using physical instruments for astronomical observation and applying physical theory to astronomical objects—Schwarzschild's interdisciplinary outlook on the foundations of science[8] enabled him to overcome the constraints imposed by specialization and deal with foundational problems on the borderline of physics and astronomy that were not in the focus of mainstream research.

3. SCHWARZSCHILD'S PRERELATIVISTIC CONSIDERATIONS ON THE RELATIVITY OF ROTATION

In 1897, while he was assistant at the Kuffner Observatory in Ottakring, Schwarzschild published a popular article entitled *Things at Rest in the Universe* (*Was in der Welt ruht,* Schwarzschild 1897). In this paper he discusses the relativity of motion and the problem of finding appropriate reference frames. In particular, he is concerned with the question of how fixed directions in space can be defined.

Schwarzschild's starting point is the observation that the motion of an object can only be perceived relative to other objects and that therefore any object may be considered at rest. The question of what thing is at rest in the universe should therefore be reformulated in a historical manner as "[w]hat things in the universe did one find useful to treat as being at rest, at different times [in history]?".[9] In the Copernican system, Schwarzschild explains, fixed directions in space were defined by reference

6 See (Blumenthal 1918).

7 See, in particular, Schwarzschild's correspondence with David Hilbert. On a postcard to Schwarzschild from October 1915, for example, Hilbert wrote: "The astronomers, I think, should now leave everything aside and only strive to confirm or refute Einstein's law of gravitation." ("Die Astronomen, meine ich, müssten nun Alles liegen lassen u. nur danach trachten, das Einsteinsche Gravitationsgesetz zu bestätigen oder widerlegen!") Hilbert to Schwarzschild, 23 October, 1915, N Briefe 331, 6r. (This and all following translations are my own, a few of them are based on the companion volumes to the Einstein edition.) Examples of this kind are also found in Schwarzschild's correspondence with Arnold Sommerfeld. A selection of Sommerfeld's scientific correspondence has recently been published (Sommerfeld 2000–2004).

8 See sec. 8.

to the system of fixed stars. Towards the end of the 17th century it became clear however that the Copernican stipulation is not unambiguous: the stars perform motions relative to one another, the so-called proper motions. Schwarzschild therefore next considers the electromagnetic aether as a candidate for a material reference of rest but comes to the conclusion that the aether too cannot serve such a purpose since it is affected by ponderable matter moving through it. Schwarzschild concludes that there are no material objects in the universe that one could reasonably consider at rest and that one can only take resort to "certain conceptually defined points and directions that may serve as a substitute to a certain extent".[10]

In order to explain how fixed directions in space may be defined on the basis of the law of inertia, Schwarzschild refers to Foucault's pendulum. By accurate observations of the rotation of the pendulum's plane of oscillation, Schwarzschild explains, one could calculate the speed of rotation of the Earth, and would then have to describe as fixed the direction with respect to which the Earth rotates with the calculated speed.

In following this idea further, Schwarzschild establishes an interesting connection between inertia and gravitation in the following way. In regarding the planets orbiting around the Sun as gigantic, diagonally pushed pendulums, he conceives an astronomical realization of the physical model of the pendulum. In analogy to Foucault's pendulum, fixed directions in space are then given by the aphelia (or perihelia) of the orbits of the different planets. However, Schwarzschild explains, astronomical observations since the middle of the 19th century reveal that the directions singled out by the orbits of the different planets rotate with respect to each other at a very slow rate, so that it is "impossible to consider all as fixed."[11] Although Schwarzschild was aware of possible astronomical explanations, such as interplanetary friction, he considered it more probable that an explanation of these small anomalies has to go further, requiring a revision of the classical law of gravitation.

In this way, Schwarzschild established a relation between the two physical phenomena, inertia and gravitation, the integration of which was later to lie at the basis of Einstein's theory of general relativity. Moreover, the observational fact by which Schwarzschild links the two phenomena—the perihelion shift of the inner planets—was later to play a crucial role in the establishment of general relativity, for some years being the only empirical fact suggesting a superiority of general relativity over the Newtonian theory.

There are, of course, fundamental aspects of general relativity that have no analogue in Schwarzschild's prerelativistic considerations. Most notably, Schwarzschild did not consider a field theory of gravitation that unifies gravitation and inertia in one

---

9 "[w]as in der Welt hat man zu verschiedenen Zeiten als ruhend zu betrachten für gut befunden?" (Schwarzschild 1897, 514). All page numbers cited for this text refer to vol. 3 of the *Collected Works* edition (Schwarzschild 1992).

10 "[...] gewisse begrifflich definierte Punkte und Richtungen, die einigermaßen als Ersatz eintreten können [...]" (Schwarzschild 1897, 516).

11 "[...] unmöglich alle als fest betrachtet werden können." (Schwarzschild 1897, 520.)

single field. In fact, in this text, Schwarzschild does not even question the origin of inertia. Unlike Mach and Einstein, he does not search for a physical cause of inertia but rather assumes inertia to be given and, on its basis, defines fixed directions in space. Most probably he therefore thought of modifications of the Newtonian law of gravitation that do not affect inertial frames. For example, it was well known at the time that the change of the exponent in Newton's inverse square law yields perihelion motions.[12] Such a motion could have easily been subtracted from the observed motions in order to obtain the "true" inertial directions in space given by the planets' orbits. There are however notes found in Schwarzschild's manuscripts that show that he was concerned with the question of the origin of inertia and that, in this context, he considered the possibility of local inertial frames rotating with respect to one another. These notes, in which Schwarzschild was again using orbits of celestial bodies in order to determine inertial directions, shall now be discussed.

As explained in sec. 2, Schwarzschild had to defend five theses, probably formulated by his teacher von Seeliger, in order to obtain his post-doctoral degree in 1899. One of these theses read: "The existence of centrifugal forces is comprehensible only under the assumption of a medium pervading all of space."[13] In a notebook of 1899 (N 11:17),[14] we find Schwarzschild's tentative defense of this thesis. In a thought experiment reminiscent of Einstein's later ones, attempting to clarify the nature of rotation, Schwarzschild imagines two planets of identical constitution rotating with different angular velocity and having atmospheres that are so dense that the outer world cannot be observed.[15] An inhabitant of one of the planets travelling to the other would have no way of understanding how the difference in the "gravitational conditions" (*Schwereverhältnisse*) arises, since he would not notice the rotation. This shows clearly, Schwarzschild explains,

12 Thus, Schwarzschild's teacher Hugo von Seeliger wrote in a letter to Arnold Sommerfeld: "that the law of attraction $1/r^n$ $(n \neq 2)$ [causes] perihelion shifts, that is known to any astronomer since time immemorial. [...] *Newton* already treated this case, or a quite similar one, in his 'Principia.'" ("[...] daß das Anziehungsgesetz $1/r^n$ $(n \neq 2)$ Perihelbewegungen [hervorruft], das ist jedem Astronomen seit jeher bekannt. [...] Schon *Newton* hat diesen oder einen ganz ähnlichen Fall in den 'Prinzipien' behandelt.") Hugo von Seeliger to Arnold Sommerfeld, May 25, 1902, Arnold Sommerfeld Nachlass, Deutsches Museum, Munich, HS 1977-28/A, 321, 1-1.

13 "Die Existenz von Centrifugalkräften ist nur unter der Annahme eines den ganzen Raum erfüllenden Mittels zu begreifen." A document naming the five theses can be found in N 21 (for an explanation of this notation see the next footnote).

14 Here and in the following, references to Karl Schwarzschild's Nachlass in the Niedersächsische Staats- und Universitätsbibliothek Göttingen are indicated by an archival number following an 'N' (e.g. N 11:17).

15 Consider, for example, Einstein's thought experiment involving two identical fluid bodies rotating with respect to one another (Einstein 1916, 771–772). Schwarzschild's thought experiment is further reminiscent of a later one by Poincaré, who also considered a planet covered by clouds so that its inhabitants cannot see the sun or stars. They therefore, Poincaré argued, would have to wait longer than we did until a "Copernicus" arrived, who could explain the centrifugal and Coriolis forces by assuming that the planet rotates (Poincaré 1902, chap. 7).

> that not only the internal relative circumstances but also the relations to the surrounding space have an influence on the processes in a system of bodies. Following Newton we could state that there is an absolute space and that the relation of motions to this absolute space has an influence on the forces appearing through this motion. Or, in other words: absolute space has an effect on the bodies. Now, we are used to thinking of anything having an effect as something real, namely something material, and from this it follows that, if we want to stick to the usual way of thinking, we have to imagine space, Newton's absolute space, filled with a substance.[16]

The hypothetical identification of space with a substance now puts Schwarzschild in a position to discuss the global validity of the locally distinguished directions:

> This substance does not have to be at absolute rest, but only in a state of motion that in some way distinguishes three fixed directions in space [...]. Then it is comprehensible that the centrifugal forces are based on a relation of the motion of the usual bodies to the motion of this substance.[17]

From the observation that the perihelia of double stars are at rest with respect to the directions that seem fixed inside the solar system, Schwarzschild concludes that the directions distinguished in their region of space have to be the same as in the solar system.

To sum up, while in the previous example, Schwarzschild had established a relation between inertia and gravitation, here he relates inertia to the structure of space, considering the possibility that the local inertial directions may vary on cosmological scales.

4. SCHWARZSCHILD'S PRERELATIVISTIC CONSIDERATIONS ON NON-EUCLIDEAN COSMOLOGY

A second example for Schwarzschild's prerelativistic treatment of foundational questions having cosmological implications is provided by his application of non-Euclidean geometry to physical space. Again, this work appears to have been inspired by one of the five theses Schwarzschild had to defend in order to attain his degree. This thesis reads: "The hypothesis that our space is curved should be rejected".[18] It is

---

16 "[...] daß auf die Vorgänge in einem Körpersystem nicht nur die inneren relativen Verhältnisse, sondern auch die Beziehungen zum Raum, der sie umgiebt, von Einfluß sind. Wir könnten mit Newton sagen, daß es einen absoluten Raum giebt und daß das Verhalten der Bewegungen zu diesem absoluten Raum auf die bei der Bewegung auftretenden Kräfte von Einfluß ist. Oder in anderen Worten: der absolute Raum hat eine Wirkung auf die Körper. Nun pflegen wir uns aber alles, was eine Wirkung hat, als etwas wirkliches, nämlich als etwas Materielles zu denken, und daraus folgt, daß wir, wenn wir überhaupt in der üblichen Denkweise bleiben wollen, uns den Raum, Newtons absoluten Raum durch einen Stoff erfüllt denken müssen." (N 11:17, 8v–9r.) There are no page numbers in this notebook. The page numbers given here refer to my pagination.

17 "Dieser Stoff muß nicht absolut ruhen, sondern nur eine Bewegungsform haben, welche auf irgend eine Weise drei besondre feste Richtungen auszeichnet [...]. Dann ist begreiflich, daß die Centrifugalkräfte auf einer Beziehung der Bewegung der gewöhnlichen Körper zur Bewegung dieses Stoffes beruhen." (N 11:17, 9r.)

18 "Die Hypothese einer Krümmung unseres Raumes ist zu verwerfen" (N 21).

plausible to assume that this thesis too was formulated by von Seeliger. In fact, the thesis seems to reflect von Seeliger's attitude toward the application of non-Euclidean geometry to physics and astronomy which was extremely sceptical as may be illustrated by the following passage from a talk by von Seeliger entitled *Remarks on the So-Called Absolute Motion, Space, and Time*:

> [...] the common and therefore very fatal misapprehension has emerged that one believed to be able to decide by measurement which geometry is the "true" one, or even, which space is the one in which we live. From the stand point taken here the latter formulation is by far the more dangerous one, since space in itself has no properties at all.[19]

In the above-mentioned notebook of Schwarzschild we find an entry in which the thesis is slightly reformulated as follows: "The assumption of a curvature of our space is without any advantage for the explanation of the structure of the system of fixed stars".[20] The note is accompanied by considerations and calculations in which Schwarzschild examines empirical consequences of a curvature of space, for example, on the parallaxes of stars. One year later, Schwarzschild published a more detailed account of these considerations, though now under a different perspective. While the notebook entries aimed at a rejection of the curvature of space—understandably so in view of their context, the defense of a thesis—the purpose of the published article is to estimate the degree of curvature that can be assumed without contradicting observation. The article bears the title *On the Permissible Scale of the Curvature of Space (Über das zulässige Krümmungsmaass des Raumes,* Schwarzschild 1900).[21]

In his article Schwarzschild mainly discusses two cases: hyperbolic space, having constant negative curvature, and spherical space, having constant positive curvature. He makes the assumption that light travels along geodesics.

As far as hyperbolic space is concerned, Schwarzschild is able to estimate a minimal radius of curvature with the help of the parallax. As is well known, the parallax of a star, for simplification assumed to be nearly perpendicular to the ecliptic, is defined as half the difference of the two angles under which the star is seen in an interval of half a year, $\pi = (\alpha - \beta)/2$ (see Fig. 1). (In Euclidean space this coincides with the angle under which the radius of the Earth's orbit is seen from that star.) In Euclidean space, therefore, a parallax of exactly zero implies that the star is infinitely far from the Earth, since parallel geodesics in Euclidean space intersect at infinity. In hyperbolic space, in contrast, neighboring geodesics diverge. Thus even stars infinitely remote from the Earth possess a certain parallax. This minimal paral-

19 "[Es] ist der verbreitete, aber gerade darum sehr verhängnisvolle Irrtum entstanden, daß man glaubte durch Messungen entscheiden zu können, welche Geometrie die "wahre" ist, oder gar, welcher Raum der ist, in dem wir leben. Von dem hier vertretenen Standpunkt aus ist die letztere Fassung die bei weitem gefährlichere, da der Raum an sich überhaupt keine Eigenschaften hat." (Seeliger 1913, 200–201.)

20 "Die Annahme einer Krümmung unseres Raumes ist ohne jeden Vorteil für die Erklärung des Baues des Fixsternsystems." (N 11:17, 4r.)

21 The article is based on a talk Schwarzschild held at the Heidelberg meeting of the *Astronomische Gesellschaft* in 1900.

lax decreases with an increase in the radius of curvature. Since for most stars no parallax can be observed, the minimal parallax is given by the accuracy of observation. This is given by Schwarzschild as 0.05″. From this he concludes that the curvature radius of hyperbolic space must be at least 4 million times the radius of the Earth's orbit.

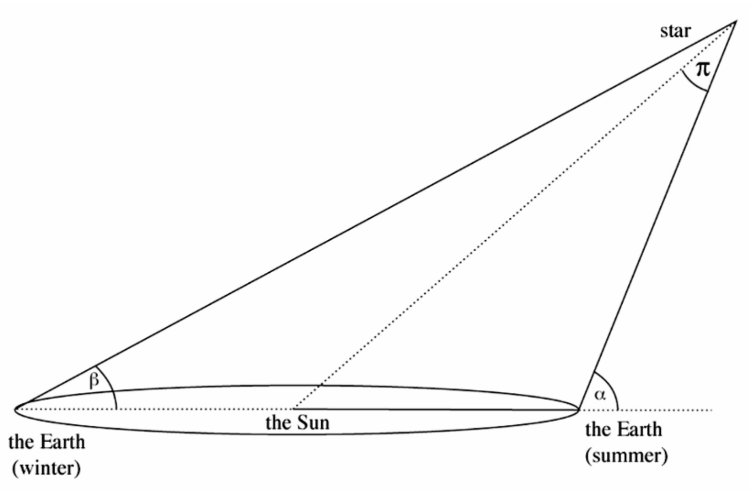


Figure 1: The annual parallax of a star nearly perpendicular to the ecliptic

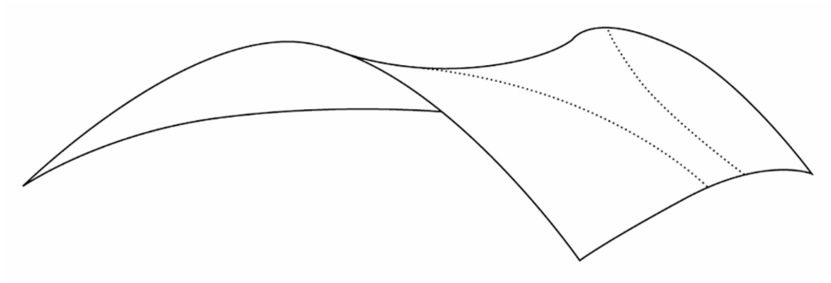


Figure 2: Two-dimensional hyperbolic space. Two geodesics are drawn as dotted lines

As concerns spherical space, Schwarzschild discusses the special case of an elliptic space. The latter can be obtained from usual spherical space by identifying antipodal points. As a consequence, two geodesics going around the world intersect at only one point. Schwarzschild's reason for preferring elliptic to spherical space is that, in the latter, light emitted at one point in space in different directions would converge on the antipodal point, a rather artificial-looking consequence which, according to Schwarzschild, one would not accept without being forced to.

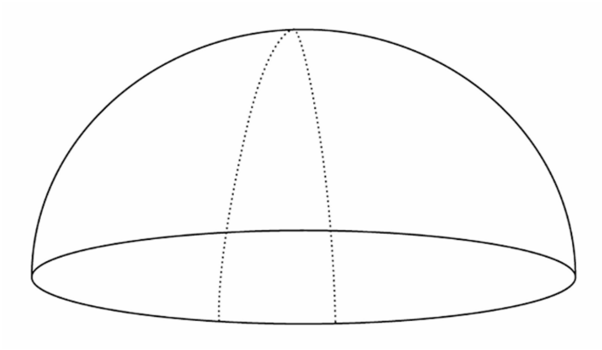


Figure 3: Two-dimensional elliptic space (the antipodal points on the circle c are to be identified). Two geodesics are drawn as dotted lines

In the case of elliptic space there is no minimal parallax and physical considerations are required in order to determine a minimal radius of curvature. Schwarzschild offers the following reasoning. In elliptic space neighboring geodesics converge and thus intersect already at a finite distance, namely at the distance $(\pi/2)R$, where R denotes the curvature radius. Stars having a parallax smaller than a certain given value, say $0.1''$, therefore have all to be located within a finite volume. Now, there are approximately 100 stars of parallax above $0.1''$. All other stars are thus to be found in this finite volume. If one assumes a uniform distribution of stars, one can determine a certain minimal radius of curvature. A weaker requirement, however, is that the stars with parallax less than $0.1''$ occupy a volume large enough so that they do not influence each other in a way that could not have escaped observation. Schwarzschild does, for instance, calculate that if the elliptic space had a curvature radius of about 30,000 times the radius of the Earth's orbit, stars at great distances from the Earth would be separated from one another by only about 40 times the radius of the Earth's orbit. The physical interactions between the stars resulting from this could hardly be concealed from observation. From these considerations Schwarzschild concludes that the minimal radius of curvature of elliptic space is of the order of 100 million times the radius of the Earth's orbit. Schwarzschild further argues that such a relatively small radius of curvature (roughly 1600 light years) is only a realistic possibility if one further assumes an absorption of the starlight of about 40 magnitudes in one circulation around the universe because it is only under this assumption that the appearance of a counter image of the Sun can be avoided.

In this article, as in a later one (Schwarzschild 1909), Schwarzschild expresses his preference for the elliptic space over the hyperbolic or even the Euclidean one, because its finiteness would make it possible in principle to investigate the macroscopic world exhaustively. This idea would have a soothing effect on the mind.

While the differences between Schwarzschild's application of non-Euclidean geometry to physical space on one side and modern cosmology on the other are obvious (application to three-dimensional space rather than to four-dimensional spacetime, no dynamics of geometry, consideration of scales that today are hardly considered cosmological), Schwarzschild's consideration also contains striking parallels to the modern treatment of the problem such as the idea that light proceeds along geodesics and, most notably, the possibility that the universe is spatially closed.[22]

5. THE PERIHELION BREAKTHROUGH

In sec. 3 we have seen how Schwarzschild put the perihelion anomalies of the inner planets into the context of the fundamental physical phenomena of inertia and gravitation. In view of this, Einstein's successful calculation of Mercury's perihelion motion on the basis of his new theory must have appeared to Schwarzschild as the realization of his earlier speculations on the relations between physics and astronomy. In this sec. it will be argued that it was indeed Einstein's perihelion result that instigated Schwarzschild's interest in general relativity. It turns out that even Schwarzschild's derivation of his first exact solution was motivated by his concern to consolidate Einstein's result.

As early as 1912 Schwarzschild had been confronted with the question of the observability of astronomical consequences of general relativity and its predecessors—in particular consequences of the principle of equivalence. Interestingly, in view of Schwarzschild's correspondence, it was not Einstein himself who first confronted Schwarzschild with the question of astronomical consequences of such a theory but rather one of his antagonists in the search for a new theory of gravitation: the theoretical physicist Max Abraham (1875–1922). Abraham, at that time holding a post as professor of rational mechanics at the University of Milan, was himself working on a new theory of gravitation on which he had already published.[23] Although Einstein and Abraham were in severe disagreement about the foundations the new theory of gravitation should build upon, some empirical consequences of Abraham's theory coincided with those Einstein had derived from his more general considerations. Thus, in his first publication on the matter, Abraham discusses the bending of light in a gravitational field that follows from Huygens' principle whenever the speed of light is assumed to be variable, and, in a footnote, points out that Einstein has drawn the astronomers' attention to the fact that the bending of star light in the gravitational field of the Sun may be observable (Abraham 1912, 2).

There are two remnants of Schwarzschild's correspondence with Abraham found in Schwarzschild's Nachlass. The first is a draft in Schwarzschild's hand of a letter most probably addressed to Abraham,[24] the second is a letter from Abraham to Schwarzschild, dated October 13, 1912 (N Briefe 5). From Schwarzschild's draft it

---

22 In an addendum to his article, Schwarzschild mentions a further possibility that later became a debated subject in relativity theory: the possible application of different topologies to physical space.

23 His first publication on that matter being (Abraham 1912).

becomes apparent that Abraham had previously raised the question whether there would be an effect recognizable through astronomical observation if the Sun's loss of inertial mass was proportional to the energy it radiates away, while the gravitational mass did not change in this proportion. In his letter from October 13, 1912, Abraham formulated another idea, arguing that the energy loss of the planets when cooling down in the process of the genesis of the solar system must have diminished the inertial and the gravitational mass in equal proportion, since otherwise Kepler's third law of planetary motion could not be valid. Finally, in their correspondence, the two discussed the possible shift of spectral lines in a gravitational field. Here, as elsewhere in his correspondence and writings prior to Einstein's perihelion result, Schwarzschild is very sceptical about the astronomical detectability of the predicted effects, although he appears to regard the nascent theory of relativity with openness. Thus, as concerns the redshift in the solar spectrum, Schwarzschild writes:

> The shift of the wavelengths on the Sun that Einstein demands, exists [...] due to a strange coincidence in exactly the right magnitude. There is, however, no doubt that it is to be blamed partly on pressure and partly on downwards motions in the solar atmosphere. To see more clearly in this respect, one has to study the lines at the different points of the solar disk. Until now, this has only been done in a really sufficient manner for the lines 3933 Å of calcium (St. John, Astrophysical J[ournal).] His results do in fact speak *against* the existence of the sought-after shift. Despite this, I do not want to claim that this is already an absolute veto against the theory. Before that, more lines would have to be equally well investigated.[25]

In 1913 Schwarzschild himself started an investigation of exactly the kind he had spoken of in his letter to Abraham, performing a series of observations of the band at 3883 Å in the solar spectrum. The continuation of these observations was foiled by the outbreak of war in 1914, but Schwarzschild reported on the results so far obtained in a communication that was presented to the Prussian Academy of Sciences by Einstein on November 5, 1914 (Schwarzschild 1914). In the introduction to this commu-

---

24 N Briefe 846. The draft is dated in another hand as "1912" and commented on as "possibly [to] W. Lorey." The contents however makes it most probable that it is the draft of a letter to Max Abraham. It may be dated September 29, 1912 (or a little earlier) on the basis of Abraham's letter to Schwarzschild from October 13, 1912, mentioning a letter from Schwarzschild from September 29: most probably the letter that was written on the basis of the draft.

25 "Die Verschiebung der Wellenlängen auf der Sonne, die Einstein fordert, besteht [...] durch einen merkwürdigen Zufall genau in der richtigen Größe. Es ist aber kein Zweifel, daß dieselbe zum Teil auf Druck, zum Teil auf absteigende Bewegungen in der Sonnenatmosphäre zu schieben sind. Um Klarheit darüber zu bekommen, muß man die Linien an den einzelnen Punkten der Sonnenscheibe studieren. Das ist in wirklich ausreichender Weise bisher nur für die Linien 3933 A. E. des Calciums geschehen (St. John, Astrophysical J. dessen Resultate sprechen durchaus *gegen* die Existenz der gesuchten Verschiebung. Trotzdem möchte ich nicht behaupten, daß hiermit schon ein absolutes Veto gegen die neue Theorie gegeben ist. Es müßten doch erst noch mehr Linien gleich gut untersucht werden." Draft of a letter from Schwarzschild to Abraham, probably September 29, 1912, N Briefe 846, 27v, 28r. St. John's publications in the *Astrophysical Journal* on the motion of calcium vapour in the solar atmosphere are (St. John 1910; 1910–11). For an account on St. John's later work on line shifts and its relation to general relativity, see (Hentschel 1993).

nication, Schwarzschild points out that the observation of the solar spectrum is of interest not only for the sake of solar physics, but "can, according to Mr. Einstein, inform us about the relativity of the world".[26] By referring to Einstein's article on the influence of gravitation on the propagation of light (Einstein 1911), Schwarzschild explicitly relates the gravitational redshift to the principle of equivalence. However, once more Schwarzschild does not conclude affirmatively, describing his preliminary results and other astronomers' observations he reports on as still being indecisive concerning the gravitational redshift.

In his correspondence with Max Planck in 1913, Schwarzschild even more clearly expresses his doubts concerning an astronomical verification of Einstein's theory. In a letter from January 31, 1913, Planck had asked Schwarzschild for an assessment of the feasibility and the expenses of the eclipse expedition that Erwin Freundlich was planning for the year 1914 and for which he was going to apply to the *Preussische Akademie der Wissenschaften* for funding (N Briefe 593, 2r, v). Freundlich intended to search for a deflection of starlight near the solar disk as predicted by Einstein. Schwarzschild commented on the observational side of the problem in the following way:

> In the problem itself I also have no particular confidence. The diminution of the frequency on the Sun and the shift to red of all spectral lines on the Sun that Einstein assumes can be regarded as refuted by the observations. The last word has not yet been spoken, but the shifts which for single lines are also to violet, can be too well interpreted as being due to pressure. Since this whole thing looks rather fishy, it won't be much different for the deflection of light rays by the Sun's gravitation.[27]

When Einstein succeeded in deriving the correct value for Mercury's perihelion shift from his theory,[28] Schwarzschild's appraisal of the new theory of relativity changed drastically. Einstein presented his calculation of Mercury's perihelion advance to the Prussian Academy of sciences on November 18, 1915. Schwarzschild was on leave from his military duties at the Russian front and attended the meeting.[29] Back in Russia, Schwarzschild wrote to Einstein:

> It is quite a wonderful thing that from such an abstract idea the Mercury anomaly emerges so stringently.[30]

---

26 "[...] kann nach Hrn. Einstein auch Auskunft über die Relativität der Welt geben." (Schwarzschild 1914, 1201.)

27 "Auch zum Probleme selbst habe ich kein besonderes Fiduz. Die Verminderung der Schwingungszahl auf der Sonne und die [...] Verschiebung aller Spektrallinien nach Rot auf der Sonne, die Einstein annimmt [...] kann als durch die Beobachtungen als widerlegt angesehn werden. Das letzte Wort ist noch nicht gesprochen, aber die [...] Verschiebungen, die [...] bei einzelnen Linien auch nach Violett gehen, lassen sich zu gut als Druckverschiebungen deuten. Da es hiermit ziemlich faul [?] steht, wird es mit der Ablenkung der Lichtstrahlen durch die Sonnengravitation auch nicht viel anders sein." Draft of a letter from Schwarzschild to Planck, after January 31, 1913, N Briefe 593, 6r. The passages omitted are crossed-out in Schwarzschild's manuscript.

28 On Einstein's derivation and its historical context, see (Earman and Janssen 1993).

29 See the minutes of the meeting on November 18, 1915, *Archiv der Berlin-Brandenburgischen Akademie der Wissenschaften* II–V, Vol. 91, 64–66.

In a letter of the same day to Arnold Sommerfeld, Schwarzschild even explicitly states that to him the perihelion result was much more convincing than the empirical consequences of Einstein's theory discussed earlier:

> Did you see Einstein's paper on the motion of Mercury's perihelion in which he obtains the observed value correctly from his last theory of gravitation? That is something much closer to the astronomers' heart than those minimal line shifts and ray bendings.[31]

It is a matter of course that the quantitatively appropriate explanation of an anomaly that had been detected by astronomers more than half a century earlier provided a stronger argument for the new theory than the hardly detectable effects that it also predicted. However, Einstein's successful calculation of the perihelion shift did not convince everybody to the same degree as Schwarzschild. More than one year after Einstein's calculation, Max von Laue still described the result as a "agreement of two single numbers"[32] which

> remarkable as it may be, does not seem to us to give sufficient reason to change the whole physical world picture in its foundations, as Einstein's theory does.[33]

In view of Schwarzschild's earlier contextualization of perihelion motions, it becomes understandable why, for him, Einstein's result signified much more than the "agreement of two single numbers." Furthermore, on the background of Schwarzschild's prerelativistic cosmological considerations that included the application of non-Euclidean geometry to physical space and the idea of mutually accelerated inertial systems, the changes brought about by Einstein's theory must have appeared less drastic to Schwarzschild than to most others, including von Laue.

Einstein's derivation of the perihelion advance which was based on an approximation had, however, one blemish: the uniqueness of the solution remained questionable. In order to consolidate Einstein's result, Schwarzschild tried to prove the uniqueness of the solution. In the above-mentioned letter to Sommerfeld, Schwarzschild reports:

> In Einstein's calculation the uniqueness of the solution remains doubtful. In the first approximation, which Einstein makes, the solution, when carried out completely, is even apparently ambiguous—one additionally gets the beginning of a divergent expansion. I have tried to derive an exact solution, and that was unexpectedly easy.[34]

---

30 "Es ist eine ganz wunderbare Sache, daß von so einer abstrakten Idee aus die Erklärung der Merkuranomalie so zwingend herauskommt." Schwarzschild to Einstein, December 22, 1915 (CPAE 8, Doc. 169).

31 "Haben Sie Einstein's Arbeit über die Bewegung des Merkurperihels gesehen, wo er den beobachteten Wert richtig aus seiner letzten Gravitationstheorie heraus bekommt? Das ist etwas, was den Astronomen viel tiefer zu Herzen geht, als die minimalen Linienverschiebungen und Strahlenkrümmungen." Schwarzschild to Sommerfeld, December 22, 1915, München, Deutsches Museum Archiv NL 89, 059, p. 1.

32 "Übereinstimmung zwischen zwei einzelnen Zahlen"

33 "[...] scheint uns, so bemerkenswert sie ist, doch kein hinreichender Grund, das gesamte physikalische Weltbild von Grund aus zu ändern, wie es die Einsteinsche Theorie tut." (Laue 1917, 269.)

And even in his publication that is today known for containing the first derivation of an exact non-trivial solution of Einstein's field equations, Schwarzschild emphasizes that, rather than the quest for an exact solution, it is the consolidation of Einstein's result which is of primary concern:

> It is always convenient to possess exact solutions of a simple form. More important is that the calculation yields, at the same time, the uniqueness of the solution about which Mr. Einstein's treatment remained doubtful and which arguably, in view of the way in which it emerges below, could hardly have been proven by such an approximative method.[35]

6. THE RELATIVITY OF ROTATION REVISITED

After his consolidation of the connection between general relativity and observational astronomy established by Einstein's perihelion calculation, Schwarzschild turned to other questions of theoretical astronomy for which general relativity appeared to provide the adequate framework. One of these questions concerned the relativity of rotation, a question we already encountered in Schwarzschild's prerelativistic work.

The major source documenting Schwarzschild's work on this question in the context of general relativity is a formerly unrecognized manuscript page in Schwarzschild's Nachlass in the University Library of Göttingen (N 2:2, 12r). A reproduction with explanations of the page is given in the Appendix. The page is full of calculations and contains hardly any text. It is found among a few similar pages, some of which contain notes on general relativity the purpose of which however is not obvious. The notes on the page under discussion are undated but obviously stem from the short period between Einstein's successful perihelion calculation in November 1915 and Schwarzschild's death in May 1916.

In these notes, Schwarzschild distinguishes an "inner" and an "outer" metric. The inner metric describes a Minkowski spacetime in a coordinate system rotating with constant angular velocity n. Using cylindrical coordinates, the outer metric can be written as

---

34 "Bei Einstein's Rechnung bleibt die Eindeutigkeit der Lösung noch zweifelhaft. In der ersten Annäherung, die Einstein macht, ist die Lösung sogar, wenn man sie vollständig macht, scheinbar mehrdeutig — man bekommt noch den Anfang einer divergenten Entwicklung herein. Ich habe versucht, eine strenge Lösung abzuleiten, und das ging unerwartet einfach." Schwarzschild to Sommerfeld, December 22, 1915, München, Deutsches Museum Archiv NL 89, 059, 1. In his letter to Einstein from December 22, 1915, Schwarzschild reports in even more detail about the motivation that led him to his exact solution, see (CPAE 8, Doc. 169).

35 "Es ist immer angenehm, über strenge Lösungen einfacher Form zu verfügen. Wichtiger ist, daß die Rechnung zugleich die eindeutige Bestimmtheit der Lösung ergibt, über die Hrn. Einsteins Behandlung noch Zweifel ließ, und die nach der Art, wie sie sich unten einstellt, wohl auch nur schwer durch ein solches Annäherungsverfahren erwiesen werden könnte." (Schwarzschild 1916a, 190.)

$$g_{\mu\nu_{\text{outer}}} = \begin{bmatrix} -f_1 & 0 & 0 & 0 \\ 0 & -f_2 & 0 & f \\ 0 & 0 & -1 & 0 \\ 0 & f & 0 & f_4 \end{bmatrix},$$

where x_1 is a radial coordinate, x_2 is an angle, x_3 is parallel to the symmetry axis of the cylinder, and x_4 is a time-like coordinate. f and f_i, $i = 1, 2, 4$, are functions of x_1 only. For the radial coordinate becoming infinitely large, Schwarzschild imposes the condition that the outer metric tends to the (non-rotating) Minkowski metric, rescaled in such a way that it satisfies the determinant condition, $|\det(g_{\mu\nu})| = 1$.[36] The spacetime Schwarzschild attempts to investigate thus consists of a cylindrical section of Minkowski space, the inner space, rotating with constant angular velocity n relative to an inertial frame at radius infinity and surrounded by an outer space, becoming Minkowskian for $x_1 \to \infty$.

Schwarzschild then obviously tries to find general expressions for the metric functions f and f_i, $i = 1, 2, 4$ (the g_{33}-component of the outer metric Schwarzschild had set to -1). In this he follows exactly the procedure he elaborated in his publication on the field of a point mass (Schwarzschild 1916a). First, he constructs the Lagrangian of a point particle

$$F = \frac{1}{2}\sum g_{\mu\nu}\frac{dx_\mu}{ds}\frac{dx_\nu}{ds}.$$

Next, he calculates $\partial F/\partial x_i$ and $\partial F/\partial \dot{x}_i$ for $i = 1, 2, 3, 4$ and puts the resulting terms into the Euler-Lagrange equation

$$\frac{\partial F}{\partial x_i} - \frac{d}{ds}\left(\frac{\partial F}{\partial \dot{x}_i}\right) = 0.$$

He then manipulates the resulting equations—further using the determinant condition $f_1(f_2 f_4 + f^2) = 1$—in such a way that he may read off the field strengths $\Gamma^\lambda_{\mu\nu}$ by a comparison of the coefficients with those appearing in the equations of motion of a point particle. He then puts the expressions for the field strengths into the field equations and manipulates them in order to determine the functions f and f_i by integra-

36 As in (Schwarzschild 1916a), Schwarzschild here employs the condition that the determinant of the metric tensor be unity. Einstein had introduced this condition in an addendum to his publication *Concerning the Theory of General Relativity* (*Zur allgemeinen Relativitätstheorie (Nachtrag)*, Einstein 1915a), and continued to use it in his paper on the perihelion motion of Mercury (Einstein 1915b). The field equations Einstein presented and used in these papers read $G_{\mu\nu} = -\kappa T_{\mu\nu}$, where $G_{\mu\nu}$ denotes the Ricci tensor and $T_{\mu\nu}$ denotes the energy-momentum tensor. These are not the field equations of the final theory, which contain a trace term on either the left or the right-hand side. However, since in vacuum the older field equations coincide with those of the final theory, Einstein and Schwarzschild's solutions still hold there.

tion. However, while in the case of the spherically symmetric vacuum field this procedure led to a simple result, in this case the differential equations become rather involved and no solution is obtained. The calculations on the page discussed end with a second order differential equation coupling f and f_1 (see the equation at the bottom right-hand side of Schwarzschild's manuscript page reproduced as Fig. 6 in the Appendix).

What is the purpose of Schwarzschild's notes? Clearly these notes are concerned with the problem of the relativity of rotation in general relativity. As was explained in sec. 3, Schwarzschild had, in his earlier work, considered the possibility of inertial reference frames rotating with respect to one another. The spacetime Schwarzschild considers here is a realization of such an arrangement in the framework of general relativity, one inertial system being given at radial infinity, the other within the rotating cylindrical section of spacetime. The difficulty in making any further-reaching statements about the physical situation Schwarzschild is trying to describe here arises from the fact that, in his notes, Schwarzschild does not specify the matter distribution he assumes.[37] It is, however, plausible to assume that the spacetime under consideration should function as a model for the spatially two-dimensional situation of a rotating disk of Minkowski space. Then Schwarzschild's calculations can be interpreted as the exploration of a simple model for the spacetime inside and outside the rotating system of fixed stars.[38] This would mean that Schwarzschild assumed the spacetime within the system to be approximately Minkowskian, an assumption that is consistent with his earlier observation that the perihelia of remote double stars do not rotate relative to the directions defined by the planetary orbits in the solar system and that therefore inertial frames inside the Galaxy do not rotate with respect to one another. At the same time, Schwarzschild must have assumed that in its rotational motion the matter content of the system of fixed stars—be it concentrated in a ring or distributed over an ellipsoidal volume—drags along the interior Minkowski space.[39]

37 A spacetime similar to the one Schwarzschild describes here is generated by an infinitely long, rotating, cylindrical shell of matter. The interior field of such a matter distribution is indeed Minkowskian; see (Davies et al. 1971). The exterior vacuum metric of such a matter distribution, including the dragging of inertial frames close to the shell, is discussed in (Frehland 1972). Non-local effects of such a matter distribution, corresponding to the Aharonov-Bohm effect in electrodynamics, are discussed in (Stachel 1983). The problem of the relativity of rotation was addressed in 1918 by Thirring who considered a rotating spherical mass shell rather than a cylindrical one (Thirring 1918, see also Lense and Thirring 1918).

38 In an earlier publication (Schwarzschild 1909, 41–42), Schwarzschild had described the solar system whose dynamics is dominated by a central mass as being of "monarchic constitution", and the Galaxy where every star is acted upon by all other stars as being of "republican constitution", and had speculated that the whole Universe might be built up from a hierarchical sequence of structures of these two basic types. While Schwarzschild's first exact solution to Einstein's field equations provides the basis for describing the monarchic constitution in the framework of general relativity, the notes considered here can be understood as an attempt to complement this with the description of the republican constitution within that framework.

39 The spacetime of an axially symmetric distribution of particles revolving with constant angular velocity was later derived by (Stockum 1937).

In fact, Schwarzschild could have hoped that, within general relativity, a problem concerning the rotation of the system of fixed stars could be resolved. On one hand, namely, Schwarzschild contended that, by analogy to other celestial motions, it must be assumed that the system of fixed stars as a whole rotates.[40] Yet, on the other hand, such a rotation had hardly been observed, as Schwarzschild explained in an earlier text:

> [...] it turns out that the average of those few thousand stars, whose proper motions are known, displays no evidence of rotation with respect to [the] directions [defined by the planetary orbits] [...].[41]

General relativity now provided a possible explanation of this phenomenon, if it was assumed that, together with the stars themselves, the global inertial system within the Galaxy was rotating. In searching for the functions f and f_i describing the outer metric, Schwarzschild would then have attempted to clarify in what sense one may speak in general relativity of a rotation of the system of fixed stars as a whole.

That Schwarzschild indeed considered the question of the rotation of the Galaxy in the context of general relativity is made evident by a letter from Einstein dated January 9, 1916.[42] In a preceding letter by Schwarzschild which is lost, Schwarzschild must have raised several questions, which Einstein answers one by one. Einstein's second point reads as follows:

> The statement that "the system of fixed stars" is free of rotation may retain a relative meaning, which is to be fixed by a comparison.
>
> The surface of the Earth is irregular, as long as I regard very small sections of it. However, it approaches the flat elementary shape when I regard larger sections of it, whose dimensions are still small in comparison to the length of the meridian. This elementary shape becomes a curved surface when I regard even larger sections of the Earth's surface.
>
> For the gravitational field things are similar. On a small scale the individual masses produce gravitational fields that, even with the most simplifying choice of the reference system, reflect the character of the quite irregular matter distribution on the small scale. If I consider larger regions, as astronomy presents them to us, the Galilean reference system provides me with the analogue to the flat elementary shape of the Earth's surface in the previous comparison. But if I consider even larger regions, there probably will be no continuation of the Galilean system to simplify the description of the universe to the same degree as on a small scale, that is, throughout which a mass point sufficiently remote from other masses moves uniformly in a straight line.[43]

Schwarzschild's response is consistent with the calculations as interpreted above:

> As concerns the inertial system, we are in agreement. You say that beyond the Milky Way system conditions may arise under which the Galilean system is no longer the simplest. I only hold that within the Milky Way system such conditions do not arise.[44]

---

40 See, for example, (Schwarzschild 1897, 519).

41 "[...] zeigt sich, dass der Durchschnitt aus jenen paar Tausend Sternen, deren Eigenbewegung man kennt, [...] keine Rotation gegen diese Richtungen aufweist." (Schwarzschild 1897, 520.)

42 Einstein to Schwarzschild, January 9, 1916, N 193, 3-5, see also (CPAE 8, Doc. 181).

In view of this exchange between Schwarzschild and Einstein, it appears obvious that Schwarzschild's calculations are related to the question he must have posed to Einstein: Does it make sense to speak of a rotation of the system of fixed stars? The calculations then document the attempt to explore the "even larger regions" outside the system of fixed stars, in which "there probably will be no continuation of the Galilean system."

7. A CLOSED UNIVERSE AS A SOLUTION OF EINSTEIN'S FIELD EQUATIONS

In his calculations, Schwarzschild had assumed the Universe to be asymptotically Minkowskian. In his correspondence with Einstein on the question of global frames of inertia, Schwarzschild mentions a further possibility. In direct continuation of the passage quoted above, he explicates:

> As concerns very large spaces, your theory has a quite similar position as Riemann's geometry, and you are certainly not unaware that one obtains an elliptic geometry from your theory, if one puts the entire universe under uniform pressure (energy tensor $-p, -p, -p, 0$).[45]

Thus, Schwarzschild was the first to entertain the possibility of a closed universe with an elliptic geometry as a solution to Einstein's field equations. Schwarzschild's remark that Einstein's theory had a similar position as Riemann's geometry thereby alludes to his prerelativistic application of elliptic geometry to the universe on the background of Riemannian geometry discussed in sec. 4.

Contrary to Schwarzschild's assumption, Einstein was, at the time, most probably unaware of such cosmological implications of his theory.[46] It was only through a debate with the Dutch astronomer Willem de Sitter (1872–1934) beginning in fall

---

43 "Die Aussage, dass "das Fixsternsystem" rotationsfrei sei, behält wohl einen relativen Sinn, der durch ein Gleichnis festgelegt sei.
Die Oberfläche der Erde ist, solange ich ganz kleine Teile derselben ins Auge fasse, unregelmässig. Sie nähert sich aber der ebenen Grundgestalt, wenn ich grössere Teile ins Auge fasse, deren Abmessungen aber immer noch klein sind gegen die Länge des Meridians. Diese Grundgestalt wird zu einer gekrümmten Fläche, wenn ich noch grössere Teile der Erdoberfläche ins Auge fasse.
So ähnlich ist es auch mit dem Gravitationsfeld. Im Kleinen liefern die einzelnen Massen Gravitationsfelder, welche auch bei möglichst vereinfachender Wahl des Bezugssystems den Charakter der ziemlich regellosen Verteilung der Materie im Kleinen widerspiegeln. Betrachte ich grössere Gebiete, wie sie uns die Astronomie bietet, so bietet mir das Galileische Bezugssystem das Analoge zu der ebenen Grundgestalt der Erdoberfläche beim vorigen Vergleich. Betrachte ich aber noch grössere Gebiete, so wird es wohl keine Fortsetzung des Galileischen Systems geben, welche in solchem Masse wie im Kleinen die Beschreibung der Welt einfach gestaltet d.h. in welchem überall der von anderen Massen hinlänglich entfernte Massenpunkt sich gradlinig gleichförmig bewegt." Einstein to Schwarzschild, January 9, 1916 (CPAE 8, Doc. 181).

44 "Was das Inertialsystem angeht, so sind wir einig. Sie sagen, daß jenseits des Milchstraßensystems sich Verhältnisse einstellen können, in denen das Galilei'sche System nicht mehr das einfachste ist. Ich behaupte nur, daß sich innerhalb des Milchstraßensystems solche Verhältnisse nicht einstellen." Schwarzschild to Einstein, February 6, 1916, N 193, 7–8, see also (CPAE 8, Doc. 188).

1916 that Einstein was led to consider a closed universe which he hesitantly proposed in 1917 (Einstein 1917).[47] The distinction between spherical and elliptic space had thereby remained obscure to him. De Sitter pointed the distinction out to Einstein, referring to Schwarzschild's 1900 paper on the curvature of space and the argument for preferring elliptic to spherical space given therein.[48]

In Einstein's debate with de Sitter, the question of the global geometry of the universe emerged from a discussion of the relativity of inertia. Strikingly, in Schwarzschild's correspondence with Einstein, the question of the global geometry of the Universe is brought up in exactly the same context.[49] It is therefore tempting to see here the commencement of an Einstein–Schwarzschild debate foreshadowing the later Einstein–de Sitter debate. Einstein appears, however, to have not yet been prepared to consider the cosmological implications of his theory at that time. And by the time he was slowly pushed into that direction in his exchange with the astronomer de Sitter, Schwarzschild had already died. Nevertheless, in view of Schwarzschild's deliberations discussed here, it seems safe to say that, had Schwarzschild lived longer, he could have made a substantial contribution to the cosmological debates emerging later.

8. SCHWARZSCHILD'S INTERDISCIPLINARY APPROACH TO THE FOUNDATIONS OF SCIENCE

Let us come back to the question raised at the beginning: Why did Schwarzschild recognize the significance of general relativity at such an early stage? Here it has been attempted to show that, already in his early astronomical work, Schwarzschild did not act as a specialist but attempted to meet the challenges resulting from the implica-

45 "Was die ganz großen Räume angeht, hat Ihre Theorie eine ganz ähnliche Stellung, wie Riemann's Geometrie, und es ist Ihnen gewiß nicht unbekannt, daß man die elliptische Geometrie aus Ihrer Theorie herausbekommt, wenn man die ganze Welt unter einem gleichförmigen Druck stehen läßt (Energietensor $-p, -p, -p, 0$). " Schwarzschild to Einstein, February 6, 1916, N 193, 7–8, see also (CPAE 8, Doc. 188). This energy tensor actually does not yield a spherical static universe. It does, however, yield the universe Schwarzschild is speaking of when the trace term in the field equations is neglected, i.e. in the context of the older field equations $G_{\mu\nu} = -\kappa T_{\mu\nu}$, where $G_{\mu\nu}$ denotes the Ricci tensor and $T_{\mu\nu}$ denotes the energy-momentum tensor (see footnote 36). It may be the case that Schwarzschild originally conceived of the tensor on the basis of these field equations and later did not modify it as the new field equations would have demanded. Schwarzschild continues his letter by explaining the solution inside a sphere of fluid with uniform energy density (energy tensor $-p, -p, -p, \rho_0$). Here, as well as in the corresponding publication (Schwarzschild 1916b, 431–432), Schwarzschild points out that inside the sphere spherical geometry applies.

46 In November 1916, Einstein still calls the question of the boundary conditions of the metric field "purely a matter of taste which will never attain a scientific meaning." ("eine reine Geschmacksfrage, die nie eine naturwissenschaftliche Bedeutung erlangen wird.") Einstein to de Sitter, November 4, 1916 (CPAE 8, Doc. 273).

47 On the Einstein–de Sitter debate, see (CPAE 8, 351–357) and the references given therein, in particular (Kerszberg 1989; 1989a).

48 De Sitter to Einstein, June 20, 1917 (CPAE 8, Doc. 355).

tions of foundational questions in physics on astronomy. Thus it comes as no surprise that Schwarzschild was also among the first to recognize that Einstein had—without being aware of it—provided the astronomers with the adequate framework for treating their questions. As a result, a clear continuity can be perceived in Schwarzschild's work on cosmology, prerelativistic and relativistic. In this context it is interesting to question the extent to which parallel cases are provided by the work of other pioneers of relativistic astronomy such as Willem de Sitter and Arthur Eddington (1882–1944). The study of this question, however, does not lie in the scope of this contribution. Here, in conclusion, it shall only be pointed out that it was no coincidence that Schwarzschild took an astronomical road to general relativity, but that this may rather be seen as the natural outcome of his interdisciplinary approach to the foundations of the exact sciences.

Indeed, not only is interdisciplinarity the hallmark of Schwarzschild's scientific work, but he also was quite aware of the general significance of interdisciplinarity for the progress of science. On many occasions Schwarzschild explained how he saw scientific progress emerging from the interplay of the different branches of science, for instance when, on the occasion of his inaugural lecture at the Prussian Academy of Sciences, he stated that "the greatest yet unsolved problem of celestial mechanics, the so-called many-body problem, most closely touches a problem of physics that concerns the foundations of its newest developments".[50] As a further example, consider the following passage from the same speech in which Schwarzschild describes the establishment of special relativity:

> [...] an important source for the electron and relativity theory lay in an astronomical problem. The astronomical aberration results from the finite propagation speed of light through the aether in combination with the Earth's motion in space. H.A. Lorentz occupied himself many times with the theory of aberration and searched for a satisfying picture of the aether's behavior when large masses, like the Earth, move through it, until he

49 In Einstein's letter to Schwarzschild from January 9, there is, in fact, a passage in which he expresses exactly the kind of strong Machian claims concerning his theory which later sparked off his debate with de Sitter. In direct continuation of the passage quoted above, Einstein explains: "According to my theory, inertia is an interaction between masses, in the end, not an effect in which, besides the mass under consideration, 'space' itself would be involved. The essence of my theory is precisely that no independent properties are attributed to space itself.
Jokingly one may put it this way. If I let all things in the world disappear, according to Newton the Galilean inertial space remains, according to my perception, however, *nothing* remains."
("Die Trägheit ist eben nach meiner Theorie im letzten Grunde eine Wechselwirkung der Massen, nicht eine Wirkung bei welcher ausser der ins Auge gefassten Masse der 'Raum' als solcher beteiligt ist. Das Wesentliche meiner Theorie ist gerade, dass dem Raum als solchem keine selbständigen Eigenschaften gegeben werden.
Man kann es scherzhaft so ausdrücken. Wenn ich alle Dinge aus der Welt verschwinden lasse, so bleibt nach Newton der Galileische Trägheitsraum, nach meiner Auffassung aber *nichts* übrig.") Einstein to Schwarzschild, January 9, 1916 (CPAE 8, Doc. 181).

50 "[...] berührt sich das höchste noch ungelöste Problem der Himmelsmechanik, das sogenannte Vielkörperproblem, aufs engste mit einem Problem der Physik, das an die Fundamente ihrer neuesten Entwicklung greift." (Schwarzschild 1913, 597.)

> finally cut the knot by consistently implementing Fresnel's assumption that the aether is absolutely rigid and cannot be brought to flow by any force acting on it. In this way the path was cleared for the electron theory. Furthermore, the completely rigid aether stepped out of the circle of the objects that can be influenced and thus can be more closely perceived, so much so that relativity theory became possible, in which the concept of the aether only appears as a spacetime concept deepened by new experience.
>
> Electron theory and relativity theory in turn have already posed various problems to astronomy as a consequence of the modifications of celestial mechanics they necessitate.[51]

Clearly Schwarzschild was equipped to take up these challenges posed to astronomy by its neighboring disciplines. He knew that scientific progress is not a matter of the advancement of isolated disciplines, as both the previous and the following quotations make clear:

> Mathematics, physics, chemistry, astronomy march in line. Whichever lags behind is pulled forward. Whichever hastens ahead pulls the others forward. The closest solidarity exists between astronomy and the whole circle of exact sciences.[52]

ACKNOWLEDGEMENTS

This paper was previously published in the journal *Science in Context* (Schemmel 2005). I am indebted to Jürgen Renn who inspired and supported my work from its inception up to its present state. For helpful discussions and comments I would also like to thank Dieter Brill, Giuseppe Castagnetti, Peter Damerow, Jürgen Ehlers, Hubert Goenner, Michel Janssen, John Norton, and John Stachel.

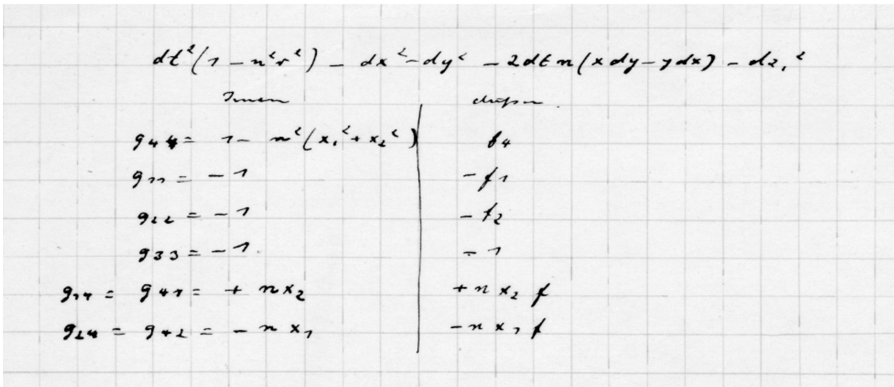
---

51 "[...] eine wichtige Quelle für die Elektronen- und Relativitätstheorie in einem astronomischen Probleme lag. Die astronomische Aberration ist eine Folge der endlichen Ausbreitungsgeschwindigkeit des Lichtes im Äther verbunden mit der Bewegung der Erde im Weltraum. H.A. Lorentz hat sich vielfach mit dem Problem der Aberration beschäftigt und nach einer befriedigenden Anschauung über das Verhalten des Äthers, wenn große Massen, wie die Erde, sich durch ihn hindurchbewegen, gesucht, bis er schließlich den Knoten zerhieb durch völlig konsequente Durchführung der alten Fresnelschen Annahme, daß der Äther absolut starr und durch keine auf ihn wirkende Kraft zum Fließen zu bringen sei. Dadurch war die Bahn frei geworden für die Elektronentheorie. Der völlig starre Äther trat ferner so sehr aus dem Kreis der beeinflußbaren und damit näher erkennbaren Objekte heraus, daß auch die Relativitätstheorie möglich wurde, bei welcher der Begriff des Äthers nur als ein durch neue Erfahrungen vertiefter Raum-Zeitbegriff erscheint.
Elektronentheorie und Relativitätstheorie haben auch Rückwärts der Astronomie schon wieder mancherlei Probleme gestellt infolge der Modifikationen der Himmelsmechanik, die sie notwendig machen." (Schwarzschild 1913, 598.)

52 "Mathematik, Physik, Chemie, Astronomie marschieren in einer Front. Wer zurückbleibt, wird nachgezogen. Wer vorauseilt, zieht die anderen nach. Es besteht die engste Solidarität der Astronomie mit dem ganzen Kreis der exakten Naturwissenschaften." (Schwarzschild 1913, 599.)

APPENDIX: ANNOTATED REPRODUCTION OF SCHWARZSCHILD'S RELATIVISTIC NOTES ON THE PROBLEM OF ROTATION

In this appendix, the manuscript page that documents Schwarzschild's relativistic calculations on the problem of rotation referred to in sec. 6 is reproduced with annotations. It is preserved as a part of Schwarzschild's Nachlass in the Niedersächsische Staats- und Universitätsbibliothek Göttingen as page 12r of folder 2:2. The page is found among a few similar pages, some of which contain further notes on general relativity. For technical reasons this reproduction is divided into three parts shown in figs. 4, 5, and 6, respectively. I am grateful to the Niedersächsische Staats- und Universitätsbibliothek Göttingen for their permission to reproduce this page.


$dt^2(1-n^2r^2) - dx^2 - dy^2 - 2dt\,n(x\,dy - y\,dx) - dz^2$

Innen	Außen
$g_{44} = 1 - n^2(x_1^2 + x_2^2)$	f_4
$g_{11} = -1$	$-f_1$
$g_{22} = -1$	$-f_2$
$g_{33} = -1$	-1
$g_{14} = g_{41} = +nx_2$	$+nx_2 f$
$g_{24} = g_{42} = -nx_1$	$-nx_1 f$

Figure 4: Schwarzschild N 2:2, 12r, upper part

Schwarzschild begins his considerations by writing down the metric for the "inner" and the "outer" spacetime in Cartesian coordinates,

$$g_{\mu\nu_{\text{inner}}} = \begin{bmatrix} -1 & 0 & 0 & nx_2 \\ 0 & -1 & 0 & -nx_1 \\ 0 & 0 & -1 & 0 \\ nx_2 & -nx_1 & 0 & 1-n^2(x_1^2+x_2^2) \end{bmatrix}; \quad g_{\mu\nu_{\text{outer}}} = \begin{bmatrix} -f_1 & 0 & 0 & nx_2 f \\ 0 & -f_2 & 0 & -nx_1 f \\ 0 & 0 & -1 & 0 \\ nx_2 f & -nx_1 f & 0 & f_4 \end{bmatrix}.$$

In the following calculation (figs. 5 and 6), he then shifts to cylindrical coordinates and absorbs the factors of n in the function f. The outer metric then reads

$$g_{\mu\nu_{\text{outer}}} = \begin{bmatrix} -f_1 & 0 & 0 & 0 \\ 0 & -f_2 & 0 & f \\ 0 & 0 & -1 & 0 \\ 0 & f & 0 & f_4 \end{bmatrix}.$$

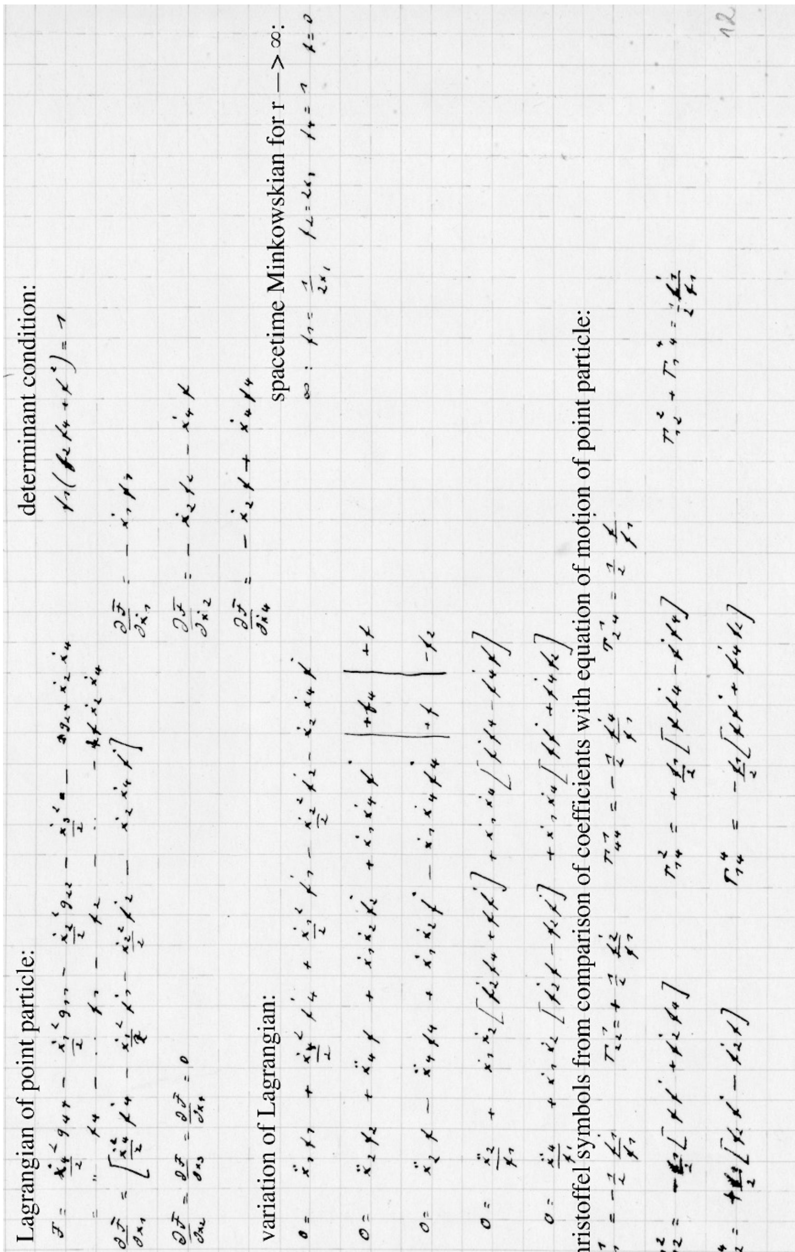


Figure 5: Schwarzschild N 2:2, 12r, middle part, with annotations

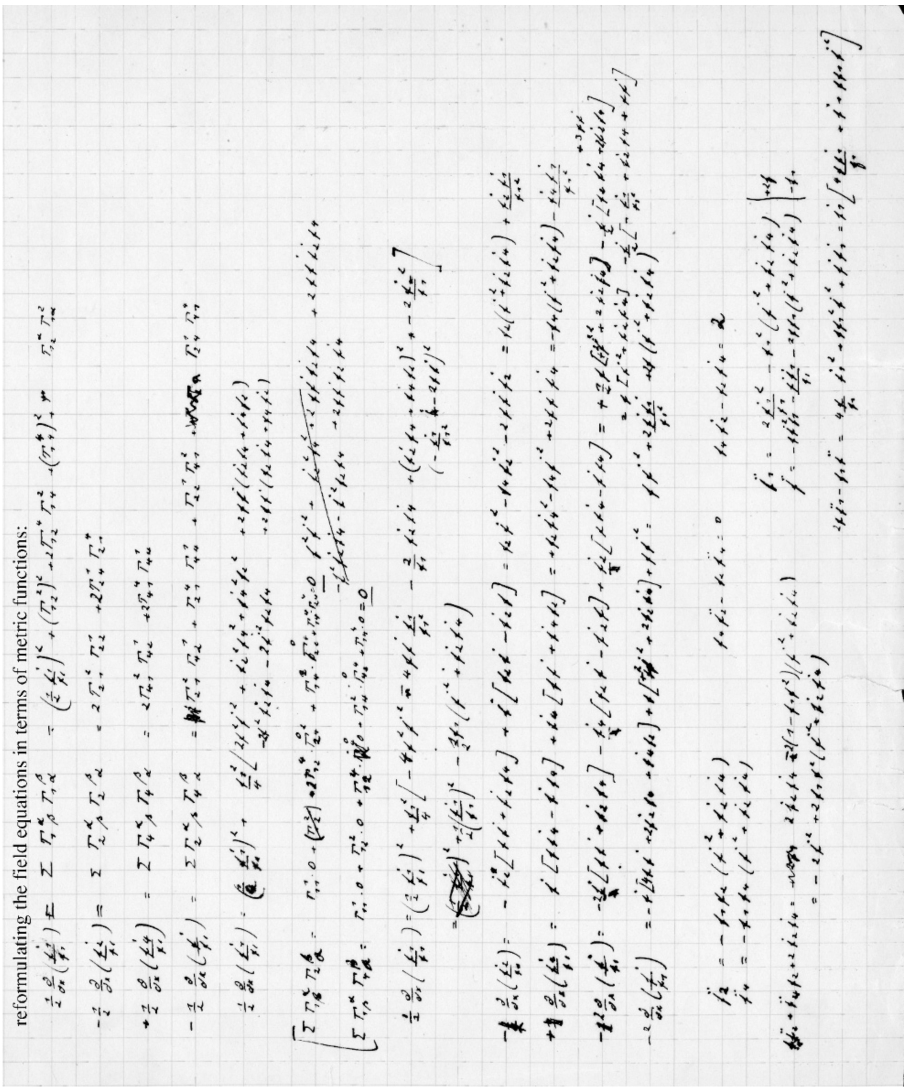


Figure 6: Schwarzschild N 2:2, 12r, lower part, with an annotation

REFERENCES

References to Karl Schwarzschild's Nachlass in the Niedersächsische Staats- und Universitätsbibliothek Göttingen are given by an archival number following an 'N' (e.g. N 2:2).

Abraham, Max. 1912. "Zur Theorie der Gravitation." *Physikalische Zeitschrift* 13: 1–4.(English translation in this volume.)

Blumenthal, Otto. 1918. "Karl Schwarzschild." *Jahresberichte der Deutschen Mathematiker-Vereinigung* 26: 56–75.

CPAE 3. 1993. Martin J. Klein, A. J. Kox, Jürgen Renn, and Robert Schulmann (eds.), *The Collected Papers of Albert Einstein*. Vol. 3. *The Swiss Years: Writings, 1909–1911*. Princeton: Princeton University Press.

CPAE 6. 1996. A. J. Kox, Martin J. Klein, and Robert Schulmann (eds.), *The Collected Papers of Albert Einstein*. Vol. 6. *The Berlin Years: Writings, 1914–1917*. Princeton: Princeton University Press.

CPAE 8. 1998. Robert Schulmann, A. J. Kox, Michel Janssen, and József Illy (eds.), *The Collected Papers of Albert Einstein*. Vol. 8. *The Berlin Years: Correspondence, 1914–1918*. Princeton: Princeton University Press.

Davies, H. and T. A. Caplan. 1971. "The Space-Time Metric inside a Rotating Cylinder." *Proceedings of the Cambridge Philosophical Society* 69: 325–327.

Earman, John and Michel Janssen. 1993. "Einstein's Explanation of the Motion of Mercury's Perihelion." In John Earman, Michel Janssen, and John D. Norton (eds.) *The Attraction of Gravitation: New Studies in the History of General Relativity* (*Einstein Studies,* vol. 5), 129–172. Boston: Birkhäuser.

Einstein, Albert. 1911. "Über den Einfluß der Schwerkraft auf die Ausbreitung des Lichtes." *Annalen der Physik* 35: 898–908, (CPAE 3, Doc. 23).

———. 1915a. "Zur allgemeinen Relativitätstheorie (Nachtrag)." *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften* 1915: 799–801, (CPAE 6, Doc. 22).

———. 1915b. "Erklärung der Perihelbewegung des Merkur aus der allgemeinen Relativitätstheorie." *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften* 1915: 831–839, (CPAE 6, Doc. 24).

———. 1916. "Die Grundlage der allgemeinen Relativitätstheorie." *Annalen der Physik* 49: 769–822, (CPAE 6, Doc. 30).

———. 1917. "Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie." *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften* 1917: 142–152, (CPAE 6, Doc. 43).

Eisenstaedt, Jean. 1982. "Histoire et singularités de la solution de Schwarzschild (1915-1923)." *Archive for History of Exact Sciences* 27: 157–198.

———. 1987. "Trajectoires et impasses de la solution de Schwarzschild." *Archive for History of Exact Sciences* 37: 275–357.

———. 1989. "The Early Interpretation of the Schwarzschild Solution." In Don Howard and John Stachel (eds.), *Einstein and the History of General Relativity.* (*Einstein Studies* vol. 1.) Boston/Basel/Berlin: Birkhäuser, 213–233.

Frehland, Eckart. 1972. "Exact gravitational field of the infinitely long rotating hollow cylinder." *Communications in mathematical physics* 26: 307–320.

Hentschel, Klaus. 1993. "The Conversion of St. John. A Case Study on the Interplay of Theory and Experiment." In Mara Beller, Robert S. Cohen, and Jürgen Renn (eds.), *Einstein in Context*. Cambridge: Cambridge University Press, 137–194.

———. 1997. *The Einstein Tower. An Intertexture of Dynamic Construction, Relativity Theory, and Astronomy.* Stanford: Stanford University Press.

Israel, Werner. 1987. "Dark Stars: the evolution of an idea." In Stephen Hawking and Werner Israel (eds.), *Three Hundred Years of Gravitation*. Cambridge: Cambridge University Press, 199–276.

Kerszberg, Pierre. 1989. "The Einstein–de Sitter Controversy of 1916–1917 and the Rise of Relativistic Cosmology." In Don Howard and John Stachel (eds.). *Einstein and the History of General Relativity.* (*Einstein Studies* vol. 1.) Boston/Basel/Berlin: Birkhäuser, 325–366.

———. 1989a. *The Invented Universe. The Einstein–De Sitter Controversy (1916–17) and the Rise of Relativistic Cosmology.* Oxford: Clarendon Press.

Laue, Max von. 1917. "Die Nordströmsche Gravitationstheorie." *Jahrbuch der Radioaktivität und Elektronik* 14: 263–313.

Lense, Joseph, and Hans Thirring. 1918. "Über den Einfluß der Eigenrotation der Zentralkörper auf die Bewegung der Planeten und Monde nach der Einsteinschen Gravitationstheorie." *Physikalische Zeitschrift* 19: 156–163.

Norton, John. 1999. "The Cosmological Woes of Newtonian Gravitation Theory." In Hubert Goenner, Jürgen Renn, Jim Ritter, and Tilman Sauer (eds.), *The Expanding Worlds of General Relativity*. (*Einstein Studies* vol. 7.) Boston/Basel/Berlin: Birkhäuser, 271–323.

Poincaré, Henri. 1902. *La Science et l'hypothèse*. Paris: E. Flammarion.

Schemmel, Matthias. 2005. "An Astronomical Road to General Relativity: The Continuity between Classical and Relativistic Cosmology in the Work of Karl Schwarzschild." *Science in Context* 18:451–478.

Schwarzschild, Karl. 1897. "Was in der Welt ruht." *Die Zeit*. Vol. 11, No. 142, 19 June 1897, Vienna, 181–183. (English translation in this volume.) [The page numbers given in the text refer to (Schwarzschild 1992) wherein this article is reprinted.]

———. 1900. "Über das zulässige Krümmungsmaass des Raumes." *Vierteljahrsschrift der Astronomischen Gesellschaft* 35: 337–347.

———. 1909. *Über das System der Fixsterne*: *Aus populären Vorträgen*. *Naturwissenschaftliche Vorträge und Schriften No.1*. Leipzig: Teubner, 39–43.

———. 1913. "Antrittsrede des Hrn. Schwarzschild." *Sitzungsberichte der königlich preussischen Akademie der Wissenschaften* 1913: 596-600.

———. 1914. "Über die Verschiebung der Bande bei 3883 Å im Sonnenspektrum." *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften* 1914: 1201–1213.

———. 1916a. "Über das Gravitationsfeld eines Massenpunktes nach der Einsteinschen Theorie." *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften* 1916: 189–196.

———. 1916b. "Über das Gravitationsfeld einer Kugel aus inkompressibler Flüssigkeit nach der Einsteinschen Theorie." *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften* 1916: 424–434.

———. 1992. *Gesammelte Werke*, Vol. 3, Hans H. Voigt (ed.). Berlin: Springer.

Seeliger, Hugo von. 1913. "Bemerkungen über die sogenannte absolute Bewegung, Raum und Zeit." *Vierteljahrsschrift der Astronomischen Gesellschaft* 48: 195–201.

Sommerfeld, Arnold. 2000–2004.*Wissenschaftlicher Briefwechsel* (2 vols.), Michael Eckert and Karl Märker (eds.). Berlin: Verlag für Geschichte der Naturwissenschaften und der Technik.

St. John, Charles E. 1910. "The General Circulation of the Mean- and High-Level Calcium Vapour in the Solar Atmosphere." *Astrophysical Journal* 32: 36–82.

———. 1910–11. "Motion and Condition of Calcium Vapour over Sun-Spots and Other Special Regions." *Astrophysical Journal* 34: 57–78, 131–153.

Stachel, John. 1983. "The gravitational fields of some rotating and nonrotating cylindrical shells of matter." *Journal of Mathematical Physics* 25: 338–341.

Stockum, W. J. van. 1937. "The Gravitational Field of a Distribution of Particles Rotating about an Axis of Symmetry." *Proceedings of the Royal Society of Edinburgh* 57: 135–154.

Thirring, Hans. 1918. "Über die Wirkung rotierender ferner Massen in der Einsteinschen Gravitationstheorie." *Physikalische Zeitschrift* 19: 33–39.

KARL SCHWARZSCHILD

THINGS AT REST IN THE UNIVERSE

Originally published as "Was in der Welt ruht" in Die Zeit, Vienna, Vol. 11, No. 142, 19 June 1897, 181–183 (1897). Reprinted in Collected Works, vol. 3, H. H. Voigt, ed., Springer-Verlag, Berlin 1992, p. 514–521. Page numbers refer to reprinted version.

Is there anything in the universe that is at rest, around which or within which the rest of the universe is constructed, or is there no hold in the unending chain of motions in which everything seems to be caught up? It is worth considering the extent to which these questions are justified and how they can be answered.

On a clear evening a few weeks ago, many people claimed to have observed how the Moon rushed across the sky with most unusual speed. On that evening a light veil of mist seems to have been blown past the Moon by the wind, more subtly creating the same illusion one believes one observes every time broken clouds move quickly past it: one assumes that the clouds stand still and that the Moon rushes through them against the wind.

This phenomenon and other much more common ones, such as the apparent rotation of long furrows when one travels through a flat landscape by train, or the enchanted room at recent fairs which leads one to believe that one is upside down and walking on the ceiling, simply provide evidence to support the immediately comprehensible statement that all perceptible motion is relative, that one can say that an object moves relative to another one but never in absolute terms that an object moves or is at rest.

With this one can already see that neither of the original two questions has been properly formulated. Namely, if only the motion of different objects relative to each other can be perceived, then without actually contradicting experience one may ascribe a completely arbitrary motion to a particular object in the universe or, as a special case, declare it to be at rest. Thus, everything and nothing is at rest in the universe. Logically, there is nothing to stop someone from stipulating, for example, that the wing tips of a buzzing insect are still. In that case one would simply have to ascribe a buzzing motion in the opposite direction to one's own body, the Earth, and all celestial objects, in order to recover the directly observable relative motion of the insect's wings with respect to the rest of the universe.

Why, however, does one smile at he who would seriously wish to claim that the Moon really did rush across the sky, or that the fields rotated when one went through

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 3
Gravitation in the Twilight of Classical Physics: Between Mechanics, Field Theory, and Astronomy.

them on the train? The reasons are of a purely practical and utilitarian nature. A clumsy assumption regarding things at rest in the universe would result in an unspeakable confusion of ideas. And for this reason, the above questions are to be replaced by the following: What things in the universe did one find it useful to treat as being at rest, at different times? Each time it becomes necessary to redefine what can be treated as being at rest, an important stage is passed in the process of development of human ideas concerning the universe. It has often been forgotten, or not even rec-
[515] ognized, that here we are | dealing with arbitrary assumptions, and at times the opinions have become doctrines which have even found their own martyrs.

When one talks of motion or rest in everyday life, one leaves out the object with respect to which motion or rest is ascribed, but one means: relative to the ground or to objects fixed with respect to the ground. It is useful to tacitly assume this particular completion of the concept of motion because most of what appears before our eyes during the course of the day is at rest relative to the ground, or at least we can name a special cause when something moves relative to the ground. However, since Copernicus science has begun to recognize other completions as useful, to ascribe motion to other objects, and to treat other things as "points at rest" in the universe.

Indeed, Copernicus suggested that all motion is relative to the center of the Sun and showed that all the complicated curves and ribbons traced by the planets in the sky found an explanation if one allows the planets to describe circles around the Sun, which is taken to be fixed.

But more precisely, Copernicus makes a further assumption, in that he stipulates that the fixed stars, located at a huge distance from our planetary system, should be considered to be at rest. In fact, the single assumption that the center of the Sun is fixed would not suffice as a definition of all motion. One could then recognize only changes in distance relative to this point but not rotations about it, just as one does not know whether a completely smooth ball, whose surface is the same all over, is stationary or rotating about its center.

Accordingly, in the Copernican system, one describes the motion of a body by giving its distance from the Sun, and direction in terms of a fixed star as seen from the Sun, at various times. The assumption that the fixed stars are at rest thereby forces one to attribute a daily rotation from west to east to the Earth, because relative to the Earth's surface the fixed stars apparently revolve from east to west within a 24 hour period. Consequently one must imagine that together with our surroundings we rush through space with the speed of a cannon ball, and that the direction of this motion continually changes. It cannot be said that at first such a way of thinking seems to be particularly in accordance with the principle of utility, which we must recognize as the single decisive principle for these questions. From this viewpoint, the reluctance of the Aristotelians at the time certainly does not seem to be as naive and ridiculous as it is generally made out to be. All the more worthy of admiration is a man such as Galileo, who, through his mechanical principles, made the adventurous elements of the new view as well as the presumed contradictions with everyday experience disappear, so that only its overwhelming advantages in the depiction of celestial motions remained. |

It was not until the end of the seventeenth century—when the use of the telescope and progress in the production of finely divided circles, used to read off the position of stars, had led to an unexpected improvement in observation skills—that the Copernican stipulation of motion was recognized as not being wholly exact and unambiguous. At that time it was noticed that the fixed stars do not actually deserve their name: they all shift in position relative to each other, and in the very distant past Cygnus and Orion once formed quite different constellations. However, these shifts in position progress extremely slowly. If Stephan's Tower would list so far in the course of one year that its top shifted by one centimeter, then this shift, as observed from Kahlenberg, would be conspicuous compared to the shift in position of the majority of the fixed stars visible to the naked eye.[1] But as soon as the shifts are at all perceptible, Copernicus' use of the fixed star system as a basis for motion is no longer possible. [516]

From those same observations concerning the change in position of the fixed stars relative to each other it has also emerged that the other Copernican point of reference, the center of the Sun, possesses no particular right to the privilege of being considered at rest. The "proper motions" of the fixed stars, such being the technical term, are indeed distributed irregularly so that neighboring stars in the sky move away in all possible directions, but on average out of several hundred stars one can see that when one looks into the sky in a certain direction, on the whole the constellations seem to spread, to get bigger, and in the opposite direction, to get smaller. From this it follows that the Sun and the former region of the sky come closer to each other. Although it would not be impossible to do, there would be no point in ascribing this motion wholly to the fixed stars and continuing to treat the Sun as being at rest, for we know that the Sun is a relatively subordinate member of the great family of stars.

Just realize into what insecurity the universe has fallen as a result, how imagination finds no place to drop anchor and no single rock in the world has a special right to be thought of as fixed and at rest.

Unfortunately we have still not reached the epoch of astronomy that will intervene offering a new fundamental definition. This will have come about when a pattern has been found in the seemingly so irregular proper motions. No effort is being spared in order to promote its appearance. Twenty observatories have joined together in order to produce a catalogue which details the position of 150,000 fixed stars relative to each other at the present time, and a major part of this huge task has already been completed. Just as many institutes have divided up the sky amongst themselves in order to establish the position of two million stars with even greater precision through photographic images. When this work is repeated | in a few decades time, [517]
such a huge number of proper motions will be known to us that a law, if it exists, must reveal itself, and that a new Copernicus, who admittedly would have no prejudice to overcome, can show from where and how these motions appear ordered. Admittedly, it could also turn out that no general order exists within the army of fixed stars. In this case, all these spheres in space are to be compared with gas molecules, which fly around completely irregularly, so irregularly that the irregularity itself

becomes a principle, according to which the effect of the gas mass as a whole can be derived through considerations of probabilities and averages.

But, for the time being, to what does the astronomer refer the heavenly motions?

Among the pieces of ponderable matter there is none, as we have seen, which can be reasonably distinguished as stationary. Thus among material things one's only chance is to look around for imponderables. As is well known, the universe is filled with an all-pervasive substance, which is weightless, neither solid nor fluid, neither visible nor invisible, and to which there is no physical description that really fits: this is the aether, the only imponderable of modern physics. Through the aether do the gentlest light waves shimmer, but the aether also mediates the mighty effects of electrical machinery. When the power of a distant waterfall is transferred to a central power station, then in a certain sense aether is the long rod impacted by the water over there and pushing the wheel of the machine here. The electrical cable is of only secondary importance; it simply maintains the energy flowing in the desired direction, rather than dispersing. Now, it would be very tempting from a philosophical standpoint to take the ever-present aether to be the basic stationary substance, and occasionally this has actually been done. But eventually, both optical and electrical phenomena have pressed upon us the conviction that the aether certainly cannot remain at rest where ponderable masses move through it, nor can it remain at rest in empty space, that rather it is traversed by internal currents. According to an investigation which Helmholtz carried out in the penultimate year of his life, the aether smoothly transfers energy from one particle to another without itself moving, as long as the energy is delivered to it evenly. Every build-up or acceleration in the energy supply, however, sets the aether into motion; and that will happen often enough.

Now that this hope has also come to naught, there remains no material object in the universe that one would have reason to consider at rest. There remain only certain conceptually defined points and directions that can serve as a substitute to a certain
[518] extent. For bringing these to its attention, astronomy owes thanks to a | science to which it is intimately related, namely mechanics. Thus we arrive at the path along which research is currently moving.

One of the basic laws of mechanics, the law of inertia, goes as follows: Each body moves in a straight line at an unchanging speed, as long as no forces act on it. This statement, however, contains more than a law based on experience; it contains at the same time a certain definition of what should be considered as being at rest in space. Since we can only identify relative motions, we can in theory ascribe any motion we choose to an individual body, even when it is unaffected by other forces. The law of inertia prescribes: Take any body (or to be more exact, so that rotations can also be recognized, any three bodies remaining at unchanging distances from each other) on which no forces act, and ascribe to these bodies a straight path and a uniform speed. The same will then apply for every other body on which no forces act. The system of inertia has so innumerably many important applications near and far that it is certainly fit and proper to make those arbitrary assumptions which lend it its simplest and only natural form.

So let us look at which ideas concerning celestial motions astronomy has to develop on the basis of this stipulation implicit in the law of inertia.

In a system comprising many mutually attracting bodies, naturally no single body describes a straight line in general, but when such a system is isolated from the effects of any forces from external, foreign bodies, still its center of gravity will describe a straight line, as shown by a simple conclusion from the law of inertia. If one treats all the bodies of the universe as constituents of one system, there would certainly be no external influence because nothing else exists beyond this system, and so its center of gravity must proceed along a straight line. Recall further an experience from mechanics, that in a train even at the highest speed, every activity can be carried out just as it can when all is at rest, provided that the speed does not alter and the train does not go around curves. This too is a special case of a general conclusion drawn from the law of inertia, which can be stated as follows: The internal processes within an isolated system of bodies are exactly the same whether its center of gravity is stationary or whether it is moving in a straight line at uniform speed. As far as the internal processes in the star system are concerned, it is therefore equally irrelevant whether its center of gravity is stationary or moving steadily. Since, furthermore, the only things we can learn about are internal processes in the star system, it is sensible to stipulate that the center of gravity of all masses in the universe and in the entire star system is the ideal point of rest.

It will, of course, look bad for the practical realization of this stipulation if, as is plausible, an infinite number of bodies of finite mass exist in the universe, all of which we then | cannot hope ever to know. However, the center of gravity of the larg- [519]
est possible system of masses is to be taken as an approximation to the definition of this ideal point of rest, especially when this system of masses is separate and remote from other such systems and as such experiences nearly no forces due to them. It is likely that the millions of stars that can be seen through a medium-sized telescope form a special system, which has been given the name "Milky Way System." The shimmering band of the Milky Way is in reality a huge ring comprised of a countless number of stars, which forms the largest mass in this system and characterizes its form. What remains of the brighter stars appear to be scattered inside this ring or to form small external appendages. Maybe other similar specimens existing in unimaginably distant regions of space are co-ordinated with our own Milky Way, but perhaps it is the only one of its kind and beyond it exist only chaotic nebular masses.

The center of gravity of the Milky Way System should then for the time being be seen as a point at rest. One will know more about its position when the great undertakings mentioned above are completed and have been sifted. However, this falls into the next phase of astronomy; for now, one makes do with a rather poor substitute. To date one knows only the proper motions of a few thousand stars, and for now one assumes that the center of gravity of these stars is at rest. As calculation shows, a speed of approximately twelve kilometres per second must then be ascribed to the Sun, a speed of similar magnitude to that with which the planets move in their orbits around the Sun.

Now there is still the matter of fixing certain directions with respect to which the rotations of each body can be reckoned. As far as the ring of the Milky Way is concerned, by analogy with all known celestial motions we must assume that it rotates in some way; therefore, from the outset it cannot be used for this purpose. However, here again new consequences of the law of inertia provide help.

Foucault's pendulum experiment is well known. If one hangs a heavy ball from a thread and causes this pendulum to oscillate from east to west, then in the course of a few hours one notices a rotation of the direction of oscillation, gradually moving in the north-south direction and, if only the pendulum swings long enough, it eventually completes a full revolution in over a day. This extremely remarkable process is usually considered to be the best proof of the Earth's rotation. For in theory, under the assumptions of the law of inertia, the 24-hour rotation of the Earth, and the Earth's gravitational pull, one finds just the amount of rotation of the oscillation direction shown by observations; whereas in the case of a stationary Earth, no rotation should take place at all. Strictly speaking, however, the Foucault pendulum experiment
[520] proves only | that a rotation of the Earth must be assumed whenever one wants to use the stipulations implicit in the law of inertia concerning that, to which motion is to be referred. Conversely, by accurate observations of the rotation of the pendulum, one could calculate the speed of rotation that has to be ascribed to the Earth according to the law of inertia, and one would then have to describe as fixed the direction with respect to which the Earth rotates with the calculated speed.

Because terrestrial pendulums are subject to too many disturbances due to air resistance and friction, one uses with a similar aim the larger pendulum experiments with which nature presents us. One can cause a pendulum to describe an elongated elliptical curve by pushing it at an angle, as in a well-known game of bowling. The planets are pendulum bobs of a similar kind. It is well known that the planets describe flattened curves, ellipses, around the Sun. The point in its path where a planet is furthest away from the Sun is called its aphelion. From the law of inertia and Newton's law of gravitation, it follows that for an isolated planet which revolves around the Sun, the direction of the aphelion is fixed in space. In reality, however, the planets are not isolated but exist in greater number, yet Newton's law allows the calculation of the small deviations from the elliptical path and the small rotations which the direction to the aphelion suffers as a result of the disturbing influences from other planets. Thus, after subtracting these disturbances from the observed aphelion directions, a direction fixed in space is obtained.

If one carries this out with the level of precision which was possible 50 years ago, then everything seems to fit together beautifully. The directions that, according to the theory, should turn out to be fixed for the different planets also appear to be fixed with respect to each other, and additionally it turns out that the average of those few thousand stars, whose proper motions we know, displays no evidence of rotation with respect to these directions within these limits of precision. According to this, the entire ring of the Milky Way can also rotate only extremely slowly. Today, however, things look different. With the level of precision which theory and observation have now achieved, one finds that the directions which one derives from the aphelions of the various planets and which one would expect to be fixed, actually perform minimal yet clearly recognizable shifts relative to each other, and thus it is impossible to treat them all as fixed. Our measure of time depends upon which directions one considers to be fixed. For we define a day to be the time it takes the Earth to revolve once around its axis and we must, of course, have a fixed direction within space in order to be able to judge when the Earth has completed a rotation. Two clocks, which agree at the beginning of a century, one tied to the direction of Jupiter's aphelion, the other to Mercury's aphelion, | would differ by three seconds at the end of it. One does not [521] really know how to explain this difference. It is possible that friction plays a role in the case of these heavenly pendulums as well, for they do not swing through an absolutely empty space. However, it is more likely that Newton's law does not describe the attractive forces of the Sun and the planets with absolute accuracy. However, there are still insufficient clues to know how a correction of Newton's law should actually read, and so one must renounce the wish to determine fixed directions with greater precision from the planetary motions using mechanical theorems. At present one prefers to hold on to the other result which, from these considerations, proved itself to be reasonably correct, and to treat those few thousand stars as being without rotation on average. That is the provisional stipulation that, in a roundabout way, one has to choose also as regards the definition of rotations in the universe.

Finally, a certain arbitrariness, with which each definition of fixed points and directions based on mechanical theorems is afflicted, is still to be pointed out. One can only base a conclusion on the law of inertia when all forces that act in a given case are known, or when it is known that no forces are present. Now there could always be forces acting in our surroundings, which spin us, together with all the neighboring stars, around arbitrarily in the universe, without however exerting any influence on the relative position of all these bodies. Such forces would, of course, completely elude our experience, and therefore the principle of utility, which alone guides us, commands us not to allow ourselves to become disconcerted by the thought of such a possibility when using theorems based on the law of inertia.

One senses a certain feeling of unease when one stops and thinks about all that is provisional, intermediate, undecided in present-day science concerning a point which is so important for a clear idea of the universe, as is the establishment of what is at rest in the universe. Yet, that is a characteristic of our times. The proud era of natural science, when it believed to have found absolute laws and to be able to give philosophy a real basis, is over. Similar instances of relativity as are found in the case of motion, exist everywhere, and with each broadening of experience, uncertainties turn up in previously accepted definitions. A profound scepticism has become fashionable, for one asks oneself even about the foundations of exact natural science, which are given the honorable name of "laws of nature": what can they be other than the most practical summary possible of what is most important for man within a limited field of experience?

EDITORIAL NOTE

[1] Mount Kahlenberg lies to the north of central Vienna, about 8 km from Stephan's Tower.

A NEW LAW OF GRAVITATION ENFORCED
BY SPECIAL RELATIVITY

SCOTT WALTER

BREAKING IN THE 4-VECTORS: THE FOUR-DIMENSIONAL MOVEMENT IN GRAVITATION, 1905–1910

INTRODUCTION

In July, 1905, Henri Poincaré (1854–1912) proposed two laws of gravitational attraction compatible with the principle of relativity and all astronomical observations explained by Newton's law. Two years later, in the fall of 1907, Albert Einstein (1879–1955) began to investigate the consequences of the principle of equivalence for the behavior of light rays in a gravitational field. The following year, Hermann Minkowski (1864–1909), Einstein's former mathematics instructor, borrowed Poincaré's notion of a four-dimensional vector space for his new matrix calculus, in which he expressed a novel theory of the electrodynamics of moving media, a space-time mechanics, and two laws of gravitational attraction. Following another two-year hiatus, Arnold Sommerfeld (1868–1951) characterized the relationship between the laws proposed by Poincaré and Minkowski, calling for this purpose both on space-time diagrams and a new 4-vector formalism.

Of these four efforts to capture gravitation in a relativistic framework, Einstein's has attracted the lion's share of attention, and understandably so in hindsight, but at the expense of a full understanding of what is arguably the most significant innovation in contemporary mathematical physics: the four-dimensional approach to laws of physics. In virtue of the common appeal made by Poincaré, Minkowski, and Sommerfeld to four-dimensional vectors in their studies of gravitational attraction, their respective contributions track the evolving form of four-dimensional physics in the early days of relativity theory.[1] The objective of this paper is to describe in terms of theorists' intentions and peer readings the emergence of a four-dimensional language for physics, as applied to the geometric and symbolic expression of gravitational action.

The subject of gravitational action at the turn of the twentieth century is well-suited for an investigation of this sort. This is not to say that the reform of Newton's

---

1 In limiting the scope of this paper to the methods applied by their authors to the problem of gravitation, four contributions to four-dimensional physics are neglected: that of Richard Hargreaves, based on integral invariants (Hargreaves 1908), two 4-vector systems due to Max Abraham (Abraham 1910) and Gilbert Newton Lewis (Lewis 1910a), and Vladimir Varičak's hyperbolic-function based approach (Varičak 1910).

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 3
Gravitation in the Twilight of Classical Physics: Between Mechanics, Field Theory, and Astronomy.

law was a burning issue for theorists. While several theories of gravitation claimed corroboration on a par with that of classical Newtonian theory, contemporary theoretical interest in gravitation as a research topic—including the Lorentz-invariant variety—was sharply curtailed by the absence of fresh empirical challenges to the inverse-square law. Rather, in virtue of the stability of the empirical knowledge base, and two centuries of research in celestial mechanics, the physics of gravitation was a well-worked, stable terrain, familiar to physicists, mathematicians and astronomers alike.[2]

The leading theory of gravitation in 1905 was the one discovered by Isaac Newton over two centuries earlier, based on instantaneous action at a distance. When Poincaré sought to bring gravitational attraction within the purview of the principle of relativity, he saw it had to propagate with a velocity no greater than that of light in empty space, such that a reformulation of Newton's law as a retarded action afforded a simple solution.

Newton's law was the principal model for Poincaré, but it was not the only one. With the success of Maxwell's theory in explaining electromagnetic phenomena (including the behavior of light) during the latter third of the nineteenth century, theories of contiguous action gained greater favor with physicists. In 1892, the Dutch theorist H. A. Lorentz produced a theory of mobile charged particles interacting in an immobile aether, that was an habile synthesis of Maxwell's field theory and Wilhelm Weber's particle theory of electrodynamics. After the discovery of the electron in 1897, and Lorentz's elegant explanation of the Zeeman effect, certain charged microscopic particles were understood to be electrons, and electrons the building-blocks of matter.[3]

In this new theoretical context of aether and electrons, Lorentz derived the force on an electron moving in microscopic versions of Maxwell's electric and magnetic fields. To determine the electromagnetic field of an electron in motion, Alfred Liénard and Emil Wiechert derived a formula for a potential propagating with finite velocity. In virtue of these two laws, both of which fell out of a Lagrangian from Karl Schwarzschild, the theory of electrons provided a means of calculating the force on a charged particle in motion due to the fields of a second charged particle in motion.[4]

An electron-based analogy to gravitational attraction of neutral mass points was then close at hand. Lorentz's electron theory was held in high esteem by early twenti-

2 For an overview of research on gravitation from 1850 to 1915, see (Roseveare 1982). On early 20th-century investigations of gravitational absorption, see de Andrade Martins (de Andrade Martins 1999). While only Lorentz-covariant theories are considered in this paper, the relative acceptance of the principle of relativity among theorists is understood as one parameter among several influencing the development of four-dimensional physics.

3 See (Buchwald 1985, 242; Darrigol 2000, 325; Buchwald and Warwick 2001).

4 Lorentz took the force per unit charge on a volume element of charged matter moving with velocity $\mathfrak{v}$ in the electric and magnetic fields $\mathfrak{d}$ and $\mathfrak{h}$ to be $\mathfrak{f} = \mathfrak{d} + \frac{1}{c}[\mathfrak{v} \cdot \mathfrak{h}]$, where the brackets indicate a vector product (Lorentz 1904c, 2:156–7). For a comparison of electrodynamic Lagrangians from Maxwell to Schwarzschild, see (Darrigol 2000, app. 9).

eth-century theorists, including both Poincaré and Minkowski, who naturally catered to the most promising research program of the moment. They each proposed two force laws: one based on retarded action at a distance, the other appealing directly to contiguous action propagated in a medium. All four particle laws were taken up in turn by Sommerfeld.[5]

Several other writers have discussed Poincaré's and Minkowski's work on gravitation. Of the first four substantial synoptic reviews of the two theories, none employed the notation of the original works, although this fact itself reflects the rapid evolution of formal approaches in physics. Early comparisons were carried out with either Sommerfeld's 4-vector formalism (Sommerfeld 1910b; Kretschmann 1914), a relative coordinate notation (de Sitter 1911), or a mix of ordinary vector algebra and tensor calculus (Kottler 1922). No further comparison studies were published after 1922, excepting one summary by North (North 1965, 49–50), although since the 1960s, the work of Poincaré and Minkowski has continued to incite historical interest.[6] Sommerfeld's contribution, while it inflected theoretical practice in general, and contemporary reception of Lorentz-covariant gravitation theory in particular, has been neglected by historians.

The present study has three sections, beginning with Poincaré's contribution, moving on in the second section to Minkowski's initial response to Poincaré's theory, and a review of his formalism and laws of gravitation. A third section is taken up by Sommerfeld's interpretation of the laws proposed by Poincaré and Minkowski. The period of study is thus bracketed on one end by the discovery of special relativity in 1905, and on the other end by Sommerfeld's paper. While the latter work did not spell the end of either 4-vector formalisms or Lorentz-covariant theories of gravitation, it was the first four-dimensional vector algebra, and represents a point of closure for a study of the emergence of a conceptual framework for four-dimensional physics.

1. HENRI POINCARÉ'S LORENTZ-INVARIANT LAWS OF GRAVITATION

Poincaré's memoir on the dynamics of the electron (Poincaré 1906), like Einstein's relativity paper of 1905, contains the fundamental insight of the physical significance of the group of Lorentz transformations, not only for electrodynamics, but for all natural phenomena. The law of gravitation, to no lesser extent than the laws of electrodynamics, fell presumably within the purview of Einstein's theory, but this is not a point that Einstein, then working full time as a patent examiner in Bern, chose to elaborate upon immediately. Poincaré, on the other hand, as Professor of Mathematical Astron-

---

5 On the Maxwellian approach to gravitation, see (North 1965, chap. 3; Roseveare 1982, 129–31; Norton 1992, 32). The distinction drawn here between retarded action at a distance and field representations reflects that of Lorentz (Lorentz 1904b), for whom this was largely a matter of convenience. On nineteenth-century conceptions of the electromagnetic field, see (Cantor and Hodge 1981).

6 On Poincaré's theory see (Cunningham 1914, 173; Whitrow and Morduch 1965, 20; Harvey 1965, 452; Cuvaj 1970, app. 5; Schwartz 1972; Zahar 1989, 192;Torretti 1996, 132). On Minkowski's theory see (Weinstein 1914, 61; Pyenson 1985, 88; Corry 1997, 287).

omy and Celestial Mechanics at the Sorbonne, could hardly finesse the question of gravitation. In particular, his address to the scientific congress at the St. Louis World's Fair, on 24 September, 1904, had pinpointed Laplace's calculation of the propagation velocity of gravitation as a potential spoiler for the principle of relativity.[7]

There may have been another reason for Poincaré to investigate a relativistic theory of gravitation. In the course of his study of Lorentz's contractile electron, Poincaré noted that the required relations between electromagnetic energy and momentum were not satisfied in general. Raised earlier by Max Abraham, the problem was considered by Lorentz to be a fundamental one for his electron theory.[8]

Solving the stability problem of Lorentz's contractile electron was a trivial matter for Poincaré, as it meant transposing to electron theory a special solution to a general problem he had treated earlier at some length: to find the equilibrium form of a rotating fluid mass.[9] He postulated a non-electromagnetic, Lorentz-invariant "supplementary" potential that exerts a binding (negative) pressure inside the electron, and reduces the total energy of the electron in an amount proportional to the volume decrease resulting from Lorentz contraction. When combined with the electromagnetic field Lagrangian, this binding potential yields a total Lagrangian invariant with respect to the Lorentz group, as Poincaré required.

In accordance with the electromagnetic world-picture and the results of Kaufmann's experiments, Poincaré supposed the inertia of matter to be exclusively of electromagnetic origin, and he set out, as he wrote in §6 of his paper,

> to determine the total energy due to electron motion, the corresponding action, and the quantity of electromagnetic momentum, in order to calculate the electromagnetic masses of the electron.

---

7 Laplace estimated the propagation velocity of gravitation to be 10^6 times that of light, and Poincaré noted that such a signal velocity would allow inertial observers to detect their motion with respect to the aether (Poincaré 1904, 312).

8 See (Poincaré 1906, 153–154; Miller 1973, 230–233). Following Abraham's account (Abraham 1905, 205), the problem may be presented in outline as follows (using modified notation and units). Consider a deformable massless sphere of radius a and uniformly distributed surface charge, and assume that this is a good model of the electron. The longitudinal mass $m_{||}$ of this sphere may be defined as the quotient of external force and acceleration, $m_{||} = d|\boldsymbol{G}|/d|\boldsymbol{v}|$, where $\boldsymbol{G}$ is the electromagnetic momentum resulting from the electron's self-fields, and $\boldsymbol{v}$ is electron velocity. Defining the electromagnetic momentum to be $\boldsymbol{G} = \int \boldsymbol{E} \times \boldsymbol{B}\, dV$, where $\boldsymbol{E}$ and $\boldsymbol{B}$ denote the electric and magnetic self-fields, and V is for volume, we let $c = 1$, and find the longitudinal mass for small velocities to be $m_{||} = \frac{e^2}{6\pi a}(1-v^2)^{-3/2}$. Longitudinal electron mass may also be defined in terms of the electromagnetic energy W of the electron's self-fields, assuming quasistationary motion: $m_{||} = \frac{1}{|\boldsymbol{v}|}\frac{dW}{d|\boldsymbol{v}|}$, where $W = \frac{e^2}{6\pi a}(1-v^2)^{-1/2} + \frac{e^2}{24\pi a}(1-v^2)^{1/2}$. This leads, however, to an expression for longitudinal mass different from the previous one: $m_{||} = \frac{e^2}{6\pi a}\left[(1-v^2)^{-3/2} + \frac{1}{4}(1-v^2)^{-1/2}\right]$. From the difference in these two expressions for longitudinal mass, Abraham concluded that the Lorentz electron required the postulation of a non-electromagnetic force and was thereby not compatible with a purely electromagnetic foundation of physics.

9 See (Poincaré 1885, 1902a, 1902b). In the limit of null angular velocity, gravitational attraction can be replaced by electrostatic repulsion, with a sign reversal in the pressure gradient.

Non-electromagnetic mass does not figure in this analysis, and consequently, one would not expect the non-electromagnetic binding potential to contribute to the tensorial electromagnetic mass of the electron, although Poincaré did not state this in so many words. Instead, immediately after obtaining an expression for the binding potential, he derived the small-velocity, "experimental" mass from the electromagnetic field Lagrangian alone, neglecting a contribution from the binding potential. The mass of the slowly-moving Lorentz electron was then equal to the electrostatic mass, just as one would want for an electromagnetic foundation of mechanics. This fortuitous result, which revised Lorentz's electron mass value downward by a quarter, was obtained independently by Einstein, using a method that did not constrain electron structure (Einstein 1905, 917).[10] Although the question of electron mass was far from resolved, Poincaré had shown that the stability problem represented no fundamental obstacle to the pursuit of a new mechanics based on the concept of a contractile electron.

With this obstacle out of the way, Poincaré proceeded as if the laws of mechanics were applicable to the experimental mass of the electron.[11] Noting that the negative pressure deriving from his binding potential is proportional to the fourth power of mass, and furthermore, that Newtonian attraction is itself proportional to mass, Poincaré conjectured that

> there is some relation between the cause giving rise to gravitation and that giving rise to the supplementary potential.

On the basis of a formal relation between experimental mass and the binding potential, in other words, Poincaré predicted the unification of his negative internal electron pressure with the gravitational force, in a future theory encompassing all three forces.[12]

On this hopeful note, Poincaré began his memoir's ninth and final section, entitled "Hypotheses concerning gravitation." Lorentz's theory, Poincaré explained, promised to account for the observed relativity of motion:

> In this way Lorentz's theory would fully explain the impossibility of detecting absolute motion, if all forces were of electromagnetic origin.[13]

---

10 Poincaré also neglected the mass contribution of the binding potential in his 1906–1907 Sorbonne lectures, according to student notes (Poincaré 1953, 233). For reviews of Poincaré's derivation of the binding potential, see (Cuvaj 1970, app. 11) and (Miller 1973). On post-Minkowskian interpretations of the binding potential (also known as Poincaré pressure), see (Cuvaj 1970, 203; Miller 1981, 382, n. 29; Yaghjian 1992).

11 In this paper Poincaré made no distinction between inertial and gravitational mass.

12 As Cuvaj points out (Cuvaj 1968, 1112), Poincaré may have found inspiration for this conjecture in Paul Langevin's remark that gravitation stabilized the electron against Coulomb repulsion. Unlike Langevin, Poincaré anticipated a unified theory of gravitation and electrons, in the spirit of theories pursued later by Gustav Mie, Gunnar Nordström, David Hilbert, Hans Reissner, Hermann Weyl and Einstein; for an overview see (Vizgin 1994).

13 "Ainsi la théorie de Lorentz expliquerait complètement l'impossibilité de mettre en évidence le mouvement absolu, si toutes les forces étaient d'origine électromagnétique" (Poincaré 1906, 166).

The hypothesis of an electromagnetic origin of gravitational force had been advanced by Lorentz at the turn of the century. On the assumption that the force between "ions" (later "electrons") of unlike sign was of greater magnitude at a given separation than that between ions of like sign (following Mossotti's conjecture), Lorentz represented gravitational attraction as a field-theoretical phenomenon analogous to electromagnetism, reducing to the Newtonian law for bodies at rest with respect to the aether. Lorentz's theory tacitly assumed negative energy density for the "gravitational" field, and a gravitational aether of huge intrinsic positive energy density, two well-known sticking-points for Maxwell. Another difficulty stemmed from the dependence of gravitational force on absolute velocities.[14]

Neither Lorentz's gravitation theory nor Maxwell's sticking-points were mentioned by Poincaré in the ninth section of his memoir. Instead, he recalled a well-known empirical fact: two bodies that generate identical electromagnetic fields need not exert the same attraction on electrically neutral masses. Although Lorentz's theory clearly accounts for this fact, Poincaré concluded that the gravitational field was distinct from the electromagnetic field. What this tells us is that Poincaré's attention was not focused on Lorentz's theory of gravitation.[15]

To Poincaré's way of thinking, it was the impossibility of an electromagnetic reduction of gravitation that had driven Lorentz to suppose that all forces transform like electromagnetic ones:

> The gravitational field is therefore distinct from the electromagnetic field. Lorentz was obliged thereby to extend his hypothesis with the assumption that *forces of any origin whatsoever, and gravitation in particular, are affected by a translation* (or, if one prefers, by the Lorentz transformation) *in the same manner as electromagnetic forces*. (Poincaré 1906, 166.)[16]

---

14 See (Lorentz 1900; Havas 1979, 83; Torretti 1996, 131). On Lorentz's precursors see (Whittaker 1951–1953, 2:149; Zenneck 1903). Lorentz's theory of gravitation failed to convince Oliver Heaviside, who had carefully weighed the analogy from electromagnetism to gravitation (Heaviside 1893). In a letter to Lorentz, Heaviside called into question the theory's electromagnetic nature, by characterizing Lorentz's gravitational force as "action at a distance of a double kind" (18 July, 1901, Lorentz Papers, Rijksarchief in Noord-Holland te Haarlem). Aware of these difficulties, Lorentz eventually discarded his theory, citing its incompatibility with the principle of relativity (Lorentz 1914, 32).

15 In his 1906–1907 Sorbonne lectures (Poincaré 1953), Poincaré discussed a different theory (based on an idea due to Le Sage) that Lorentz had proposed in the same paper, without mentioning the Mossotti-style theory. His first discussion of the latter theory was in 1908, when he considered it to be an authentic relativistic theory, and one in which the force of gravitation was of electromagnetic origin (Poincaré 1908, 399).

16 Poincaré's account of Lorentz's reasoning should be taken with a grain of salt, as Lorentz made no mention of his theory of gravitation in the 1904 publication referred to by Poincaré, "Electromagnetic phenomena in a system moving with any velocity less than that of light." While the electron theory developed in the latter paper did not address the question of the origin of the gravitational force, it admitted the possibility of a reduction to electromagnetism (such as that of his own theory) by means of the additional hypothesis referred to in the quotation: all forces of interaction transformed in the same way as electric forces in an electrostatic system (Lorentz 1904a, §8). The contraction hypothesis formerly invoked to account for the null result of the Michelson-Morley experiments, Lorentz added, was subsumed by the new hypothesis.

It was the cogency of the latter hypothesis that Poincaré set out to examine in detail, with respect to gravitational attraction. The situation was analogous to the one Poincaré had encountered in the case of electron energy and momentum mentioned above, where he had considered constraining internal forces of the electron to be Lorentz-invariant. Such a constraint solved the problem immediately, but Poincaré recognized that it was inadmissible nonetheless, because it violated Maxwell's theory (p. 136). A similar violation in the realm of mechanics could not be ruled out in the case of gravitation, such that a careful analysis of the admissibility of the formal requirement of Lorentz-invariance was called for.

Poincaré set out to determine a general expression for the law of gravitation in accordance with the principle of relativity. A relativistic law of gravitation, he reasoned, must obey two constraints distinguishing it from the Newtonian law. First of all, the new force law could no longer depend solely on the masses of the two gravitating bodies and the distance between them. The force had to depend on their velocities, as well. Furthermore, gravitational action could no longer be considered instantaneous, but had to propagate with some finite velocity, so that the force acting on the passive mass would depend on the position and velocity of the active mass at some earlier instant in time. A gravitational propagation velocity greater than the speed of light, Poincaré observed, would be "difficult to understand," because attraction would then be a function of a position in space not yet occupied by the active mass (p. 167).

These were not the only conditions Poincaré wanted to satisfy. The new law of gravitation had also (1) to behave in the same way as electromagnetic forces under a Lorentz transformation, (2) to reduce to Newton's law in the case of relative rest of the two bodies, and (3) to come as close as possible to Newton's law in the case of small velocities. Posed in this way, Poincaré noted, the problem remains indeterminate, save in the case of null relative velocity, where the propagation velocity of gravitation does not enter into consideration. Poincaré reasoned that if two bodies have a common rectilinear velocity, then the force on the passive mass is orthogonal to an ellipsoid, at the center of which lies the active mass.

Undeterred by the indeterminacy of the question in general, Poincaré set about identifying quantities invariant with respect to the Lorentz group, from which he wanted to construct a law of gravitation satisfying the constraints just mentioned. To assist in the identification and interpretation of these invariants, Poincaré referred to a space of four dimensions. "Let us regard," he wrote,

$$\begin{array}{cccc} x, & y, & z, & t\sqrt{-1} \\ \delta x, & \delta y, & \delta z, & \delta t\sqrt{-1} \\ \delta_1 x, & \delta_1 y, & \delta_1 z, & \delta_1 t\sqrt{-1}, \end{array}$$

> as the coordinates of 3 points P, P', P'', in space of 4 dimensions. We see that the Lorentz transformation is merely a rotation in this space about the origin, regarded as fixed. Consequently, we will have no distinct invariants apart from the 6 distances between the 3 points P, P', P'', considered separately and with the origin, or, if one prefers, apart from the 2 expressions:

$$x^2 + y^2 + z^2 - t^2, \quad x\delta x + y\delta y + z\delta z - t\delta t,$$

> or the 4 expressions of like form deduced by arbitrary permutation of the 3 points P, P', P''.[17]

Here Poincaré formed three quadruplets representing the differential displacement of two point masses, with respect to a certain four-dimensional vector space, later called a pseudo-Euclidean space.[18] By introducing such a 4-space, Poincaré simplified the task of identifying quantities invariant with respect to the Lorentz transformations, the line interval of the new space being formally identical to that of a Euclidean 4-space. He treated his three points P, P', and P'' as 4-vectors, the scalar products of which are invariant, just as in Euclidean space. In fact, Poincaré did not employ vector terminology or notation in his study of gravitation, but provided formal definitions of certain objects later called 4-vectors.

Poincaré's habit, and that of the overwhelming majority of his French colleagues in mathematical physics well into the 1920s, was to express ordinary vector quantities in Cartesian coordinate notation, and to forgo notational shortcuts when differentiating, writing these operations out in full.[19] Although he did not exclude symbols such as Δ or $\Box$ from his scientific papers and lectures, he employed them parsimoniously.[20] In line with this practice, Poincaré did little to promote vector methods from his chair at the Sorbonne. In twenty volumes of lectures on mathematical physics and celestial mechanics, there is not a single propadeutic on quaternions or vector algebra.[21] Poincaré deplored the "long calculations rendered obscure by notational

---

17 "Regardons $x, y, z, t\sqrt{-1}, \delta x, \delta y, \delta z, \delta t\sqrt{-1}, \delta_1 x, \delta_1 y, \delta_1 z, \delta_1 t\sqrt{-1}$, comme les coordonnées de 3 points P, P', P'' dans l'espace à 4 dimensions. Nous voyons que la transformation de Lorentz n'est qu'une rotation de cet espace autour de l'origine, regardée comme fixe. Nous n'aurons donc pas d'autres invariants distincts que les six distances des trois points P, P', P'' entre eux et à l'origine, ou, si l'on aime mieux, que les 2 expressions: $x^2 + y^2 + z^2 - t^2, x\delta x + y\delta y + z\delta z - t\delta t$, ou les 4 expressions de même forme qu'on en déduit en permutant d'une manière quelconque les 3 points P, P', P''" (Poincaré 1906, 168–9).

18 Poincaré's three points P, P', P'' may be interpreted in modern terminology as follows. Let the spacetime coordinates of the passive mass point be $A = (x_0, y_0, z_0, t_0)$, with ordinary velocity $\xi = (\delta x/\delta t, \delta y/\delta t, \delta z/\delta t)$, such that at time $t_0 + \delta t$ it occupies the spacetime point $A' = (x_0 + \delta x, y_0 + \delta y, z_0 + \delta z, t_0 + \delta t)$. Likewise for the active mass point, $B = (x_0 + x, y_0 + y, z_0 + z, t_0 + t)$, with ordinary velocity $\xi_1 = (\delta_1 x/\delta_1 t, \delta_1 y/\delta_1 t, \delta_1 z/\delta_1 t)$, such that at time $t_0 + t + \delta_1 t$, it occupies the spacetime point $B' = (x_0 + x + \delta_1 x, y_0 + y + \delta_1 y, z_0 + z + \delta_1 z, t_0 + t + \delta_1 t)$. Poincaré's three quadruplets may now be expressed as position 4-vectors: $P = B - A, P' = B' - B, P'' = A' - A$.

19 While the first German textbook on electromagnetism to employ vector notation systematically dates from 1894 (Föppl 1894), the first comparable textbook in French was published two decades later by Jean-Baptiste Pomey (1861–1943), instructor of theoretical electricity at the *École supérieure des Postes et Télégraphes* in Paris (Pomey 1914–1931, vol. 1).

20 The Laplacian was expressed generally as $\nabla^2 = \partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2$, but by Poincaré as Δ. The d'Alembertian, $\Box \equiv \partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2 - \partial^2/\partial t^2$, became in Poincaré's notation: $\Box \equiv \Delta - d^2/dt^2$. Poincaré employed $\Box$ in his lectures on electricity and optics (Poincaré 1901, 456), and was the first to employ it in a relativistic context.

complexity" in W. Voigt's molecular theory of light, and seems to have been of the opinion that in general, new notation only burdened the reader.[22]

The point of forming quadruplets was to obtain a set of Lorentz-invariants corresponding to the ten variables entering into the right-hand side of the new force law, representing the squared distance in space and time of the two bodies and their velocities $(\xi, \eta, \zeta, \xi_1, \eta_1, \zeta_1)$. How did Poincaré obtain his invariants? According to the method cited above, six invariants were to be found from the distances between P, P', P'', and the origin, or from the scalar products of P, P', and P''. These six intermediate invariants were then to be combined to obtain homogeneous invariants depending on the duration of propagation of gravitational action and the velocities of the two point masses. Poincaré skipped over the intermediate step and produced the following four invariants, in terms of squared distance, distance and velocity (twice), and the velocity product:

$$\sum x^2 - t^2, \frac{t - \sum x\xi}{\sqrt{1 - \sum \xi^2}}, \frac{t - \sum x\xi_1}{\sqrt{1 - \sum \xi_1{}^2}}, \frac{1 - \sum \xi\xi_1}{\sqrt{(1 - \sum \xi^2)(\sqrt{1 - \sum \xi_1{}^2})}}. \tag{1}$$

The Lorentz-invariance and geometric significance of these quantities are readily verified.[23] These four invariants (1), the latter three of which were labeled A, B, and C, formed the core of Poincaré's constructive approach to the law of gravitation. (For convenience, I refer to Poincaré's four invariants [1] as his "kinematic" invariants.)

Inspection of the signs of these invariants reveals an inconsistency, the reason for which is apparent once the intermediate calculations have been performed. Instead of constructing his four invariants out of scalar products, Poincaré introduced an inversion for A, B, and C.[24] This sign inconsistency had no consequence on his search

21 Poincaré's manuscript lecture notes for celestial mechanics, however, show that he saw fit to introduce the quaternionic method to his students (undated notebook on quaternions and celestial mechanics, 32 pp., private collection, Paris; hpcd 76, 78, 93, Henri Poincaré Archives, Nancy).

22 Manuscript report of the PhD. thesis submitted by Henri Bouasse, 13 December, 1892, AJ[16]5535, *Archives Nationales*, Paris. From Poincaré's conservative habits regarding formalism, he appears as an unlikely candidate at best for the development of a four-dimensional calculus circa 1905; cf. H. M. Schwartz's counterfactual conjecture: if Poincaré had adopted the ordinary vector calculus by the time he wrote his *Rendiconti* paper, "he would have in all likelihood introduced explicitly ... the convenient four-dimensional vector calculus" (Schwartz 1972, 1287, n. 7).

23 The invariants (1) may be expressed in ordinary vector notation, letting $\Sigma x = \boldsymbol{x}, \Sigma\xi = \boldsymbol{v}, \Sigma\xi_1 = \boldsymbol{v}_1$, and for convenience, $k = 1/\sqrt{1-\Sigma\xi^2}, k_1 = 1/\sqrt{1-\Sigma\xi_1{}^2}$, such that the four quantities (1) read as follows: $\boldsymbol{x}^2 - t^2, k(t - \boldsymbol{xv}), k_1(t - \boldsymbol{xv}_1), kk_1(1 - \boldsymbol{vv}_1)$.

24 Poincaré's four kinematic invariants (1) are functions of the following six intermediate invariants: $a = x^2 + y^2 + z^2 - t^2$, $b = x\delta x + y\delta y + z\delta z - t\delta t$, $c = x\delta_1 x + y\delta_1 y + z\delta_1 z - t\delta_1 t$, $d = \delta x\delta_1 x + \delta y\delta_1 y + \delta z\delta_1 z - \delta t\delta_1 t$, $e = \delta x^2 + \delta y^2 + \delta z^2 - \delta t^2$, $f = \delta_1 x^2 + \delta_1 y^2 + \delta_1 z^2 - \delta_1 t^2$. In terms of the latter six invariants, the four kinematic invariants (1) may be expressed as follows: $\Sigma x^2 - t^2 = a$, $A = -b/\sqrt{-e}$, $B = -c/\sqrt{-f}$, and $C = -d/(\sqrt{-e}\sqrt{-f})$. For a slightly different reconstruction of Poincaré's kinematic invariants, see (Zahar 1989, 193).

for a relativistic law of gravitation, although it affected his final result, and perplexed at least one of his readers, as I will show in section 3.

What Poincaré needed next for his force law was a Lorentz-invariant expression for the force itself. Up to this point, he had neither a velocity 4-vector nor a force 4-vector definition on hand. Presumably, the search for Lorentz-invariant expressions of force led him to define these 4-vectors. Earlier in his memoir (p. 135), Poincaré had determined the Lorentz transformations of force density, but now he was interested in the Lorentz transformations of force at a point. The transformations of force density:

$$X' = k(X + \varepsilon T), \quad Y' = Y, \quad Z' = Z, \quad T' = k(T + \varepsilon X), \tag{2}$$

where k is the Lorentz factor, $k = 1/\sqrt{1-\varepsilon^2}$, and ε designates frame velocity, led Poincaré to define a fourth component of force density, T, as the product of the force density vector with velocity, $T = \sum X\xi$.[25] He gave the same definition for the temporal component of force at a point: $T_1 = \sum X_1\xi$.[26] Next, dividing force density by force at a point, Poincaré obtained the charge density ρ. Ostensibly from the transformation for charge density, Poincaré singled out the Lorentz-invariant factor:[27]

$$\frac{\rho}{\rho'} = \frac{1}{k(1+\xi\varepsilon)} = \frac{\delta t}{\delta t'}. \tag{3}$$

The components of a 4-velocity vector followed from the foregoing definitions of position and force density:

> The Lorentz transformation ... will act in the same way on $\xi, \eta, \zeta, 1$ as on $\delta x, \delta y, \delta z, \delta t$, with the difference that these expressions will be multiplied moreover by the *same* factor $\delta t/\delta t' = 1/k(1+\xi\varepsilon)$.[28]

Concerning the latter definition, Poincaré observed a formal analogy between the force and force density 4-vectors, on one hand, and the position and velocity 4-vectors, on the other hand: these pairs of vectors transform in the same way, except that one member is multiplied by $1/k(1+\xi\varepsilon)$. While this analogy may seem mathematically transparent, it merits notice, as it appears to have eluded Poincaré at first.

With these four kinematic 4-vectors in hand, Poincaré defined a fifth quadruplet Q with components of force density $(X, Y, Z, T\sqrt{-1})$. Just as in the previous case, the scalar products of his four quadruplets P, P', P'', and Q were to deliver four new Lorentz-invariants in terms of the force acting on the passive mass (X_1, Y_1, Z_1):[29]

---

25 This definition was remarked by (Pauli 1921, 637).

26 The same subscript denotes the *force* acting on the *passive* mass, ΣX_1, and the *velocity* of the *active* mass, ξ_1.

27 The ratio ρ/ρ' is equal to the Lorentz factor, since in Poincaré's configuration, $\varepsilon = -\xi$. Some writers hastily attribute a 4-current vector to Poincaré, the form $\rho(\xi, \eta, \zeta, i)$ being implied by Poincaré's 4-vector definitions of force density and velocity.

28 "La transformation de Lorentz ... agira sur $\xi, \eta, \zeta, 1$ de la même manière que sur $\delta x, \delta y, \delta z, \delta t$, avec cette différence que ces expressions seront en outre multipliées par le *même* facteur $\delta t/\delta t' = 1/k(1+\xi\varepsilon)$" (Poincaré 1906, 169).

$$\frac{\sum X_1^2 - T_1^2}{1-\sum \xi^2},\ \frac{\sum X_1 x - T_1 t}{\sqrt{1-\sum \xi^2}},\ \frac{\sum X_1 \xi_1 - T_1}{\sqrt{1-\sum \xi^2 \sqrt{1-\sum \xi_1^2}}},\ \frac{\sum X_1 \xi - T_1}{\sqrt{1-\sum \xi^2}}. \tag{4}$$

The fourth invariant in (4) was always null by definition of $T_{1,}$ leaving only three invariants, denoted $M, N,$ and P. (In order to distinguish these invariants from the kinematic invariants, I will refer to [4] as Poincaré's "force" invariants.)

Comparing the signs of the kinematic invariants (1) with those of the force invariants (4), we see that Poincaré obtained consistent signs only for the latter invariants. He must not have computed his force invariants in the same way as his kinematic invariants, for reasons that remain obscure. It is not entirely unlikely that in the course of his analysis of the transformations of velocity and force, Poincaré realized that he could compute the force invariants directly from the scalar products of four 4-vectors. Two facts, however, argue against this reading. In the first place, Poincaré did not mention that his force invariants were the scalar products of position, velocity and force 4-vectors. Secondly, he did not alter the signs of his kinematic invariants to make them correspond to scalar products of position and velocity 4-vectors.[30] The fact that Poincaré's kinematic invariants differ from products of 4-position and 4-velocity vectors leads us to believe that when forming these invariants he was *not* thinking in terms of 4-vectors.[31]

From this point on, Poincaré worked exclusively with arithmetic combinations of three force invariants (M, N, P) and four kinematic invariants $(\sum x^2 - t^2, A, B, C)$ in order to come up with a relativistic law of gravitation. He had no further use, in particular, for the four quadruplets he had identified in the process of constructing

---

29 The invariants (4) may be expressed in ordinary vector notation, recalling the definitions of note 23, and letting $\Sigma X_1 = \boldsymbol{f}_1$, and $T_1 = \boldsymbol{f}_1 \boldsymbol{v}$: $\mathrm{k}^2 \boldsymbol{f}_1^2(1-\boldsymbol{v}^2), k\boldsymbol{f}_1(\boldsymbol{x}-\boldsymbol{v}t), kk_1\boldsymbol{f}_1(\boldsymbol{v}_1-\boldsymbol{v}), k^2\boldsymbol{f}_1(v-v)$. The fourth invariant is obviously null in this form.

30 Poincaré's force invariants (4) are functions of the following six intermediate invariants: $m = k(X_1\delta x + Y_1\delta y + Z_1\delta z - T_1\delta t)$, $n = k(X_1\delta_1 x + Y_1\delta_1 y + Z_1\delta_1 z - T_1\delta_1 t)$, $o = k(X_1 x + Y_1 y + Z_1 z - T_1 t)$, $p = k^2(X_1^2 + Y_1^2 + Z_1^2 - T_1^2)$, $q = \delta x^2 + \delta y^2 + \delta z^2 - \delta t^2$, and $s = \delta_1 x^2 + \partial_1 y^2 + \partial_1 z^2 - \partial_1 t^2$. Let the four force invariants (4) be denoted by $M, N, P,$ and S, then $M = p, N = o, P = n/\sqrt{-s}$, and $S = m/\sqrt{-q}$.
The same force invariants (4) are easily calculated using 4-vectors. Recalling the definitions in notes 23 and 29, let $\mathfrak{R} = (\boldsymbol{x}, it), U = k(\boldsymbol{v}, i), U_1 = k_1(\boldsymbol{v}_1, i)$, and $F_1 = k(\boldsymbol{f}_1, i\boldsymbol{f}_1\boldsymbol{v})$, where $\sqrt{-1} = i$. Then the force invariants (4) may be expressed as scalar products of 4-vectors: $M = F_1F_1$, $N = F_1\mathfrak{R}$, $P = F_1U_1$, and $S = F_1U$.

31 The kinematic invariants (1) obtained by Poincaré differ from those obtained from the products of 4-position and 4-velocity, contrary to Zahar's account (Zahar 1989, 194). Recalling the 4-vectors $\mathfrak{R}, U, U_1$, from n. 30, we form the products: $\mathfrak{R}\mathfrak{R}, \mathfrak{R}U, \mathfrak{R}U_1$, and UU_1. In Poincaré's notation, the latter four products are expressed as follows:

$$\Sigma x^2 - t^2, -\frac{t - \Sigma x\xi}{\sqrt{1-\Sigma\xi^2}}, -\frac{t-\Sigma x\xi_1}{\sqrt{1-\Sigma\xi_1^2}}, -\frac{t - \Sigma\xi\xi_1}{\sqrt{(1-\Sigma\xi^2)(1-\Sigma\xi_1^2)}}.$$

These invariants differ from those of Poincaré (1) only by the sign of $A, B,$ and C, as noted by (Sommerfeld 1910b, 686).

these same invariants (corresponding to modern 4-position, 4-velocity, 4-force-density and 4-force vectors), although in the end he expressed his laws of gravitation in terms of 4-force components.

To find a law applicable to the general case of two bodies in relative motion, Poincaré introduced constraints and approximations designed to reduce the complexity of his seven invariants and recover the form of the Newtonian law in the limit of slow motion ($\xi_1 \ll 1$). Poincaré naturally looked first to the velocity of propagation of gravitation. He briefly considered an emission theory, where the velocity of gravitation depends on the velocity of the source. Although the emission hypothesis was compatible with his invariants, Poincaré rejected this option because it violated his initial injunction barring a hyperlight velocity of gravitational propagation.[32] That left him with a propagation velocity of gravitation less than or equal to that of light, and to simplify his invariants Poincaré set it equal to that of light in empty space, such that $t = -\sqrt{\sum x^2} = -r$. This stipulation reduced the total number of invariants from seven to six.

With the propagation velocity of gravitation decided, Poincaré proceeded to construct a force law for point masses. He tried two approaches, the first of which is the most general. The basic idea of both approaches is to neglect terms in the square of velocity occurring in the invariants, and to compare the resulting approximations with their Newtonian counterparts. In the Newtonian scheme, the coordinates of the active mass point differ from those in the relativistic scheme (cf. note 18); Poincaré took the former to be $(x_0 + x_1, y_0 + y_1, z_0 + z_1)$ at the instant of time t_0, where the subscript 0 corresponds to the position of the passive mass point, and the coordinates with subscript 1 are found by assuming uniform motion of the source:

$$x = x_1 - \xi_1 r,\ \ y = y_1 - \eta_1 r,\ \ z = z_1 - \zeta_1 r,\ \ r = r_1 - \sum x\xi_1. \tag{5}$$

In the first approach, Poincaré made use of both the kinematic and force invariants. Substituting the values (2) into the kinematic invariants A, B, and C from (1) and the force invariants M, N, and P from (33), neglecting terms in the square of velocity, Poincaré obtained their sought-after Newtonian counterparts. Replacing the force vector occurring in the transformed force invariants by Newton's law ($\sum X_1 = -1/r_1^2$), and rearranging, Poincaré obtained three quantities in terms of distance and velocity.[33] He then re-expressed these quantities in terms of two of his original kinematic invariants, A and B, and equated the three resulting kinematic invariants to their corresponding original force invariants (4). He now had the solution in hand; three expressions relate his force invariants (containing the force vector $\sum X_1$) to two of his kinematic invariants:

32 An emission theory was proposed a few years later by Walter Ritz; see (Ritz 1908).

33 Using (5), Poincaré found the transformed force invariants $1/r_1^4, -1/r_1 - \Sigma x_1(\xi - \xi_1)/r_1^2$, and $\Sigma x_1(\xi - \xi_1)/r_1^3$.

$$M = \frac{1}{B^4}, \quad N = \frac{+A}{B^2}, \quad P = \frac{A-B}{B^3}. \tag{6}$$

He noted that complementary terms could be entertained for the three relations (6), provided that they were certain functions of his kinematic invariants $A, B,$ and C. Then without warning, he cut short his demonstration, remarking that the gravitational force components would take on imaginary values:

> The solution (6) appears at first to be the simplest, nonetheless, it may not be adopted. In fact, since M, N, P are functions of X_1, Y_1, Z_1, and $T_1 = \sum X_1 \xi$, the values of X_1, Y_1, Z_1 can be drawn from these three equations (6), but in certain cases these values would become imaginary.[34]

The quoted remark seems to suggest that for selected values of the particle velocities, the force turns out to be imaginary. However, the real difficulty springs from the equation $M = 1/B^4$, which allows for a repulsive force. The general approach failed to deliver.[35]

The fact that Poincaré published the preceding derivation may be understood in one of two ways. On the one hand, there is a psychological explanation: Poincaré's habit, much deplored by his peers, was to present his findings more or less in the order in which he found them. The case at hand may be no different from the others. On the other hand, Poincaré may have felt it worthwhile to show that the general approach breaks down. From the latter point of view, Poincaré's result is a positive one.

For his second attack on the law of gravitation, Poincaré adopted a less general approach. He knew where his first approach had become unsuitable, and consequently, leaving aside his three force invariants, he fell back on the form of his basic force 4-vector, which he now wrote in terms of his kinematic invariants, re-expressed in terms of $r = -t$, $k_0 = 1/\sqrt{1-\xi^2}$, and $k_1 = 1/\sqrt{1-\xi_1^2}$.[36] He assumed the gravitational force on the passive mass (moving with velocity ξ, η, ζ) to be a function of the distance separating the two mass points, the velocity of the passive mass point, and the velocity of the source, with the form:

34 "Au premier abord, la solution (6) paraît la plus simple, elle ne peut néanmoins être adoptée; en effet, comme M, N, P sont des fonctions de X_1, Y_1, Z_1, et de $T_1 = \Sigma X_1 \xi$, on peut tirer de ces trois équations (6) les valeurs de X_1, Y_1, Z_1; mais dans certains cas ces valeurs deviendraient imaginaires" (Poincaré 1906, 172).

35 Replacing A and B in (6) by their definitions results in the three equations: $M = k^2 \boldsymbol{f}_1^2(1-\boldsymbol{v}^2) = 1/k^4(r+\boldsymbol{x}\boldsymbol{v}_1)^4$, $N = \boldsymbol{f}_1(\boldsymbol{x}+\boldsymbol{v}r) = -(r+\boldsymbol{x}\boldsymbol{v})/[k_1^2(r+\boldsymbol{x}\boldsymbol{v}_1)^2]$, $P = kk_1\boldsymbol{f}_1(\boldsymbol{v}_1-\boldsymbol{v}) = [k(r+\boldsymbol{x}\boldsymbol{v})- -k_1(r+\boldsymbol{x}\boldsymbol{v}_1)]/k_1^3(r+\boldsymbol{x}\boldsymbol{v}_1)^3]$. Equations N and P imply an attractive force for all values of $\boldsymbol{v}$ and $\boldsymbol{v}_1$, while M leads to the ambiguously-signed solution: $\boldsymbol{f}_1 = \pm(1/[k^2(r+\boldsymbol{x}\boldsymbol{v}_1)^2])$. Presumably, the superfluous plus sign in (6) is an indication of Poincaré's preoccupation with obtaining a force of correct sign.

36 $A = -k_0(r+\Sigma x\xi)$, $B = -k_1(r+\Sigma x\xi_1)$, and $C = k_0 k_1(1-\Sigma x\xi\xi_1)$.

$$X_1 = x\frac{\alpha}{k_0} + \xi\beta + \xi_1\frac{k_1}{k_0}\gamma, \quad Z_1 = z\frac{\alpha}{k_0} + \zeta\beta + \zeta_1\frac{k_1}{k_0}\gamma,$$
$$Y_1 = y\frac{\alpha}{k_0} + \eta\beta + \eta_1\frac{k_1}{k_0}\gamma, \quad T_1 = -r\frac{\alpha}{k_0} + \beta + \frac{k_1}{k_0}\gamma, \tag{7}$$

where $\alpha, \beta,$ and γ denote functions of the kinematic invariants.[37] By definition, the component T_1 is the scalar product of the ordinary force and the velocity of the passive mass point, $T_1 = \sum X_1\xi$, such that the three functions α, β, γ satisfy the equation:

$$-A\alpha - \beta - C\gamma = 0. \tag{8}$$

Poincaré further assumed $\beta = 0$, thereby eliminating a term depending on the velocity of the passive mass, and fixing the value of γ in terms of α. Applying the same slow-motion approximation and translation (5) as in his initial approach, Poincaré found $X_1 = \alpha x_1$, and by comparison with Newton's law, α reduces to $-1/r_1^3$. In terms of the kinematic invariants (1), this relation was expressed as $\alpha = 1/B^3$, and the law of gravitation (7) took on the form:[38]

$$X_1 = \frac{x}{k_0B^3} - \xi_1\frac{k_1}{k_0}\frac{A}{B^3C}, \quad Z_1 = \frac{z}{k_0B^3} - \zeta_1\frac{k_1}{k_0}\frac{A}{B^3C},$$
$$Y_1 = \frac{y}{k_0B^3} - \eta_1\frac{k_1}{k_0}\frac{A}{B^3C}, \quad T_1 = \frac{r}{k_0B^3} - \frac{k_1}{k_0}\frac{A}{B^3C}. \tag{9}$$

Inspection of Poincaré's gravitational force (9) reveals two components: one parallel to the position 4-vector between the passive mass and the retarded source, and one parallel to the source 4-velocity. The law was not unique, Poincaré noted, and it neglected possible terms in the velocity of the passive mass.

Poincaré underlined the open-ended nature of his solution by proposing a second gravitational force law. Rearranging (9) and replacing the factor $1/B^3$ by C/B^3, such that the force depended linearly on the velocity of the passive mass, Poincaré arrived at a second law of gravitation:[39]

---

37 Using modern 4-vector notation, and denoting Poincaré's gravitational force 4-vector $F_1 = k_0(X_1, Y_1, Z_1, iT_1)$, equation (7) may be expressed: $F_1 = \alpha\mathfrak{R} + \beta U + \gamma U_1$, where $\mathfrak{R}$ denotes a light-like 4-vector between the mass points, α, β, γ stand for undetermined functions of the three kinematic invariants A, B, and C, while $U = k_0(\boldsymbol{v}, i), U_1 = k_1(\boldsymbol{v}_1, i)$ designate the 4-velocities of the passive and active mass points, respectively.

38 In ordinary vector form, recalling the definitions in notes 23 and 29, the spatial part of Poincaré's law is expressed as follows: $\boldsymbol{f}_1 = -[(\boldsymbol{x} + r\boldsymbol{v}_1) + \boldsymbol{v} \times (\boldsymbol{v}_1 \times \boldsymbol{x})]/[kk_1^3(r + \boldsymbol{x}\boldsymbol{v}_1)^3(1 - \boldsymbol{v}\boldsymbol{v}_1)]$. Cf. (Zahar 1989, 199).

39 This law may be reformulated using the vectors defined in notes 23 and 29, and neglecting (with Poincaré) the component T_1: $\boldsymbol{f}_1 = -([(\boldsymbol{x} + r\boldsymbol{v}_1) + \boldsymbol{v} \times (\boldsymbol{v}_1 \times \boldsymbol{x})]/[k_1^2(r + \boldsymbol{x}\boldsymbol{v}_1)^3])$. Cf. (Zahar 1989, 199). Comparable expressions were developed by Lorentz and Kottler (Lorentz 1910, 1239; Kottler 1922, 169).

$$
\begin{aligned}
X_1 &= \frac{\lambda}{B^3} + \frac{\eta\nu' - \zeta\mu'}{B^3}, \\
Y_1 &= \frac{\mu}{B^3} + \frac{\zeta\lambda' - \xi\nu'}{B^3}, \\
Z_1 &= \frac{\nu}{B^3} + \frac{\xi\mu' - \eta\lambda'}{B^3},
\end{aligned}
\tag{10}
$$

where

$$
\begin{aligned}
k_1(x + r\xi_1) &= \lambda, \qquad k_1(y + r\eta_1) = \mu, \qquad k_1(z + r\zeta_1) = \nu, \\
k_1(\eta_1 z - \zeta_1 y) &= \lambda', \quad k_1(\zeta_1 x - \xi_1 z) = \mu', \quad k_1(\xi_1 y - x\eta_1) = \nu'.
\end{aligned}
$$

Poincaré neglected to write down the expression for T_1, probably because of its complicated form. (For the sake of simplicity, I refer to [9] and [10], including the latter's neglected fourth component, as Poincaré's first and second law.) The unprimed triplet $B^{-3}(\lambda, \mu, \nu)$ supports what Poincaré termed a "vague analogy" with the mechanical force on a charged particle due to an electric field, while the primed triplet $B^{-3}(\lambda', \mu', \nu')$ supports an analogy to the mechanical force on a charged particle due to a magnetic field. He identified the fields as follows:

> Now λ, μ, ν, or $\frac{\lambda}{B^3}, \frac{\mu}{B^3}, \frac{\nu}{B^3}$, is an electric field of sorts, while λ', μ', ν', or rather $\frac{\lambda'}{B^3}, \frac{\mu'}{B^3}, \frac{\nu'}{B^3}$, is a magnetic field of sorts.[40]

While Poincaré wrote freely of a "gravity wave" [*onde gravifique*], he abstained from speculating on the nature of the field referred to here. As one of the first theorists (with FitzGerald and Lorentz) to have employed retarded potentials in Maxwellian electrodynamics, Poincaré must have considered the possibility of introducing a corresponding gravitational 4-potential (Whittaker 1951–1953, 1: 394, n. 3).[41] But as matters stood when Poincaré submitted this paper for publication in July, 1905, he was not in a position to elaborate the physics of fields in four-dimensional terms, since he possessed neither a 4-potential nor a 6-vector.

Poincaré had realized the objective of formulating a Lorentz-invariant force of gravitation. As we have seen, he surpassed this objective by identifying not one but

---

40 "Alors λ, μ, ν, ou $\lambda/B^3, \mu/B^3, \nu/B^3$, est une espèce de champ électrique, tandis que λ', μ', ν', ou plutôt $\lambda'/B^3, \mu'/B^3, \nu'/B^3$, est une espèce de champ magnétique" (Poincaré 1906, 175).

41 A 4-potential corresponding to Poincaré's second law (10) was given by Kottler (Kottler 1922, 169). Additional assumptions are required in order to identify a "gravito-magnetic" field with a term arising from the Lorentz transformation of force: $\boldsymbol{v} \times (\boldsymbol{v}_1 \times \boldsymbol{x})$, or the second term of the 3-vector version of (10) (neglecting the global factor; see note 39). In particular, it must be assumed that when the sources of the "gravito-electric" field $B^{-3}(\lambda, \mu, \nu)$ are at rest, the force on a mass point m is $\boldsymbol{f} = mB^{-3}(\lambda, \mu, \nu)$, independent of the velocity of m. For a detailed discussion, see (Jackson 1975, 578).

two such force laws. Designed to reduce to Newton's law in the first order of approximation in ξ_1 (or particle velocity divided by the speed of light), Poincaré's laws could diverge from Newton's only in second-order terms. The argument satisfied Poincaré, who did not report any precise numerical results, explaining that this would require further investigation. Instead, he noted that the disagreement would be ten thousand times smaller than a first-order difference stemming from the assumption of a propagation velocity of gravitation equal to that of light, "*ceteris non mutatis*" (p. 175). His result contradicted Laplace, who had predicted an observable first-order effect arising from just such an assumption. At the very least, Poincaré had demonstrated that Laplace's argument was not compelling in the context of the new dynamics.[42]

On several occasions over the next seven years, Poincaré returned to the question of gravitation and relativity, without ever comparing the predictions of his laws with observation. During his 1906–1907 Sorbonne lectures, for example, when he developed a general formula for perihelion advance, Poincaré used a Lagrangian approach, rather than one or the other of his laws (Poincaré 1953, 238). Student notes of this course indicate that he stopped short of a numerical evaluation for the various electron models (perhaps leaving this as an exercise). However, Poincaré later provided the relevant numbers in a general review of electron theory. Lorentz's theory called for an extra 7" centennial advance by Mercury's perihelion, a figure slightly greater than the one for Abraham's non-relativistic electron theory.[43] According to the best available data, Mercury's anomalous perihelial advance was 42", prompting Poincaré to remark that another explanation would have to be found in order to account for the remaining seconds of arc. Astronomical observations, Poincaré concluded soberly, provided no arguments in favor of the new electron dynamics (Poincaré 1908, 400).[44]

Poincaré capsulized the situation of his new theory in a fable in which Lorentz plays the role of Ptolemy, and Poincaré that of an unknown astronomer appearing sometime between Ptolemy and Copernicus. The unknown astronomer notices that all the planets traverse either an epicycle or a deferent in the same lapse of time, a regularity later captured in Kepler's second law. The analogy to electron dynamics turns on a regularity discovered by Poincaré in his study of gravitation:

> If we were to admit the postulate of relativity, we would find the same number in the law of gravitation and the laws of electromagnetism, which would be the velocity of light; and we would find it again in all the other forces of any origin whatsoever.[45]

---

42 Poincaré reviewed Laplace's argument in his 1906–1907 lectures (Poincaré 1953, 194). For a contemporary overview of the question of the propagation velocity of gravitation see (Tisserand 1889–1896, 1: 511).

43 Fritz Wacker, a student of Richard Gans in Tübingen, published similar results in (Wacker 1906).

44 Poincaré explained to his students that Mercury's anomalous advance could plausibly be attributed to an intra-Mercurial matter belt (Poincaré 1953, 265), an idea advanced forcefully by Hugo von Seeliger in 1906 (Roseveare 1982, 78). In a lecture delivered in September, 1909, Poincaré revised his estimate of the relativistic perihelial advance downward slightly to 6" (Poincaré 1909).

45 "[S]i nous admettions le postulat de relativité, nous trouverions dans la loi de gravitation et dans les lois électromagnétiques un nombre commun qui serait la vitesse de la lumière; et nous le retrouverions encore dans toutes les autres forces d'origine quelconque" (Poincaré 1906, 131).

This common propagation velocity of gravitational action, of electromagnetic fields, and of any other force, could be understood in one of two ways:

> Either everything in the universe would be of electromagnetic origin, or this aspect—shared, as it were, by all physical phenomena—would be a mere epiphenomenon, something due to our methods of measurement.[46]

If the electromagnetic worldview were valid, all particle interactions would be governed by Maxwell's equations, featuring a constant propagation velocity. Otherwise, the common propagation velocity of forces had to be a result of a measurement convention. In relativity theory, as Poincaré went on to point out, the measurement convention to adopt was one defining lengths as equal if and only if spanned by a light signal in the same lapse of time, as this convention was compatible with the Lorentz contraction. There was a choice to be made between the electromagnetic worldview (as realized in the electron models of Abraham and Bucherer-Langevin) and the postulate of relativity (as upheld by the Lorentz-Poincaré electron theory). Although Poincaré favored the latter theory, he felt that its destiny was to be superseded, just as Ptolemaic astronomy was superseded by Copernican heliocentrism.

The failure of his Lorentz-invariant law of gravitation to explain the anomalous advance of Mercury's perihelion probably fed Poincaré's dissatisfaction with the Lorentz-Poincaré theory in general, but what he found particularly troubling at the time was something else altogether: the discovery of magneto-cathode rays. There is no place in the Lorentz-Poincaré electron theory for rays that are both neutral (as Paul Villard reported in June, 1904) and deflected by electric and magnetic fields, which is probably why Poincaré felt the "entire theory" to be "endangered" by magneto-cathode rays.[47]

Uncertainty over the empirical adequacy of the Lorentz-Poincaré electron theory may explain why the *Rendiconti* memoir was Poincaré's last in the field of electron physics. But is it enough to explain his disinterest in the development of a four-dimensional formalism? One year after the publication of his article on electron dynamics, Poincaré commented:

> A translation of our physics into the language of four-dimensional geometry does in fact appear to be possible; the pursuit of this translation would entail great pain for limited profit, and I will just cite Hertz's mechanics, where we see something analogous. Meanwhile, it seems that the translation would remain less simple than the text and would always have the feel of a translation, and that three-dimensional language seems the best suited to the description of our world, even if one admits that this description may be carried out in another idiom.[48]

---

46 "Ou bien il n'y aurait rien au monde qui ne fût d'origine électromagnétique. Ou bien cette partie qui serait pour ainsi dire commune à tous les phénomènes physiques ne serait qu'une apparence, quelque chose qui tiendrait à nos méthodes de mesure" (Poincaré 1906, 131–132).

47 See (Poincaré 1906, 132; Stein 1987, 397, n. 29). On the history of magneto-cathode rays, see (Carazza and Kragh 1990).

Poincaré clearly saw in his own work the outline of a four-dimensional formalism for physics, yet he saw no future in its development, and this, entirely apart from the question of the empirical adequacy of the Lorentz-Poincaré theory.

Why did Poincaré discount the value of a language tailor-made for relativity? Three sources of disinterest in such a prospect spring to mind, the first of which stems from his conventionalist philosophy of science. Poincaré recognized an important role for notation in the exact sciences, as he famously remarked with respect to Edmond Laguerre's work on quadratic forms and Abelian functions that

> in the mathematical sciences, having the right notation is philosophically as important as having the right classification in the life sciences.[49]

More than likely, Poincaré was aware of the philosophical implications of a four-dimensional notation for physics, although he had yet to make his views public. But given his strong belief in the immanence of Euclidean geometry's fitness for physics, he must have considered the chances for success of such a language to be vanishingly small.[50]

A second source for Poincaré's disinterest in four-dimensional formalism is his practice of physics. As mentioned above, Poincaré dispensed with vectorial systems (and most notational shortcuts); he even avoided writing i for $\sqrt{-1}$. When considered in conjunction with his conventionalist belief in the suitability of Euclidean geometry for physics, this conservative habit with respect to notation makes Poincaré appear all the less likely to embrace a four-dimensional language for physics.

The third possible source of discontent is Poincaré's vexing experience with invariants of pseudo-Euclidean 4-space. As shown above (p. 205), Poincaré's first approach to the construction of a law of gravitation ended unsatisfactorily, and the failure of Poincaré's intuition in this instance may well have colored his view of the prospects for a four-dimensional physics.

An immediate consequence of Poincaré's refusal to work out the form of four-dimensional physics was that others could readily pick up where he left off. Roberto Marcolongo (1862–1945), Professor of Mathematical Physics in Messina, and a leading proponent of vectorial analysis, quickly discerned in Poincaré's paper a potential

---

48 "Il semble bien en effet qu'il serait possible de traduire notre physique dans le langage de la géométrie à quatre dimensions; tenter cette traduction ce serait se donner beaucoup de mal pour peu de profit, et je me bornerai à citer la mécanique de Hertz où l'on voit quelque chose d'analogue. Cependant, il semble que la traduction serait toujours moins simple que le texte, et qu'elle aurait toujours l'air d'une traduction, que la langue des trois dimensions semble la mieux appropriée à la description de notre monde, encore que cette description puisse se faire à la rigueur dans un autre idiome" (Poincaré 1907, 15). See also (Walter 1999b, 98), and for a different translation, (Galison 1979, 95). On Hertz's mechanics, see (Lützen 1999).

49 "[D]ans les Sciences mathématiques, une bonne notation a la même importance philosophique qu'une bonne classification dans les Sciences naturelles" (Poincaré 1898–1905, 1:x).

50 Poincaré's analysis of the concepts of space and time in relativity theory appeared in 1912 (Poincaré 1912). On the cool reception among mathematicians of Poincaré's views on physical geometry, see (Walter 1997).

for formal development. Marcolongo referred, like Poincaré, to a four-dimensional space with one imaginary axis, but defined the fourth coordinate as the product of time t and the negative square root of -1 (i.e., $-t\sqrt{-1}$ instead of $t\sqrt{-1}$). After forming a 4-vector potential out of the ordinary vector and scalar potentials, and defining a 4-current vector, he expressed the Lorentz-covariance of the equations of electrodynamics in matrix form. No other applications were forthcoming from Marcolongo, and a failure to produce further 4-vector quantities and functions limited the scope of his contribution, which went unnoticed outside of Italy (Marcolongo 1906).[51] Nothing further on Poincaré's method appeared in print until April, 1908, when Hermann Minkowski's paper on the four-dimensional formalism and its application to the problem of gravitation appeared in the *Göttinger Nachrichten*.

2. HERMANN MINKOWSKI'S SPACETIME LAWS OF GRAVITATION

The young Hermann Minkowski, second son of an immigrant family of Russian Jews, attended the Altstädtische Gymnasium in Königsberg (later Kaliningrad). Shortly after graduation, Minkowski submitted an essay for the Paris Academy's 1882 Grand Prize in Mathematical Sciences. His entry on quadratic forms shared top honors with a submission by the seasoned British mathematician Henry J. S. Smith, his senior by thirty-eight years.[52] The young mathematician went on to study with Heinrich Weber in Königsberg, and with Karl Weierstrass and Leopold Kronecker in Berlin. In the years following the prize competition, Minkowski became acquainted with Poincaré's writings on algebraic number theory and quadratic forms, and in particular, with a paper in Crelle's *Journal* containing some of the results from Minkowski's prize paper, still in press. To his friend David Hilbert he confided the "angst and alarm" brought on by Poincaré's entry into his field of predilection; with his "swift and versatile" energy, Poincaré was bound to bring the whole field to closure, or so it seemed to him at the time.[53] From the earliest, formative years of his scientific career, Minkowski found in Poincaré—his senior by a decade—a daunting intellectual rival.

While Minkowski had discovered in Poincaré a rival, he was soon to find that the Frenchman could also be a teacher, from whom he could learn new analytical skills and methods. Named Privatdozent in Bonn in 1887, Minkowski contributed to the abstract journal *Jahrbuch der Fortschritte der Mathematik*, and in 1892, took on the considerable task of summarizing the results of the paper for which Poincaré was

---

51 This paper later gave rise to a priority claim for a slightly different substitution: $u = it$ (Marcolongo to Arnold Sommerfeld, 5 May, 1913, Archives for History of Quantum Physics 32). On Marcolongo's paper see also (Maltese 2000, 135).

52 See (Rüdenberg 1973; Serre 1993; Strobl 1985).

53 Minkowski to Hilbert, 14 February, 1885, (Minkowski 1973, 30). Minkowski's fears turned out to be for naught, as Poincaré pursued a different line of research (Zassenhaus 1975, 446). On Minkowski's early work on the geometry of numbers see (Schwermer 1991); on later developments, see (Krätzel 1989).

awarded the King Oscar II Prize (Minkowski 1890–1893). The mathematics Poincaré created in his prize paper (the study of homoclinic points in particular) was highly innovative, and at the same time, difficult to follow. Among those whom we know had trouble understanding certain points of Poincaré's prize memoir were Charles Hermite, Gustav Mittag-Leffler, and Karl Weierstrass, who happened to constitute the prize committee.[54] Minkowski, however, welcomed the review as a learning opportunity, as he wrote to his friend and former teacher, Adolf Hurwitz:

> Poincaré's prize paper is also among the works I have to report on for the *Fortschritte*. I am quite fond of it. It is a fine opportunity for me to get acquainted with problems I have not worried about too much up to now, since I will naturally set a positive goal of making my case well.[55]

In the 1890s, building on his investigations of the algebraic theory of quadratic forms, Minkowski developed the geometric analog to this theory: geometrical number theory. A high point of his efforts in this new field, and one which contributed strongly to the establishment of his reputation in mathematical circles, was the publication of *Geometrie der Zahlen* (Minkowski 1896). The same year, Minkowski accepted a chair at Zurich Polytechnic, whereby he rejoined Hurwitz. Minkowski's lectures on mathematics and mathematical physics attracted a small following of talented and ambitious students, including the future physicists Walter Ritz and Albert Einstein, and the budding mathematicians Marcel Grossmann and Louis Kollros.[56]

Minkowski's lectures on mechanics in Zurich throw an interesting light on his view of symbolic methods in physics at the close of the nineteenth century. The theory of quaternions, he noted in 1897, was used nowhere outside of England, due to its "relatively abstract character and inherent difficulty."[57] Two of its fundamental concepts, scalars and vectors, had nevertheless gained broad approval among physicists, Minkowski wrote, and had found "frequent application especially in the theory of electricity."[58] Applications of quaternions to problems of physics were advanced in Germany with the publication of Felix Klein and Arnold Sommerfeld's *Theorie des Kreisels*, a work referred to in Minkowski's lecture notes of 1898–1899 (Klein and Sommerfeld 1879–1910).[59] Minkowski admired Klein and Sommerfeld's text,

---

54 See (Gray 1992) and the reception study by Barrow-Green (Barrow-Green 1997, chap. 6).

55 Minkowski to Hurwitz, 5 January, 1892, Cod. Ms. Math. Arch. 78: 188, Handschriftenabteilung, Niedersächsische Staats- und Universitätsbibliothek (NSUB). On Minkowski's report see also (Barrow-Green 1997, 143).

56 Minkowski papers, Arc. 4° 1712, Jewish National and University Library (JNUL); Minkowski to Hilbert, 11 March, 1901, (Minkowski 1973, 139).

57 Vorlesungen über analytische Mechanik, Wintersemester 1897/98, p. 29, Minkowski papers, Arc. 4° 1712, JNUL.

58 Loc. cit. note 57. The concepts of scalar and vector mentioned by Minkowski were those introduced by W. R. Hamilton (1805–1865), the founder of quaternion theory. Even in Britain, vectors were judged superior to quaternions for use in physics, giving rise to spirited exchanges in the pages of *Nature* during the 1890s, as noted by (Bork 1966) and (Crowe 1967, chap. 6). On the introduction of vector analysis as a standard tool of the physicist during this period, see (Jungnickel and McCormmach 1986, 2:342), and for a general history, see (Crowe 1967).

expressing "great interest" in the latter to Sommerfeld, along with his approval of the fundamental significance accorded to the concept of momentum. However, their text did not make the required reading list for Minkowski's course in mechanics.[60]

In 1899, at the request of Sommerfeld, who a year earlier had agreed to edit the physics volumes of Felix Klein's ambitious *Encyclopedia of the Mathematical Sciences including Applications* (hereafter *Encyklopädie*), Minkowski agreed to cover a topic in molecular physics he knew little about, but one perfectly suited to his skills as an analyst: capillarity.[61] The article that appeared seven years later represented his second contribution to physics, after a short note on theoretical hydrodynamics published in 1888, but which, ten years later, Minkowski claimed no one had read—save the abstracter (Minkowski 1888, 1907).[62]

When Minkowski accepted Göttingen's newly-created third chair of pure mathematics in the fall of 1902, the pace of his research changed brusquely. The University of Göttingen at the turn of the last century was a magnet for talented young mathematicians and physicists.[63] Minkowski soon was immersed in the activities of Göttingen's Royal Society of Science, its mathematical society, and research seminars. Several faculty members, including Max Abraham, Gustav Herglotz, Eduard Riecke, Karl Schwarzschild, and Emil Wiechert, actively pursued theoretical or experimental investigations motivated by the theory of electrons, and it was not long before Minkowski, too, took up the theory. During the summer semester of 1905 he co-led a seminar with Hilbert on electron theory, featuring reports by Wiechert and Herglotz, and by Max Laue, who had just finished a doctoral thesis under Max Planck's supervision.[64]

Along with seminars on advanced topics in physics and analytical mechanics, Göttingen featured a lively mathematical society, with weekly meetings devoted to presentations of work-in-progress and reports on scientific activity outside of Göttingen. The electron theory was a frequent topic of discussion in this venue. For instance, the problem of gravitational attraction was first addressed by Schwarzschild in December, 1904, in a report on Alexander Wilken's recent paper on the compatibility of Lorentz's electron theory with astronomical observations.[65]

---

59 Vorlesungen über Mechanik, Wintersemester 1898/99, 47, 59, Minkowski papers, Arc. 4° 1712, JNUL. Minkowski referred to Klein and Sommerfeld's text in relation to the concept of force and its anthropomorphic origins, the kinetic theory of gas, and the theory of elasticity.

60 Minkowski to Sommerfeld, 30 October, 1898, MSS 1013A, Special Collections, National Museum of American History. An extensive reading list of mechanics texts is found in Minkowski's course notes for the 1903–1904 winter semester, Mechanik I, 9, Minkowski papers, Arc. 4° 1712, JNUL.

61 Minkowski to Sommerfeld, 30 October, 1898, loc. cit. note 60; Minkowski to Sommerfeld, 18 November, 1899, Nachlass Sommerfeld, Arch HS1977-28/A, 233, Deutsches Museum München; research notebook, 12 December, 1899, Arc. 4° 1712, Minkowski papers, JNUL.

62 Minkowski to Sommerfeld, 30 October, 1898, loc. cit. note 60.

63 On Göttingen's rise to preeminence in these fields, see (Manegold 1970; Pyenson 1985, chap. 7; Rowe 1989, 1992).

64 Nachlass Hilbert 570/9, Handschriftenabteilung, NSUB; (Pyenson 1985, chap. 5).

65 *Jahresbericht der deutschen Mathematiker-Vereinigung* 14, 61.

A focal point of sorts for the mathematical society, Poincaré's scientific output fascinated Göttingen scientists in general, and Minkowski in particular, as mentioned above.[66] Minkowski reported to the mathematical society on Poincaré's publications on topology, automorphic functions, and capillarity, devoting three talks in 1905–1906 to Poincaré's 1888–1889 Sorbonne lectures on this subject (Poincaré 1895). Others reporting on Poincaré's work were Conrad Müller on Poincaré's St. Louis lecture on the current state and future of mathematical physics (31 January, 1905), Hugo Broggi on probability (27 October, 1905), Ernst Zermelo on a boundary-value problem (12 December, 1905), Erhard Schmidt on the theory of differential equations (19 December, 1905), Max Abraham on the Sorbonne lectures (6 February, 1906) and Paul Koebe on the uniformization theorem (19 November, 1907). One gathers from this list that the Göttingen mathematical society paid attention to Poincaré's contributions to celestial mechanics, mathematical physics, and pure mathematics, all subjects intersecting with the ongoing research of its members. It also appears that no other member of the mathematical society was quite as assiduous in this respect as Minkowski.[67]

When Einstein's relativity paper appeared in late September, 1905, it drew the attention of the Bonn experimentalist Walter Kaufmann, a former Göttingen Privatdozent and friend of Max Abraham, but neither Abraham nor any of his colleagues rushed to report on the new ideas to the mathematical society.[68] Poincaré's long memoir on the dynamics of the electron, published in January, 1906, fared better, although nearly two years went by before Minkowski found an occasion to comment on Poincaré's gravitation theory, and to present his own related work-in-progress. Minkowski's typescript has been conserved, and is the source referred to here.[69]

On the occasion of the 5 November, 1907, meeting of the mathematical society, Minkowski began his review of Poincaré's work by observing that gravitation remained an "important question" in relativity theory, since it was not yet known "how the law of gravitation is arranged for in the realm of the principle of relativity."[70] The basic problem of gravitation and relativity, in other words, had not been solved by Poincaré. Eliding mention of Poincaré's two laws, Minkowski recognized

---

66 Although Poincaré spoke on celestial mechanics in Göttingen in 1895 (Rowe 1992, 475) and was invited back in 1902, he did not return until 1909, a few months after Minkowski's sudden death. See Hilbert to Poincaré, 6 November, 1908 (Dugac 1986, 209); Klein to Poincaré, 14 Jan., 1902 (Dugac 1989, 124–125). Sponsored by the Wolfskehl Fund, Poincaré's 1909 lecture series took place during "Poincaré week", in the month of April. His lectures were published the following year (Poincaré 1910) in a collection launched in 1907, based on an idea of Minkowski's (Klein 1907, IV).

67 *Jahresbericht der deutschen Mathematiker-Vereinigung* 14: 128, 586; 15: 154–155; 17: 5.

68 On Kaufmann's cathode-ray deflection experiments, see (Miller 1981, 226) and (Hon 1995). Readings of Kaufmann's articles are discussed at length by Richard Staley (Staley 1998, 270).

69 Undated typescript of a lecture on a new form of the equations of electrodynamics, *Math. Archiv* 60: 3, Handschriftenabteilung, NSUB. This typescript differs significantly from the posthumously-published version (Minkowski 1915).

70 "Es entsteht die grosse Frage, wie sich denn das Gravitationsgesetz in das Reich des Relativitätsprinzipes einordnen lässt" (p. 15).

in his work only one positive result: by considering gravitational attraction as a "pure mathematical problem," he said, Poincaré had found gravitation to propagate with the speed of light, thereby overturning the standard Laplacian argument to the contrary.[71]

Minkowski expressed dissatisfaction with Poincaré's approach, allowing that Poincaré's was "only one of many" possible laws, a fact stemming from its construction out of Lorentz-invariants. Consequently, Poincaré's investigation "had by no means a definitive character."[72] A critical remark of this sort often introduces an alternative theory, but in this instance none was forthcoming, and as I will show in what follows, there is ample reason to doubt that Minkowski was actually in a position to improve on Poincaré's investigation. Nonetheless, at the end of his talk Minkowski set forth the possibility of elaborating his report.

Minkowski's lecture was not devoted entirely to Poincaré's investigation of Lorentz-invariant gravitation. The purpose of his lecture, according to the published abstract, was to present a new form of the equations of electrodynamics leading to a mathematical redescription of physical laws in four areas: electricity, matter, mechanics, and gravitation.[73] These laws were to be expressed in terms of the differential equations used by Lorentz as the foundation of his successful theory of electrons (Lorentz 1904a), but in a form taking greater advantage of the invariance of the quadratic form $x^2 + y^2 + z^2 - c^2t^2$. Physical laws, Minkowski stated, were to be expressed with respect to a four-dimensional manifold, with coordinates x_1, x_2, x_3, x_4, where units were chosen such that $c = 1$, the ordinary Cartesian coordinates x, y, and z, went over into the first three, and the fourth was defined to be an imaginary time coordinate, $x_4 = it$. Implicitly, then, Minkowski took as his starting point the four-dimensional vector space described in the last section of Poincaré's memoir on the dynamics of the electron.

Minkowski acknowledged, albeit obliquely, a certain continuity between Poincaré's memoir and his own program to reform the laws of physics in four-dimensional terms. By formulating the electromagnetic field equations in four-dimensional notation, Minkowski said he was revealing a symmetry not realized by his predecessors, not even by Poincaré himself (Walter 1999b, 98). While Poincaré had not sought to modify the standard form of Maxwell's equations, Minkowski felt it was time for a change. The advantage of expressing Maxwell's equations in the new notation, Minkowski informed his Göttingen colleagues, was that they were then "easier to grasp" (p. 11).

His reformulation naturally began in the electromagnetic domain, with an expression for the potentials. He formed a 4-vector potential denoted (ψ) by taking the

---

71 Actually, Poincaré postulated the light-like propagation velocity of gravitation, as mentioned above, (p. 204).

72 "Poincaré weist ein solches Gesetz auf, indem er auf die Betrachtung von Invarianten der Lorentzschen Gruppe eingeht, doch ist das Gesetz nur eines unter vielen möglichen, und die betreffenden Untersuchungen tragen in keiner Weise einen definitiven Charakter" (p. 16). See also (Pyenson 1973, 233).

73 *Jahresbericht der deutschen Mathematiker-Vereinigung* 17 (1908), *Mitt. u. Nachr.*, 4–5.

ordinary vector potential over for the first three components, and setting the fourth component equal to the product of i and the scalar potential. The same method was applied to obtain a four-component quantity for current density: for the first three components, Minkowski took over the convection current density vector, $\rho\mathfrak{w}$, or charge density times velocity, and defined the fourth component to be the product of i and the charge density. Rewriting the potential and current density vectors in this way, Minkowski imposed what is now known as the Lorentz condition, $\mathrm{Div}(\psi) = 0$, where Div is an extension of ordinary divergence. This led him to the following expression, summarizing two of the four Maxwell equations:

$$\Box\psi_j = -\rho_j \quad (j = 1, 2, 3, 4), \tag{11}$$

where $\Box$ is the d'Alembertian, employed earlier by Poincaré (cf. note 20).

Of the formal innovations presented by Minkowski to the mathematical society, the most remarkable was what he called a *Traktor*, a six-component entity used to represent the electromagnetic field.[74] He defined the six components via the 4-vector potential, using a two-index notation: $\psi_{jk} = \partial\psi_k/\partial\psi_j - \partial\psi_j/\partial\psi_k$, noting the anti-symmetry relation $\rho_{kj} = -\rho_{jk}$, and zeros along the diagonal $\psi_{jj} = 0$. In this way, the Traktor components $\psi_{14}, \psi_{24}, \psi_{34}, \psi_{23}, \psi_{31}, \psi_{12}$ match up with the field quantities $-i\mathfrak{E}_x, -i\mathfrak{E}_y, -i\mathfrak{E}_z, \mathfrak{h}_x, \mathfrak{h}_y, \mathfrak{h}_z$.[75]

The Traktor first found application when Minkowski turned to his second topic: the four-dimensional view of matter. Ignoring the electron theories of matter of Lorentz and Joseph Larmor, Minkowski focused uniquely on the macroscopic electrodynamics of moving media.[76] For this subject he introduced a *Polarisationstraktor*, (p), along with a 4-current-density, (σ), defined by the current density vector $\boldsymbol{i}$ and the charge density ρ: $(\sigma) = i_x, i_y, i_z, i\rho$ (p. 9). Recalling (11), Minkowski wrote Maxwell's source equations in covariant form:

$$\frac{\partial p_{1j}}{\partial x_1} + \frac{\partial p_{2j}}{\partial x_2} + \frac{\partial p_{3j}}{\partial x_3} + \frac{\partial p_{4j}}{\partial x_4} = \sigma_j + \Box\psi_j. \tag{12}$$

Minkowski's relativistic extension of Maxwell's theory was all the simpler in that it elided the covariant expression of the constitutive equations, which involves 4-velocity.[77] While none of his formulas invoked 4-velocity, Minkowski acknowledged that his theory required a "velocity vector of matter $(w) = w_1, w_2, w_3, w_4$" (p. 10).

---

74 The same term was employed by Cayley to denote a line which meets any given lines, in a paper of 1869 (Cayley 1869).

75 When written out in full, one obtains, for example, $\psi_{23} = \partial\psi_3/\partial x_2 - \partial\psi_2/\partial x_2 = \mathfrak{h}_x$. Minkowski later renamed the Traktor a *Raum-Zeit-Vektor II. Art* (Minkowski 1908, §5) but it is better known as either a 6-vector, an antisymmetric 6-tensor, or an antisymmetric, second-rank tensor. As the suite of synonyms suggests, this object found frequent service in covariant formulations of electrodynamics.

76 For a comparison of the Lorentz and Larmor theories, see (Darrigol 1994).

77 On the four-dimensional transcription of Ohm's law see (Arzeliès and Henry 1959, 65–67).

In order to express the "visible velocity of matter in any location," Minkowski needed a new vector as a function of the coordinates x, y, z, t (p. 7). Had he understood Poincaré's 4-velocity definition (see above, p. 202), he undoubtedly would have employed it at this point. Instead, following the same method of generalization from three to four components successfully applied in the case of 4-vector potential, 4-current density, and 4-force density, Minkowski took over the components of the velocity vector $\mathfrak{w}$ for the spatial elements of the quadruplet designated w_1, w_2, w_3, w_4:

$$\mathfrak{w}_x, \quad \mathfrak{w}_y, \quad \mathfrak{w}_z, \quad i\sqrt{1-\mathfrak{w}^2}. \tag{13}$$

There are two curious aspects to Minkowski's definition. First of all, its squared magnitude does not vanish when ordinary velocity vanishes; even a particle at rest with respect to a reference frame is described in that frame by a 4-velocity vector of nonzero length. This is also true of Poincaré's 4-velocity definition, and is a feature of relativistic kinematics. Secondly, the components of Minkowski's quadruplet do not transform like the coordinates x_1, x_2, x_3, x_4, and consequently lack what he knew to be an essential property of a 4-vector.[78]

The most likely source for Minkowski's blunder is Poincaré's paper. We recall that Poincaré's derivation of his kinematic invariants ignored 4-vectors (see above, p. 202), and what is more, his paper features a misleading misprint, according to which the spatial part of a 4-velocity vector is given to be the ordinary velocity vector.[79] Other sources of error can easily be imagined, of course.[80] It is strange that Minkowski did not check the transformation properties of his 4-velocity definition, but given its provenance, he probably had no reason to doubt its soundness.

Minkowski's mistake strongly suggests that at the time of his lecture, he did not yet conceive of particle motion in terms of a world line parameter. Such an approach to particle motion would undoubtedly have spared Minkowski the error, since it renders trivial the definition of 4-velocity.[81] As matters stood in November, 1907, however, Minkowski could proceed no further with his project of reformulation.[82] The development of four-dimensional mechanics was hobbled by Minkowski's spare

78 Minkowski mentions this very property on p. 6.

79 The passage in question may be translated as follows: "Next we consider $X, Y, Z, T\sqrt{-1}$, as the coordinates of a fourth point Q; the invariants will then be functions of the mutual distances of the five points O, P, P', P'', Q, and among these functions we must retain only those that are 0th degree homogeneous with respect, on one hand, to $X, Y, Z, T, \delta x, \delta y, \delta z, \delta t$ (variables that can be further replaced by $X_1, Y_1, Z_1, T_1, \xi, \eta, \zeta, 1$), and on the other hand, with respect to $\delta_1 x, \delta_1 y, \delta_1 z, 1$ (variables that can be further replaced by $\xi_1, \eta_1, \zeta_1, 1$)" (Poincaré 1906, 170). The misprint is in the next-to-last set of variables, where instead of 1 we should have $\delta_1 t$.

80 One other obvious source for Minkowski's error is Lorentz's transformation of charge density: $\rho' = \rho/\beta l^3$, where $1/\beta = \sqrt{1-v^2/c^2}$, and l is a constant later set to unity (Lorentz 1904a, 813), although this formula was carefully corrected by Poincaré.

81 Let the differential parameter $d\tau$ of a world line be expressed in Minkowskian coordinates by $d\tau^2 = dx_1^2 + dx_2^2 + dx_3^2 + dx_4^2$. The 4-velocity vector U_μ is naturally defined to be the first derivative with respect to this parameter, $U_\mu = dx_\mu/d\tau$ $(\mu = 1, 2, 3, 4)$.

stock of 4-vectors even more than that of electrodynamics. Although Minkowski defined a force-density 4-vector, the fourth component of which he correctly identified as the energy equation, he did not go on to define 4-force at a point.[83] Once again, the definition of a force 4-vector at a point would have been trivial, had Minkowski possessed a correct 4-velocity definition. No more than a review of Planck's recent investigation (Planck 1907), Minkowski's discussion of mechanics involved no 4-vectors at all. Likewise for the subsequent section on gravitation, which reviewed Poincaré's theory, as shown above (p. 214). Without a valid 4-vector for velocity, Minkowski's electrodynamics of moving media was severely hobbled; without a point force 4-vector, his four-dimensional mechanics and theory of gravitation could go nowhere.

The difficulty encountered by Minkowski in formulating a four-dimensional approach to physics is surprising in light of the account he gave later of the background to his discovery of spacetime (Minkowski 1909). Minkowski presented his spacetime view of relativity theory as a simple application of group methods to the differential equations of classical mechanics. These equations were known to be invariant with respect to uniform translations, just as the squared sum of differentials $dx^2 + dy^2 + dz^2$ was known to be invariant with respect to rotations and translations of Cartesian axes in Euclidean 3-space, and yet no one, he said, had thought of compounding the two corresponding transformation groups. When this is done properly (by introducing a positive parameter c), one ends up with a group Minkowski designated G_c, with respect to which the laws of physics are covariant. (The group G_c is now known as the Poincaré group.) Presumably, the four-dimensional approach appeared simple to Minkowski in hindsight, several months after his struggle with 4-velocity.

In summary, while Minkowski formulated the idea of a four-dimensional language for physics based on the form-invariance of the Maxwell equations under the transformations of the Lorentz group, his development of this program beyond electrodynamics was hindered by a misunderstanding of the four-dimensional counterpart of an ordinary velocity vector. This was to be only a temporary obstacle.

On 21 December, 1907, Minkowski presented to the Royal Society of Science in Göttingen a memoir entitled "The Basic Equations for Electromagnetic Processes in Moving Bodies," which I will refer to for brevity as the *Grundgleichungen*.[84] Minkowski's memoir revisits in detail most of the topics introduced in his 5 November lecture to the mathematical society, but employs none of the jargon of spaces,

---

82 The incongruity noted by Pyenson (Pyenson 1985, 84) between Minkowski's announcement of a four-dimensional physics on one hand, and on the other hand, a trifle of 4-vector definitions and expressions, is to be understood as a indication of Minkowski's gradual ascent of the learning curve of four-dimensional physics.

83 Minkowski defined the spatial components of the empty space force density 4-vector $\mathfrak{X}_j$ in terms of the ordinary force density components $\mathfrak{X}, \mathfrak{Y}, \mathfrak{Z}$, and their product with velocity: $\mathfrak{A} = \mathfrak{X}\mathfrak{w}_x, \mathfrak{Y}\mathfrak{w}_y, \mathfrak{Z}\mathfrak{w}_z$, such that $\mathfrak{X}_j = \mathfrak{X}, \mathfrak{Y}, \mathfrak{Z}, i\mathfrak{A}$. He also expressed the force density 4-vector as the product of 4-current-density and the Traktor: $\mathfrak{X}_j = \rho_1\psi_{j1} + \rho_2\psi_{j2} + \rho_3\psi_{j3} + \rho_4\psi_{j4}$.

geometries, and manifolds. What it emphasizes instead—in agreement with its title—is the achievement of the first theory of electrodynamics of moving bodies in full conformance to the principle of relativity. Also underlined is a second result described as "very surprising": the laws of mechanics follow from the postulate of relativity and the law of energy conservation alone. On the four-dimensional world and the new form of the equations of electrodynamics, both topics headlined in his November lecture, Minkowski remained coy. Curiously, the introduction mentions nothing about a new formalism, even though all but one of fourteen sections introduce and employ new notation or calculation rules (not counting the appendix).

The added emphasis on the laws of mechanics in Minkowski's introduction, on the other hand, reflects Minkowski's recent discovery of correct definitions of 4-velocity and 4-force, along with geometric interpretations of these entities. It was in the *Grundgleichungen* that Minkowski first employed the term "spacetime" [*Raumzeit*].[85] For example, he introduced 4-current density as the exemplar of a "spacetime vector of the first kind" (§5), and used it to derive a velocity 4-vector. Identifying $\rho_1, \rho_2, \rho_3, \rho_4$ with $\rho\mathfrak{w}_x, \rho\mathfrak{w}_y, \rho\mathfrak{w}_z, i\rho$, just as he had done in his lecture of 5 November, Minkowski wrote the transformation to a primed system moving with uniform velocity $q < 1$:

$$\rho' = \rho\left(\frac{-q\mathfrak{w}_z + 1}{\sqrt{1-q^2}}\right),\ \rho'\mathfrak{w}'_{z'} = \rho\left(\frac{\mathfrak{w}_z - q}{\sqrt{1-q^2}}\right),\ \rho'\mathfrak{w}'_{x'} = \rho\mathfrak{w}_x, \rho'\mathfrak{w}'_{y'} = \rho\mathfrak{w}_y. \quad (14)$$

Observing that this transformation did not alter the expression $\rho^2(1-w^2)$, Minkowski announced an "important remark" concerning the relation of the primed to the unprimed velocity vector (§4). Dividing the 4-current density by the positive square root of the latter invariant, he obtained a valid definition of 4-velocity,

$$\frac{\mathfrak{w}_x}{\sqrt{1-\mathfrak{w}^2}},\quad \frac{\mathfrak{w}_y}{\sqrt{1-\mathfrak{w}^2}},\quad \frac{\mathfrak{w}_z}{\sqrt{1-\mathfrak{w}^2}},\quad \frac{i}{\sqrt{1-\mathfrak{w}^2}}, \quad (15)$$

the squared magnitude of which is equal to –1. Minkowski seemed satisfied with this definition, naming it the spacetime velocity vector [*Raum-Zeit-Vektor Geschwindigkeit*].

---

84 Minkowski's manuscript was delivered to the printer on 21 February, 1908, corrected, and published on 5 April, 1908 (*Journal für die "Nachrichten" der Gesellschaft der Wissenschaften zu Göttingen, mathematische-naturwissenschaftliche Klasse* 1894–1912, *Scient.* 66, Nr. 1, 471, *Archiv der Akademie der Wissenschaften zu Göttingen*). I thank Tilman Sauer for pointing out this source to me.

85 While the published version of Minkowski's 5 November lecture refers on one occasion to a "*Raumzeitpunkt*" (Minkowski 1915, 934) the term occurs nowhere in the archival typescript. The source of this addition is unknown. A manuscript annotation of the first page of the typescript bears Sommerfeld's initials, and indicates that he compared parts of the typescript to the proofs, as Lewis Pyenson correctly points out (Pyenson 1985, 82). Pyenson errs, however, in attributing to Sommerfeld the authorship of the remaining annotations, which were all penned in Minkowski's characteristic cramped hand.

The significance of the spacetime velocity vector, Minkowski observed, lies in the relation it establishes between the coordinate differentials and matter in motion, according to the expression

$$\sqrt{-(dx_1^2 + dx_2^2 + dx_3^2 + dx_4^2)} = dt\sqrt{1-\mathfrak{w}^2}. \tag{16}$$

The Lorentz-invariance of the right-hand side of (16), signaled earlier by both Poincaré and Planck, now described the relation of the sum of the squares of the coordinate differentials to the components of 4-velocity.

The latter relation plays no direct role in Minkowski's subsequent development of the electrodynamics of moving media, and in this it is unlike the 4-velocity definition. Rewriting the right-hand side of (16) as the ratio of the coordinate differential dx_4 to the temporal component of 4-velocity, w_4, Minkowski defined the spacetime integral of (16) as the "proper time" [*Eigenzeit*] pertaining to a particle of matter. The introduction of proper time streamlined Minkowski's 4-vector expressions, for instance, 4-velocity was now expressed in terms of the coordinate differentials, the imaginary unit, and the differential of proper time, $d\tau$:

$$\frac{dx}{d\tau}, \quad \frac{dy}{d\tau}, \quad \frac{dz}{d\tau}, \quad i\frac{dt}{d\tau}. \tag{17}$$

Along with the notational simplification realized by the introduction of proper time, Minkowski signaled a geometric interpretation of 4-velocity. Since proper time is the parameter of a spacetime line (or as he later called it, a world line), it follows that 4-velocity is equal to the slope of a world line at a given spacetime point, much like ordinary three-velocity is described by the slope of a displacement curve in classical kinematics. What Minkowski pointed out, in other words, is that 4-velocity is tangent to a world line at a given spacetime point (p. 108).

In order to develop his mechanics, Minkowski needed a workable definition of mass. He adapted Einstein's and Planck's notion of rest mass to the arena of spacetime by considering that a particle of matter sweeps out a hypertube in spacetime. Conservation of particle mass m was then expressed as invariance of the product of rest mass density with the volume slices of successive constant-time hypersurfaces over the length of the particle's world line, such that $dm/d\tau = 0$. Minkowski did not consider the case of variable rest mass density, which arises, for instance, in the case of heat exchange.

Minkowski's decision to adopt a constant rest mass density is linked to his view of the electrodynamics of moving media. Recall that he had introduced a six-vector in his 5 November lecture to represent the field. The product of the field and excitation six-vectors, he noted, leads to an interesting 4 by 4 matrix, combining the Maxwell stresses, Poynting vector, and electromagnetic energy density. He did not assign a name to this object, known later as the energy-momentum tensor, and often viewed as one of Minkowski's greatest achievements in electrodynamics.[86] Of special interest to Minkowski was the fact that the 4-divergence of this matrix, denoted lor S, is a 4-vector, K: [87]

$$K = \text{lor}\, S. \tag{18}$$

This 4-divergence (18) was used to define the "ponderomotive" force density, or generalized force per unit volume, neither mechanical nor non-mechanical in the pure sense of these terms. The 4-vector K is not normal, in general, to the velocity w of a given volume element, so to ensure that the ponderomotive force acts orthogonally to w, Minkowski added a component containing a velocity term:

$$K + (w\bar{K})w. \tag{19}$$

The parentheses in (19) indicate a scalar product, and $\bar{K}$ stands for the transpose of K. By defining the ponderomotive force density in this way, Minkowski effectively opted for an equation of motion in which 4-acceleration is normal to 4-velocity.[88] It appears that Minkowski let this view of force and acceleration guide his development of spacetime mechanics. In the latter domain, he formed a 4 by 4 matrix S in the force density and energy of an elastic media with the same transformation properties as the energy-momentum tensor S of (18), and used the 4-divergence of this tensor to express the equations of motion of a volume element of constant rest mass density ν (p. 106):

$$\nu\frac{dw_h}{d\tau} = K_h + \kappa w_h \quad (h = 1, 2, 3, 4). \tag{20}$$

The factor κ was determined by the definition of 4-velocity to be equal to the scalar product $(K\bar{w})$, much like the definition of ponderomotive force (19). In sum, it may be supposed that the non-orthogonality with respect to a given volume element of the 4-divergence of Minkowski's asymmetric energy-momentum tensor for moving media led Minkowski to introduce a velocity term to his definition of ponderomotive force. This definition was then ported to spacetime mechanics, where for the sake of consistency, Minkowski held rest mass density constant in the equations of motion (20).

Minkowski's stipulation of constant rest mass density was eventually challenged by Max Abraham (Abraham 1909, 739) and others, for reasons that do not concern us here. Despite its obvious drawbacks, it greatly simplified the tasks of outlining the mechanics of spacetime and developing a theory of gravitation. For example, it per-

---

86 While Minkowski's tensor is traceless, it is also asymmetric, a fact which led to criticism and rejection by leading theorists of the day. His asymmetric tensor was later rehabilitated; for a technical discussion with reference to the original papers, see (Møller 1972, 219). In the absence of matter, his tensor assumes a symmetric form; in this form, it was hailed by theorists.

87 Minkowski defined the energy-momentum tensor S in two ways: as the product of six-vectors, $fF = S - L$, where L is the Lagrange density, and in component form via the equations for Maxwell stresses, the Poynting vector, and electromagnetic energy density (Minkowski 1908, 96).

88 Minkowski's alternative between a 4-force definition and the "natural" spacetime equations of motion was underlined by Pauli (Pauli 1921, 664).

mitted him to define the equations of motion of a particle in terms of the product of rest mass and 4-acceleration, where the latter is the derivative of 4-velocity with respect to proper time. Since 4-velocity is orthogonal to 4-acceleration, for constant proper mass it is also orthogonal to a 4-vector Minkowski called a "driving force" [*bewegende Kraft*, p. 108]. Minkowski wrote four equations defining this force:

$$m\frac{d}{d\tau}\frac{dx}{d\tau} = R_x,\ m\frac{d}{d\tau}\frac{dy}{d\tau} = R_y,\ m\frac{d}{d\tau}\frac{dz}{d\tau} = R_z,\ m\frac{d}{d\tau}\frac{dt}{d\tau} = R_t. \tag{21}$$

The first three expressions differ from Planck's equations of motion, in that Planck defined force as change in *momentum*, instead of mass times acceleration. It was only a few months later that Minkowski explicitly defined four-momentum as the product of 4-velocity with proper mass (Planck 1906, eq. 6; Minkowsi 1909, §4).[89] By dividing Minkowski's first three equations by a Lorentz factor, one obtains Planck's equations. Minkowski's fourth equation, R_t, formally dependent on the other three, expresses the law of energy conservation.[90] From energy conservation and the relativity postulate alone, Minkowski concluded, one may derive the equations of motion. This is the single "surprising" result of his investigation of relativistic mechanics, referred to at the outset of his paper (see above, p. 218).

If Minkowski found few surprises in spacetime mechanics, many of his readers were taken aback by his four-dimensional approach. For example, the first physicists to comment on his work, Albert Einstein and Jakob Laub, rewrote Minkowski's expressions in ordinary vector notation, sparing the reader the "sizable demands" [*ziemlich große Anforderungen*] of Minkowski's mathematics (Einstein and Laub 1908, 532). They did not specify the nature of the demands, but the abstracter of their paper pointed to the "special knowledge of the calculation methods" required for assimilation of Minkowski's equations.[91] In other words, for Minkowski's readers, his novel matrix calculus was the principal technical obstacle to overcome. Where Poincaré pushed rejection of formalism to an extreme, Minkowski pulled in the other direction, introducing a formalism foreign to the practice of physics. What motivated this brash move is unclear, and his choice is all the more curious because he knowingly defied the German trend of vector notation in electrodynamics.[92] As mentioned above, Minkowski was ill-disposed toward quaternions, although he admitted in print that they could be brought into use for relativity instead of matrix calculus. He spoke here from experience, as manuscript notes reveal that he used quaternions (in addition to Cartesian-coordinate representation and ordinary vector analysis) to investigate the electrodynamics of moving media.[93] In the end, however, he felt that for his purposes quaternions were "too limited and cumbersome" [*zu eng und schwerfällig*, p. 79].

---

89 In the latter lecture, Minkowski proposed the modern definition of kinetic energy as the temporal component of 4-momentum times c^2, or $mc^2 dt/d\tau$. The "spatial" part of the driving force (21) was referred to by Lorentz (Lorentz 1910, 1237) as a "Minkowskian force" [*Minkowskische Kraft*], differing from the Newtonian force by a Lorentz factor. Lorentz complemented the Minkowskian force with a "Minkowskian mass" [*Minkowskische Masse*].

As far as notation is concerned, Minkowski's treatment of differential operations broke cleanly with then-current practice. It also broke with the precedent of his 5 November lecture, where he had introduced, albeit parsimoniously, both $\square$ and Div (see above, p. 215). For the *Grundgleichungen* he adopted a different approach, extending the ∇ to four dimensions, and labeling the resulting operator *lor*, already encountered above in (18). The name is short for Lorentz, and the effect is the operation: $|\partial/\partial x_1, \partial/\partial x_2, \partial/\partial x_3, \partial/\partial x_4|$. When applied to a 6-vector, *lor* results in a 4-vector, in what Minkowski described as an appropriate translation of the matrix product rule (p. 89); it also mimics the effect of the ordinary ∇. Transforming as a 4-vector, *lor* is liberally employed in the second part of the *Grundgleichungen*, to the exclusion of any and all particular 4-vector functions.[94] The use of *lor* made for a presentation of electrodynamics elegant in the extreme, at the expense of legibility for German physicists more used to thinking in terms of gradients, divergences, and curls (or rotations).

Minkowski's equations of electrodynamics departed radically in form with those of the old electrodynamics, shocking the thought patterns of physicists, and creating a phenomenon of rejection that took several years—and a new formalism—to overcome.[95] Why did Minkowski break with this tradition? Did he feel that the new physics of spacetime required a clean break with nineteenth-century practice? Perhaps,

---

90 Minkowski's argument may be summarized as follows. From the definition of a 4-vector, the following orthogonality relation holds for the driving force R:

$$R_x\frac{dx}{d\tau} + R_y\frac{dy}{d\tau} + R_z\frac{dz}{d\tau} = R_t\frac{dt}{d\tau}. \tag{22}$$

Integration of the rest-mass density over the hypersurface normal to the world line of the mass point results in the driving force components (21), but if the integration is to be performed instead over a constant-time hypersurface, proper time is replaced by coordinate time, such that the fourth equation reads: $md/dt(dt/d\tau) = R_t d\tau/dt$. From (22) we obtain an expression for R_t, which we multiply by $d\tau/dt$:

$$m\frac{d}{dt}\left(\frac{dt}{d\tau}\right) = \mathfrak{w}_x R_x\frac{d\tau}{dt} + \mathfrak{w}_y R_y\frac{d\tau}{dt} + \mathfrak{w}_z R_z\frac{d\tau}{dt}. \tag{23}$$

Minkowski reasoned that since the right-hand side of (23) describes the rate at which work is done on the particle, the left-hand side must be the rate of change of the particle's kinetic energy, such that (23) represents the law of energy conservation. He immediately related (23) to the kinetic energy of the particle:

$$m\left(\frac{dt}{d\tau} - 1\right) = m\left(\frac{1}{\sqrt{1-\mathfrak{w}^2}} - 1\right) = m\left(\frac{1}{2}|\mathfrak{w}|^2 + \frac{3}{8}|\mathfrak{w}|^4 + \cdots\right). \tag{24}$$

Minkowski did not justify the latter expression, but in virtue of his definition of proper time, $d\tau = dt\sqrt{1-\mathfrak{w}^2}$, the left-hand side of (23) may be rewritten as $m(d/dt)(1/\sqrt{1-\mathfrak{w}^2})$, such that upon integration the particle's kinetic energy is $m/\sqrt{1-\mathfrak{w}^2} + C$, where C is a constant. For agreement with the Newtonian expression of kinetic energy in case of small particle velocities ($\mathfrak{w} \ll 1$), we let $C = -m$, which accords both with (24) and the definition of kinetic energy given in a later lecture (cf. note 89).

91 *Jahrbuch über die Fortschritte für Mathematik* 39, 1908, 910.

but he must have recognized that the old methods would prove resistant to change. His own subsequent practice shows as much: after writing the *Grundgleichungen* Minkowski did not bother with *lor* during his private explorations of the formal side of electrodynamics, preferring the coordinate method.[96]

He also relied largely—but not exclusively—on a Cartesian-coordinate approach during his preliminary investigations of the subjects treated in the *Grundgleichungen*. His surviving research notes, made up almost entirely of symbolic calculations, shed an interesting light on Minkowski's path to both a theory of the electrodynamics of moving media, and a theory of gravitation, or more generally to his process of discovery. Notably, where the subjects of mechanics and gravitation are relegated to the appendix of the *Grundgleichungen*, these notes show that Minkowski pursued questions of electrodynamics and gravitation in parallel, switching from one topic to the other three times in the course of 163 carefully numbered pages. At least fifteen of these pages are specifically concerned with gravitation; the notes are undated, but those concerning gravitation are certainly posterior to the typescript of the 5 November lecture, because unlike the latter, they feature valid definitions of 4-velocity and 4-force.

Minkowski's attempt to capture gravitational action in terms of a 4-scalar potential is of particular interest. We recall that Minkowski had expressed Maxwell's equations in terms of a 4-vector potential (11) during his lecture of 5 November, and on this basis, it was natural for him to investigate the possibility of representing gravitational force on a point mass in a fashion analogous to that of the force on a point charge moving in an electromagnetic field. In his scratch notes, Minkowski defined a 4-scalar potential Φ, in terms of which he initially devised the law of motion:

---

92 This trend is described by Darrigol (Darrigol 1993, 270). The sharp contrast between the importance assigned to vector methods in France and Germany may be linked to the status accorded to applied mathematics in these two nations, as discussed by H. Gispert in her review of the French version of Klein's *Encyklopädie* (Gispert 2001).

93 At one point during his calculations Minkowski seemed convinced of the utility of this formalism, remarking that electrodynamics is "predestined for application of quaternionic calculations" (*Math. Archiv* 60: 6, 21, Handschriftenabteilung, NSUB).

94 A precedent for Minkowski's exclusive use of *lor* may be found in (Gibbs and Wilson 1901), where ∇ is similarly preferred to vector functions.

95 Cf. Max von Laue's remark that physicists understood little of Minkowski's work because of its unfamiliar mathematical expression (Laue 1951, 515) and Chuang Liu's account of the difficulty experienced by Max Abraham and Gunnar Nordström in applying Minkowski's formalism (Liu 1991, 66). While Minkowski's calculus is a straightforward extension of Cayley's formalism (for a summary, see (Cunningham 1914, chap. 8), the latter formalism was itself unfamiliar to physicists.

96 *Math. Archiv* 60: 5, Handschriftenabteilung, NSUB. This 82-page set of notes dates from 23 May to 6 October, 1908. A posthumously published paper on the electron-theoretical derivation of the basic equations of electrodynamics for moving media, while purported to be from Minkowski's Nachlass, was written entirely by Max Born, as he acknowledged (Minkowski and Born 1910, 527). In the latter publication *lor* makes only a brief appearance.

$$\frac{d}{d\tau}\frac{1}{\sqrt{1-v^2}}-\frac{\partial\Phi}{\partial t} = 0,\quad \frac{d}{d\tau}\frac{-y'}{\sqrt{1-v^2}}-\frac{\partial\Phi}{\partial y} = 0,$$
$$\frac{d}{d\tau}\frac{-x'}{\sqrt{1-v^2}}-\frac{\partial\Phi}{\partial x} = 0,\quad \frac{d}{d\tau}\frac{-z'}{\sqrt{1-v^2}}-\frac{\partial\Phi}{\partial z} = 0, \tag{25}$$

where constants are neglected, τ denotes proper time, and primes indicate differentiation with respect to coordinate time t (i.e. $x' = dx/dt$).[97] This generalization of the Newtonian potential to a 4-scalar potential appears to be one of the first paths explored by Minkowski in his study of gravitation, but his investigation is inconclusive. In particular, there is no indication in these notes of a recognition on Minkowski's part that a four-scalar potential conflicts with the postulates of invariant rest mass and light velocity.[98] Nor is there any evidence that he considered suspending either one of these postulates.

Likewise, in the *Grundgleichungen* there is no question of adopting either a variable mass density or a gravitational 4-potential. Once he had established the foundations of spacetime mechanics, Minkowski took up the case of gravitational attraction. The problem choice is significant, in that the same question had been treated at length by Poincaré (although not to Minkowski's satisfaction, as mentioned above, p. 215). Implicitly, Minkowski encouraged readers to compare methods and results. Explicitly, proceeding in what he described (in a footnote) as a "wholly different way" from Poincaré, Minkowski said he wanted to make "plausible" the inclusion of gravitation in the scheme of relativistic mechanics (p. 109). It will become clear in what follows that his project was more ambitious than the modest elaboration of a plausibility argument, as it was designed to validate his spacetime mechanics.

The point of departure for Minkowski's theory of gravitation was quite different from that of Poincaré, because the latter's results were integrated into the former's formalism. For example, where Poincaré initially assumed a finite propagation velocity of gravitation no greater than that of light, only to opt in the end for a velocity equal to that of light, Minkowski assumed implicitly from the outset that this velocity was equal to that of light. Similarly, Poincaré initially supposed the gravitational

97 *Math. Archiv* 60: 6, 10, Handschriftenabteilung, NSUB.

98 This "peculiar" consequence of Minkowski's spacetime mechanics was underlined by Maxwell's German translator, the Berlin physicist Max B. Weinstein (Weinstein 1914, 42). In Minkowski spacetime, 4-acceleration is orthogonal to 4-velocity: $U_\mu dU_\mu/d\tau = 0$, $\mu = 1, 2, 3, 4$, where τ is the proper time. We assume a 4-scalar potential Φ such that the gravitational 4-force $F_\mu = -m\partial\Phi/\partial x_\mu$. If we consider a point mass with 4-velocity U_μ subjected to a 4-force F_μ derived from this potential, we have $U_\mu F_\mu = -U_\mu m\partial\Phi/\partial x_\mu$. Writing 4-velocity as $dx_\mu/d\tau$, and substituting in the latter expression, we obtain

$$U_\mu F_\mu = -m\frac{dx_\mu}{d\tau}\frac{\partial\Phi}{\partial x_\mu} = -m\frac{d\Phi}{d\tau} = 0,$$

and consequently, $d\Phi/d\tau = 0$, which means that the law of motion describes the trajectory of the passive mass m only in the trivial case of constant Φ along its world line.

force to be Lorentz covariant, only to opt in the end for an analog of the Lorentz force, where Minkowski required implicitly from the outset that all forces transform like the Lorentz force.

Combining geometric and symbolic arguments, Minkowski's exposition of his theory of gravitation introduces a new geometric object, the three-dimensional "ray form" [*Strahlgebilde*] of a given spacetime point, known today as a light hypercone (or lightcone). For a fixed spacetime point $B^* = (x^*, y^*, z^*, t^*)$, the lightcone of B^* is defined by the sets of spacetime points $B = (x, y, z, t)$ satisfying the equation

$$(x - x*)^2 + (y - y*)^2 + (z - z*)^2 = (t - t*)^2, \quad t - t* \geqq 0. \tag{26}$$

For all the spacetime points B of this lightcone, B^* is what Minkowski called *B's lightpoint*. Any world line intersects the lightcone in one spacetime point only, Minkowski observed, such that for any spacetime point B on a world line there exists one and only one lightpoint B^*. Minkowski remarked in a later lecture that the lightcone divides four-dimensional space into three regions: time-like, space-like and light-like.[99]

Using this novel insight to the structure of four-dimensional space, in combination with the 4-vector notation set up in earlier in his memoir, Minkowski presented and applied his law of gravitational attraction in two highly condensed pages. Minkowski's geometric argument employs non-Euclidean relations that were unfamiliar to physicists, yet he provided no diagrams. Visually-intuitive arguments had fallen into disfavor with mathematicians by this time, with the rise of the axiomatic approach to geometry favored by David Hilbert (Rowe 1997), yet Minkowski never renounced the use of figures in geometry; he employed them in earlier works on number geometry, and went on to publish spacetime diagrams in the sequel to the *Grundgleichungen*.[100] For the purposes of my reconstruction, I refer to a spacetime diagram (Fig. 1) of the sort Minkowski employed in the sequel (reproduced in Fig. 3).[101]

99 Minkowski introduced the terms *zeitartig* and *raumartig* in (Minkowski 1909).

100 There is little agreement on where to situate Minkowski's work on relativity along a line running from the intuitive to the formal. Peter Galison (Galison 1979, 89) for example, underlines Minkowski's visual thinking (i.e., reasoning that appeals to figures or diagrams), while Leo Corry (Corry 1997, 275; 2004, chap. 4) considers Minkowski's work in the context of Hilbert's axiomatic program for physics.

101 Two spatial dimensions are suppressed in Fig. 1, and lightcones are represented by broken lines with slope equal to ± 1, the units being chosen so that the propagation velocity of light is unity $(c = 1)$. In this model of Minkowski space, orthogonal coordinate axes appear oblique in general, for example, the spatial axes x^*, y^*, z^* are orthogonal to the tangent B^*C^* at spacetime point B^* of the central line of the filament F^* described by a particle of proper mass m^*.

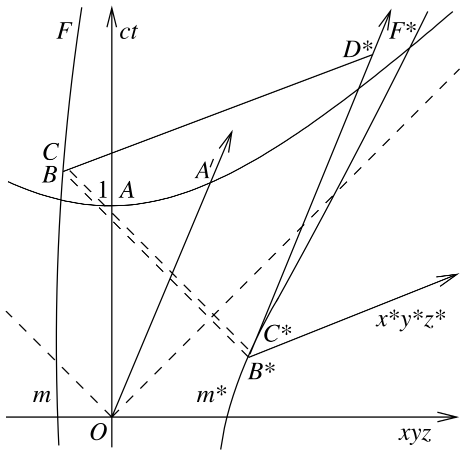


Figure 1: Minkowski's geometry of gravitation, with source in arbitrary motion.

On the assumption that the force of gravitation is a 4-vector normal to the 4-velocity of the passive mass m, Minkowski derived his law of attraction in the following way. The trajectories of two particles of mass m and m^* correspond to two spacetime filaments F and F^*, respectively. Minkowski's arguments refer to world lines he called central lines [*Hauptlinien*] of these filaments, which pass through points on the successive constant-time hypersurfaces delimited by the respective particle volumes. The central lines of the filaments F and F^* are shown in Fig. 1. An infinitesimal element of the central line of F is labeled BC, and the two lightpoints corresponding to the endpoints B and C are labeled B^* and C^* on the central line of F^*. From the origin of the rest frame O, a 4-vector parallel to B^*C^* intersects at A' the three-dimensional hypersurface defined by the equation $-x^2-y^2-z^2+t^2=1$. Finally, a space-like 4-vector BD^* extends from B to a point D^* on the world line tangent to the central line of F^* at B^*.

Referring to the latter configuration of seven spacetime points, two central lines, a lightcone and a calibration hypersurface, Minkowski expressed the spatial components of the driving force of gravitation exerted by m^* on m at B,

$$mm^*\left(\frac{OA'}{B^*D^*}\right)^3 BD^*. \tag{27}$$

Minkowski's gravitational driving force is composed of the latter 4-vector (27) and a second 4-vector parallel to B^*C^* at B, such that the driving force is always orthogonal to the 4-velocity of the passive mass m at B. (For reasons of commodity, I will refer to this law of force as Minkowski's first law.)

The form of Minkowski's first law of gravitation is comparable to that of his ponderomotive force for moving media (19), in that the driving force has two components, only one of which depends on the motion of the test particle. In the gravitational case, however, Minkowski did not write out the 4-vector components in terms of matrix products. Instead, he relied on spacetime geometry and the definition of a 4-vector. The only way physicists could understand (27) was by reformulating it in terms of ordinary vectors referring to a conveniently chosen inertial frame, and even then, they could not rely on Minkowski's description alone, as it is incomplete.[102]

Even without spacetime diagrams or a transcription into ordinary vector notation, the formal analogy of (27) to Newton's law is readily apparent, and this is probably why Minkowski wrote it this way. In doing so, however, he passed up an opportunity to employ the new matrix machinery at his disposal. Had he seized this opportunity, he would have gained a simple, self-contained, coordinate-free expression of the law of gravitation, and provided readers with a more elaborate example of his calculus in action, but the latter desiderata must not have been among his primary objectives.[103]

---

102 The 4-vector OA' in (27) has unit magnitude by definition in all inertial frames, while B^*D^* is a time-like 4-vector tangent to the central line of F^* at B^*. Consequently, B^*D^* may be taken to coincide with the temporal axis of a frame instantaneously at rest with m^* at B^*, such that it has only one nonzero component: the difference in proper time between the points B^* and D^*. It is assumed that the rest frame may be determined unambiguously for a particle in arbitrary motion, as asserted without proof by Minkowski in a later lecture (Minkowski 1909, §III); subsequently, Max Born (Born 1909, 26) remarked that any motion may be approximated by what he called hyperbolic motion, and noted that such motion is characterized by an acceleration of constant magnitude (as measured in an inertial frame). If we locate the origin of this frame at B^*, and let $D^* = (0, 0, 0, t)$, then $B^*D^* = (0, 0, 0, it)$, and $(B^*D^*)^3 = -it^3$. Likewise in this same frame, $A = A' = (0, 0, 0, 1)$, and $OA' = OA = (0, 0, 0, i)$. Minkowski understood the term (OA'/B^*D^*) as the ratio [*Verhältnis*] of two parallel 4-vectors, an operation familiar from the calculus of quaternions, but one not defined for 4-vectors. While modern vector systems ignore vector division, in Hamilton's quaternionic calculus the quotient of vectors is unambiguously defined; see, for example, (Tait 1882–1884, chap. 2). Accordingly, the quotient in (27) is the ratio of lengths, $(OA'/B^*D^*) = 1/t$, and the cubed ratio is t^{-3}. The point B lies on the same constant-time hypersurface as D^*, so we assign it the value $(x, y, z, t) = (\boldsymbol{r}, t)$. This assignment determines the value of the 4-vector $BD^* = (-x, -y, -z, 0) = (-\boldsymbol{r}, 0)$. Since B^* is a lightpoint of B, we can apply (26) to obtain $x^2 + y^2 + z^2 = t^2 = r^2$, and consequently, $t^3 = r^3$. Substituting for t^3 results in $(OA'/B^*D^*)^3 = 1/t^3 = 1/r^3$. The 4-vector B^*D^* is space-like, such that its projection on the constant-time hypersurface orthogonal to B^*D^* at D^* is the ordinary vector $(-x, -y, -z) = -\boldsymbol{r}$. In terms of ordinary vectors and scalars measured in the rest frame of m^*, Minkowski's expression (27) is equivalent to Newton's law (neglecting the gravitational constant):

$$-mm^*\frac{\boldsymbol{r}}{r^3}. \tag{28}$$

Neither (27) nor (28) contains any velocity-dependent terms, while the time-like component of Minkowski's first law depends on the velocity of the passive mass. Newton's law (28) thus coincides with Minkowski's first law only in the case of relative rest.

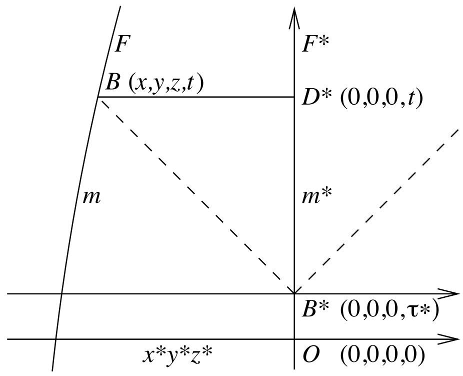


Figure 2: Minkowski's geometry of gravitation, with source in uniform motion.

Minkowski was not yet finished with his law of gravitation. Unlike Poincaré, after writing his law of gravitation, Minkowski went on to apply it to the particular case of uniform rectilinear motion of the source m^*. He considered the latter in a comoving frame, in which the temporal axis is chosen to coincide with the tangent to the central line of F^* at B^* (cf. the situation described in note 102). Referring to the reconstructed spacetime diagram in Fig. 2, the temporal axis is represented by a vertical line F^*, such that the origin is established in a frame comoving with m^*. To the retarded position of m^*, denoted B^*, Minkowski assigned the coordinates $(0, 0, 0, \tau^*)$, and to the position B of the passive mass m he assigned the coordinates (x, y, z, t). The geometry of this configuration fixes the location of D^* at $(0, 0, 0, t)$, from which the 4-vectors $BD^* = (-x, -y, -z, 0)$ and $B^*D^* = (0, 0, 0, i(t-\tau^*))$ are determined. In this case, Minkowski pointed out, (26) reduces to:

$$x^2 + y^2 + z^2 = (t - \tau^*)^2. \tag{29}$$

Substituting the above values of BD^* and B^*D^* into Minkowski's formula (27), the spatial components of the 4-acceleration of the passive mass m at B due to the active mass m^* at B^* turn out to be:[104]

$$\frac{d^2x}{d\tau^2} = -\frac{m^*x}{(t-\tau^*)^3}, \quad \frac{d^2y}{d\tau^2} = -\frac{m^*y}{(t-\tau^*)^3}, \quad \frac{d^2z}{d\tau^2} = -\frac{m^*z}{(t-\tau^*)^3}. \tag{30}$$

103 Minkowski's driving force may be expressed in his notation as a function of scalar products of 4-velocities and 4-position:

$$-mm^*\frac{(w\bar{w}^*)\mathfrak{R} - (w\bar{\mathfrak{R}})w^*}{(\mathfrak{R}\bar{w}^*)^3(w\bar{w}^*)}.$$

Here I let w and w^* designate 4-velocity at the passive and active mass points, while $\mathfrak{R}$ is the associated 4-position, the parentheses denote a scalar product, and the bar indicates transposition.

From (30) and (29), the corresponding temporal component at B may be determined:[105]

$$\frac{d^2t}{d\tau^2} = -\frac{m^*}{(t-\tau^*)^2}\frac{d(t-\tau^*)}{dt}. \tag{33}$$

Inspecting (30), it appears that the only difference between these acceleration components and those corresponding to Newtonian attraction is a replacement in the latter of coordinate time t by proper time τ.[106]

The formal similarity of (30) to the Newtonian law of motion under a central force probably suggested to Minkowski that his law induces Keplerian trajectories. With the knowledge gained from (30), to the effect that the only difference between classical and relativistic trajectories is that arising from the substitution of proper time for coordinate time, Minkowski demonstrated the compatibility of his relativistic law of gravitation with observation using only Kepler's equation and the definition of 4-velocity.

Writing Kepler's equation in terms of proper time yields:

$$n\tau = E - e\sin E, \tag{34}$$

where $n\tau$ denotes the mean anomaly, e the eccentricity, and E the eccentric anomaly. Minkowski referred to (34) and to the norm of a 4-velocity vector:

$$\left(\frac{dx}{d\tau}\right)^2 + \left(\frac{dy}{d\tau}\right)^2 + \left(\frac{dz}{d\tau}\right)^2 = \left(\frac{dt}{d\tau}\right)^2 - 1, \tag{35}$$

---

104 The intermediate calculations can be reconstructed as follows. Let the driving force be designated F_μ, $\mu = 1, 2, 3, 4$. Since $(OA'/B^*D^*)^3 = t^{-3}$, and $BD^* = (-x, -y, -z, 0)$, equations (21) and (27) yield: $F_1/m = d^2x/d\tau^2 = -m^*x/(t-\tau^*)^3$, $F_2/m = d^2y/d\tau^2 = -m^*y/(t-\tau^*)^3$, $F_3/m = d^2z/d\tau^2 = -m^*z/(t-\tau^*)^3$.

105 Minkowski omitted the intermediate calculations, which may be reconstructed in modern notation as follows. Let the 4-velocity of the passive mass point be designated $U_\mu = (dx/d\tau, dy/d\tau, dz/d\tau, idt/d\tau)$, while the first three components of its 4-acceleration, designated A_μ, at B due to the source m^* are given by (30). From the orthogonality of 4-velocity and 4-acceleration we have:

$$U_\mu A_\mu = -\frac{dx}{d\tau}\frac{m^*x}{(t-\tau^*)^3} - \frac{dy}{d\tau}\frac{m^*y}{(t-\tau^*)^3} - \frac{dz}{d\tau}\frac{m^*z}{(t-\tau^*)^3} - \frac{idt}{d\tau}\frac{id^2t}{d\tau^2} = 0.$$

Rearranging (31) results in an expression for the temporal component of 4-acceleration:

$$\frac{d^2t}{d\tau^2} = -\frac{m^*}{(t-\tau^*)^3}\left(\frac{xdx}{dt} + \frac{ydy}{dt} + \frac{zdz}{dt}\right).$$

Differentiating (29) with respect to dt results in $xdx/dt + ydy/dt + zdz/dt = (t-\tau^*)d(t-\tau^*)/dt$, the right-hand side of which we substitute in (32) to obtain (33).

106 A young Polish physicist in Göttingen, Felix Joachim de Wisniewski later studied this case, but with equations differing from (30) by a Lorentz factor (de Wisniewski 1913a, 388). In a postscript to the second installment of his paper (Wisniewski 1913b, 676) he employed Minkowski's matrix notation, becoming, with Max Born, one of the rare physicists to adopt this notation.

in order to determine the difference between the mean anomaly in coordinate time nt and the mean anomaly in proper time $n\tau$. From (35), Minkowski deduced:[107]

$$\left(\frac{dt}{d\tau}\right)^2 - 1 = \frac{m^*}{ac^2}\frac{1 + e\cos E}{1 - e\cos E}. \tag{37}$$

Solving (37) for the coordinate time dt, expanding to terms in c^{-2}, and multiplying by n led Minkowski to the expression:

$$ndt = nd\tau\left(1 + \frac{1}{2}\frac{m^*}{ac^2}\frac{1 + e\cos E}{1 - e\cos E}\right). \tag{38}$$

Recalling (34), Minkowski managed to express the difference between the mean anomaly in coordinate time and proper time:[108]

$$nt + \text{const.} = \left(1 + \frac{1}{2}\frac{m^*}{ac^2}\right)n\tau + \frac{m^*}{ac^2}e\sin E. \tag{39}$$

Evaluating the relativistic factor $(m^*)/ac^2$ for solar mass and the Earth's semi-major axis to be 10^{-8}, Minkowski found the deviation from Newtonian orbits to be negligible in the solar system. On this basis, he concluded that

> a decision *against* such a law and the proposed modified mechanics in favor of the Newtonian law of attraction with Newtonian mechanics would not be deducible from astronomical observations.[109]

According to the quoted remark, there was more at stake here for Minkowski than just the empirical adequacy of his law of gravitational attraction, as his claim is for parity between Newton's law and classical mechanics, on one hand, and the *system*

---

107 The intermediate calculations were omitted by Minkowski, but figure among his research notes (*Math. Archiv* 60: 6, 126–127, Handschriftenabteilung, NSUB). Following the method outlined by Otto Dziobek (Dziobek 1888, 12), Minkowski began with the energy integral of Keplerian motion:

$$\left(\frac{dt}{dW}\right)^2 - 1 = \frac{2}{\ell^2}\left(\frac{M}{R} - C\right), \tag{36}$$

where ℓ denotes the velocity of light, M is the sum of the masses times the gravitational constant, $M = k^2(m + m^*)$, R is the radius, and C is a constant. The left-hand side of (36) is the same as the right-hand side of (35) for $W = \tau$. In order to express dt/dW (which is to say $dt/d\tau$) in terms of E, Minkowski considered a conic section in polar coordinates, with focus at the origin: $R = a(1 - e^2)/(1 + e\cos\varphi) = a(1 - \cos E)$, where a denotes the semi-major axis, and φ is the true anomaly. By eliminating φ in favor of E and e, and differentiating (34), Minkowski obtained an expression equivalent to (37).

108 I insert the eccentricity e in the second term on the right-hand side, correcting an obvious omission in Minkowski's paper (Minkowski 1908, 111, eq. 31).

109 "... eine Entscheidung *gegen* ein solches Gesetz und die vorgeschlagene modifizierte Mechanik zu Gunsten des Newtonschen Attraktionsgesetzes mit der Newtonschen Mechanik aus den astronomischen Beobachtungen nicht abzuleiten sein" (Minkowski 1908, 111).

composed of the law of gravitation and spacetime mechanics on the other hand. This new system, Minkowski claimed, was verified by astronomical observations at least as well as the classical system formed by the Newtonian law of attraction and Newtonian mechanics.

Instead of comparing his law with one or the other of Poincaré's laws, Minkowski noted a difference in *method*, as mentioned above. In light of Minkowski's emphasis on the methodological difference with Poincaré, and the hybrid geometric-symbolic nature of Minkowski's exposition, it is clear that the point of reexamining the problem of relativity and gravitation in the *Grundgleichungen* was not simply to make plausible the inclusion of gravitation in a relativistic framework. Rather, since gravitational attraction was the only example Minkowski provided of his formalism in action, his line of argument served to *validate* his four-dimensional calculus, over and above the requirements of plausibility.

From the latter point of view, Minkowski had grounds for satisfaction, although one imagines that he would have preferred to find that his law diverged from Newton's law just enough to account for the observed anomalies. It stands to reason that if Minkowski had been fully satisfied with his first law, he would not have proposed a second law in his next paper—which turned out to be the last he would finish for publication. The latter article developed out of a well-known lecture entitled "Space and Time" (*Raum und Zeit*), delivered in Cologne on 21 September, 1908, to the mathematics section of the German Association of Scientists and Physicians in its annual meeting (Walter 1999a, 49).

In the final section of his Cologne lecture, Minkowski took up the Lorentz-Poincaré theory, and showed how to determine the field due to a point charge in arbitrary motion. On this occasion, just as in his earlier discussion of gravitation in the *Grundgleichungen*, Minkowski referred to a spacetime diagram, but this time he provided the diagram (Fig. 3). Identifying the 4-vector potential components for the source charge on this diagram, Minkowski remarked that the Liénard-Wiechert law was a consequence of just these geometric relations.[110]

---

110 Minkowski's explanation of the construction of his spacetime diagram (Fig. 3) may be paraphrased in modern terminology as follows. Suppressing the *z*-axis, we associate two world lines with two point charges e_1 and e. The world line of e_1 passes through the point at which we wish to determine the field, P_1. To find the retarded position of the source e, we draw the retrograde lightcone (with broken lines) from P_1, which intersects the world line of e at P, where there is a hyperbola of curvature ρ with three infinitely-near points lying on the world line of e; it has its center at M. The coordinate origin is established at P, by letting the *t*-axis coincide with the tangent to the world line. A line from P_1 intersects this axis orthogonally at point Q; it is space-like, and if its projection on a constant-time hypersurface has length r, the length of the 4-vector PQ is r/c. The 4-vector potential has magnitude e/r and points in the direction of PQ (i.e., parallel to the 4-velocity of e at P). The *x*-axis lies parallel to QP_1, such that N is the intersection of a line through M normal to the *x*-axis.

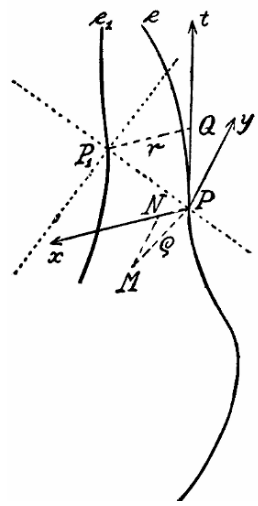


Figure 3: Minkowski's spacetime diagram of particle interaction (Minkowski 1909, 86).

Minkowski then described the driving force between two point charges. Adopting dot notation for differentiation with respect to proper time, he wrote the driving force exerted on an electron of charge e_1 at point P_1 by an electron of charge e:

$$-ee_1\left(\dot{t}_1 - \frac{\dot{x}_1}{c}\right)\mathfrak{K}, \tag{40}$$

where $\dot{t}_1$ and $\dot{x}_1$ are 4-velocity components of the test charge e_1 and $\mathfrak{K}$ is a certain 4-vector. This was the first such description of the electrodynamic driving force due to a 4-vector potential, the simplicity of which, Minkowski claimed, compared favorably with the earlier formulations of Schwarzschild and Lorentz.[111]

In the same celebratory tone, Minkowski finished his article with a discussion of gravitational attraction. The "reformed mechanics", he claimed, dissolved the disturbing disharmonies between Newtonian mechanics and electrodynamics. In order to provide an example of this dissolution, he asked how the Newtonian law of attraction would sit with his principle of relativity. Minkowski continued:

> I will assume that if two point masses m, m_1 describe world lines, a driving force vector is exerted by m on m_1, exactly like the one in the expression just given for the case of electrons, except that instead of $-ee_1$, we must now put in $+mm_1$.

---

111 Minkowski noted four conditions on $\mathfrak{K}$: it is normal to the 4-velocity of e_1 at P_1, $c\mathfrak{K}_t - \mathfrak{K}_x = 1/r^2$, $\mathfrak{K}_y = \ddot{y}/(c^2 r)$, and $\mathfrak{K}_z = 0$, where r is the space-like distance between the test charge e_1 at P_1 and the advanced position Q of the source e, and $\ddot{y}$ is the y-component of e's 4-acceleration at P. For a derivation of the 4-potential and 4-force corresponding to Minkowski's presentation, see (Pauli 1921, 644–645).

Applying the substitution suggested by Minkowski to (40), we obtain:

$$mm_1\left(\dot{t}_1 - \frac{\dot{x}_1}{c}\right)\mathfrak{K}, \tag{41}$$

where the coefficients m and m_1 refer to proper masses. Minkowski's new law of gravitation (41) fully expresses the driving force, unlike the formula (27) of his first law, which describes only one component. In addition, the 4-vectors are immediately identifiable from the notation alone. (In order to distinguish the law given in the *Grundgleichungen* from that of the Cologne lecture [41], I will call [41] Minkowski's second law.)

Since (40) was obtained from Lorentz-Poincaré theory via a 4-vector potential, the law of gravitation (41) ostensibly implied a 4-vector potential as well; in other words, following the example set by Poincaré's second law (10), Minkowski appealed in turn to a Maxwellian theory of gravitation similar to those of Heaviside, Lorentz, and Gans.[112] Although Minkowski made no effort to attach his law to these field theories, it was understood by Sommerfeld to be a formal consequence of just such a theory, as I will show in the next section.

What were the numerical consequences of this new law? Minkowski spared the reader the details, noting only that in the case of uniform motion of the source, the only divergence from a Keplerian orbit would stem from the replacement of coordinate time by proper time. He indicated that the numbers for this case had been worked out earlier, and his conclusion with respect to this new law was naturally the same: combined with the new mechanics, it was supported by astronomical observations to the same extent as the Newtonian law combined with classical mechanics.

Curiously enough, Minkowski offered no explanation of the need for a second law of attraction. Furthermore, by proposing two laws instead of one, Minkowski tacitly acknowledged defeat; despite his criticism of the Poincaré's approach (see above, p. 215), he could hardly claim to have solved unambiguously the problem of gravitation. It may also seem strange that Minkowski discarded the differences between his new law (41) and the one he had proposed earlier.[113]

Minkowski revealed neither the motivation behind a second law of gravitation, nor why he neglected the differences between his two laws, but there is a straightfor-

112 See above, p. 198, (Heaviside 1893), and (Gans 1905). Theories in which the gravitational field is determined by equations having the form of Maxwell's equations were later termed vector theories of gravitation by Max Abraham (Abraham 1914, 477). For a more recent version of such a theory, see (Coster and Shepanski 1969).

113 Minkowski's neglect of the differences between his two theories may explain why historians have failed to distinguish them. The principal difference between the two laws stems from the presence of acceleration effects in the second law. By 1905 it was known that accelerated electrons radiate energy, such that by formal analogy, a Maxwellian theory of gravitation should have featured accelerated point masses radiating "gravitational" energy. For a brief overview of research performed in the first two decades of the twentieth century on the energy radiated from accelerated electrons, see (Whittaker 1951–1953, 2:246).

ward way of explaining both of these mysteries. First, we recall the circumstances of his Cologne lecture, the final section of which Minkowski devoted to the theme of restoring unity to physics. What he wanted to stress on this occasion was that mechanics and electrodynamics harmonized in his four-dimensional scheme of things:

> In the mechanics reformed according to the world postulate, the disturbing disharmonies between Newtonian mechanics and modern electrodynamics fall out on their own.[114]

To support this view, Minkowski had to show that his reformed mechanics was a synthesis of classical mechanics and electrodynamics. A Maxwellian theory of gravitation fit the bill quite well, and consequently, Minkowski brought out his second law of gravitation (41). Clearly, this was not the time to point out the *differences* between his two laws. On the contrary, it was the perfect occasion to observe that a law of gravitation derived from a 4-vector potential formally identical to that of electrodynamics was observationally indistinguishable from Newton's law. Naturally, Minkowski seized this opportunity.

Sadly, Minkowski did not live long enough to develop his ideas on gravitation and electrodynamics; he died on 12 January, 1909, a few days after undergoing an operation for appendicitis. At the time, no objections to a field theory of gravitation analogous to Maxwell's electromagnetic theory were known, apart from Maxwell's own sticking-points (see above, p. 198). However, additional objections to this approach were raised by Max Abraham in 1912, after which the Maxwellian approach withered on the vine, as Gustav Mie and others pursued unified theories of electromagnetism and gravitation.[115]

Minkowski's first law of gravitation fared no better than his second law, but the four-dimensional language in which his two laws were couched had a bright future. The first one to use Minkowski's formal ideas to advantage was Sommerfeld, as we will see next.

3. ARNOLD SOMMERFELD'S HYPER-MINKOWSKIAN LAWS OF GRAVITATION

Neither Poincaré's nor Minkowski's work on gravitation and relativity drew comment until 25 October, 1910, when the second installment of Arnold Sommerfeld's vectorial version of Minkowski's calculus, entitled "Four-dimensional vector analysis" [*Vierdimensionale Vektoranalysis*], appeared in the *Annalen der Physik* (Sommerfeld 1910b). Sommerfeld's contribution differs from those of Poincaré and Minkowski in that it is openly concerned with the presentation of a new formalism, much as its title

114 "In der dem Weltpostulate gemäß reformierten Mechanik fallen die Disharmonien, die zwischen der Newtonschen Mechanik und der modernen Elektrodynamik gestört haben, von selbst aus" (Minkowski 1909, §5).

115 Abraham showed that a mass set into oscillation would be unstable due to the direction of energy flow (Norton 1992, 33). On the early history of unified field theories, see the reference in note 12.

indicates. In this section, I discuss Sommerfeld's interest in vectors, the salient aspects of his 4-vector formalism, and his portrayal of Poincaré's and Minkowski's laws of gravitation.

Sommerfeld displayed a lively interest in vectors, beginning with his editorship of the physics volume of Klein's six-volume *Encyklopädie* in the summer of 1898.[116] He imposed a certain style of vector notation on his contributing authors, including typeface, terminology, symbolic representation of operations, units and dimensions, and the choice of symbols for physical quantities. Articles 12 to 14 of the physics volume appeared in 1904, and were the first to implement the notation scheme backed by Sommerfeld, laid out the same year in the *Physikalische Zeitschrift*.[117] While Sommerfeld belonged to the Vector Commission formed at Felix Klein's behest in 1902, it was clear to him as early as 1901 that the article on Maxwell's theory (commissioned to Lorentz) would serve as a "general directive" for future work in electrodynamics.[118] His intuition turned out to be correct: the principal "vector" of influence was Lorentz's Article 13 (Lorentz 1904b), featuring sections on vector notation and algebra, which set a *de facto* standard for vector approaches to electrodynamics.

As mentioned above (p. 210), only one effort to extend Poincaré's four-dimensional approach beyond the domain of gravitation was published prior to Minkowski's *Grundgleichungen*. By 1910, the outlook for relativity theory had changed due to the authoritative support of Planck and Sommerfeld, the announcement of experimental results favoring Lorentz's electron theory, and the broad diffusion (in 1909) of Minkowski's Cologne lecture. Dozens of physicists and mathematicians began to take an interest in relativity, resulting in a leap in relativist publications.[119]

The principal promoter of Minkowskian relativity, Sommerfeld must have felt by 1910 that it was the right moment to introduce a four-dimensional formalism. He was not alone in feeling this way, for three other formal approaches based on Minkowski's work appeared in 1910. Two of these were 4-vector systems, similar in some respects to Sommerfeld's, and worked out by Max Abraham and the American physical chemist Gilbert Newton Lewis, respectively. A third, non-vectorial approach was proposed by the Zagreb mathematician Vladimir Varičak. Varičak's was a real, four-dimensional, coordinate-based approach relying on hyperbolic geometry. Sommerfeld probably viewed this system as a potential rival to his own approach; although he did not mention Varičak, he wrote that a non-Euclidean approach was

116 Sommerfeld's work on the *Encyklopädie* is discussed in an editorial note to his scientific correspondence (Sommerfeld 2001–2004, 1:40).

117 See (Reiff and Sommerfeld 1904; Lorentz 1904b, 1904c, Sommerfeld 1904). The scheme proposed by Sommerfeld differed from that published in articles 12 to 14 of the *Encyklopädie* only in that the operands of scalar and vector products were no longer separated by a dot.

118 Sommerfeld to Lorentz, 21 March, 1901, (Sommerfeld 2001–2004, 1:191). On Sommerfeld's participation on the Commission see (Reich 1996) and (Sommerfeld 2001–2004, 1:144).

119 For bibliometric data, and discussions of Sommerfeld's role in the rise of relativity theory, see (Walter 1999a, 68–73, 1999b, 96, 108).

possible but could not be recommended (Sommerfeld 1910a, 752, note 1). Of the three alternatives to Sommerfeld's system, the non-Euclidean style pursued by Varičak and others was the only one to obtain even a modest following. An investigation of the reasons for the contemporary neglect of these alternative four-dimensional approaches is beyond the purview of our study; for what concerns us directly, none of these methods was applied to the problem of gravitation.[120]

Sommerfeld's paper, like those of Abraham, Lewis, and Varičak, emphasized formalism, and in this it differed from the *Grundgleichungen*, as mentioned above. Like the latter work, it focused attention on the problem of gravitation. Following the example set by both Poincaré and Minkowski, Sommerfeld capped his two-part *Annalen* paper with an application to gravitational attraction, which consisted of a reformulation, comparison and commentary of their work in his own terms. Not only was Sommerfeld's comparison of Poincaré's and Minkowski's laws of gravitation the first of its kind, it also proved to be the definitive analysis for his generation.

Sommerfeld's four-dimensional vector algebra and analysis offered no new 4-vector or 6-vector definitions, but it introduced a suite of 4-vector functions, notation, and vocabulary. The most far-reaching modification with respect to Minkowski's calculus was the elimination of *lor* (cf. pp. 223–223) in favor of extended versions of ordinary vector functions. In Sommerfeld's notational scheme, the ordinary vector functions div, rot, and grad (used by Lorentz in his *Encyklopädie* article on Maxwell's theory) were replaced by 4-vector counterparts marked by a leading capital letter: Div, Rot, and Grad. These three functions were joined by a 4-vector divergence marked by German typeface, $\mathfrak{Div}$. Sommerfeld chose to retain $\Box$ (cf. note 20), while noting the equivalence to his 4-vector functions: $\Box = \text{Div Grad}$. The principal advantage of the latter functions was that their meaning was familiar to physicists. In the same vein, Sommerfeld supplanted Minkowski's unwieldy terminology of "spacetime vectors of the first and second type" [*Raum-Zeit-Vektoren I^ter^ und II^ter^ Art*] with the more succinct "four-vector [*Vierervektor*] and "six-vector" [*Sechservektor*]. The result was a compact and transparent four-dimensional formalism differing as little as possible from the ordinary vector algebra employed in the physics volume of the *Encyklopädie*.[121]

To show how his formalism performed in action, Sommerfeld first took up the geometric interpretation and calculation of the electrodynamic 4-vector potential and 4-force. In the new notation, Sommerfeld wrote the electrodynamic 4-force $\mathfrak{K}$ between two point charges e and e_0 in terms of three components in the direction of the light-like 4-vector $\mathfrak{R}$, the source 4-velocity $\mathfrak{B}$, and the 4-acceleration $\dot{\mathfrak{B}}$:

120 See (Abraham 1910; Lewis 1910a, 1910b; Varičak 1910). On Varičak's contribution see (Walter 1999b).

121 Not all of Sommerfeld's notational choices were retained by later investigators; Laue, for instance, preferred a notational distinction between 4-vectors and 6-vectors. For a summary of notation used by Minkowski, Abraham, Lewis, and Laue, see (Reich 1994).

$$
\begin{aligned}
4\pi\mathfrak{K}_{\mathfrak{R}} &= \frac{ee_0}{c(\mathfrak{R}\mathfrak{B})^2}\left(\frac{c^2-(\mathfrak{R}\dot{\mathfrak{B}})}{(\mathfrak{R}\mathfrak{B})}(\mathfrak{B}_0\mathfrak{B})+\mathfrak{B}_0\dot{\mathfrak{B}}\right)\mathfrak{R},\\
4\pi\mathfrak{K}_{\mathfrak{B}} &= \frac{-ee_0}{c(\mathfrak{R}\mathfrak{B})^2}\frac{c^2-(\mathfrak{R}\dot{\mathfrak{B}})}{(\mathfrak{R}\mathfrak{B})}(\mathfrak{B}_0\mathfrak{R})\mathfrak{B},\\
4\pi\mathfrak{K}_{\dot{\mathfrak{B}}} &= \frac{-ee_0}{c(\mathfrak{R}\mathfrak{B})^2}\frac{c^2-(\mathfrak{R}\dot{\mathfrak{B}})}{(\mathfrak{R}\mathfrak{B})}(\mathfrak{B}_0\mathfrak{R})\dot{\mathfrak{B}},
\end{aligned}
\tag{42}
$$

where parentheses indicate scalar products. Sommerfeld was careful to note the equivalence between (42) and what he called Minkowski's "geometric rule" (40).

In the ninth and final section of his paper, Sommerfeld took up the law of electrostatics and the classical law of gravitation. The former was naturally considered to be a special case of (42), with two point charges relatively at rest. The same was true for the law of gravitation, as Sommerfeld noted that Minkowski had proposed a formal variant of (40) as a law of gravitational attraction (what I call Minkowski's second law, [41]). Sommerfeld's expression of the electrodynamic 4-force is unwieldy, but takes on a simpler form in case of uniform motion of the source ($\dot{\mathfrak{B}} = 0$). Neglecting the 4π factor, and substituting $-mm_0$ for $+ee_0$, Sommerfeld expressed the corresponding version of Minkowski's second law:

$$
-mm_0 c\frac{(\mathfrak{B}_0\mathfrak{B})\mathfrak{R}-(\mathfrak{B}_0\mathfrak{R})\mathfrak{B}}{(\mathfrak{R}\mathfrak{B})^3}. \tag{43}
$$

The latter law is compact and self-contained, in that its interpretation depends only on the definitions and rules of the algebraic formalism. In this sense, (43) improves on the Minkowskian (41), even if it represents only a special case of the latter law.

Once Sommerfeld had expressed Minkowski's second law in his own terms, he turned to Poincaré's two laws. The transformation of Poincaré's first law was more laborious than the transformation of Minkowski's second law. First of all, Sommerfeld transcribed Poincaré's first law (9) into his 4-vector notation, while retaining the original designation of invariants. This step itself was not simple: in order to cast Poincaré's kinematic invariants as scalar products of 4-vectors, Sommerfeld had to adjust the leading sign of (9), to obtain:

$$
\frac{k_0\mathfrak{K}}{mm'} = -\frac{1}{B^3C}\left(C\mathfrak{R}-\frac{1}{c}A\mathfrak{B}\right). \tag{44}
$$

Sommerfeld noted the "correction" of what he called an "obvious sign error" in (9).[122] The difference is due to Poincaré's irregular derivation of the kinematic invariants (1), as mentioned above (p. 203), although from Sommerfeld's remark it is not clear that he saw it this way.

122 "Mit Umkehr des bei Poincaré offenbar versehentlichen Vorzeichens" (Sommerfeld 1910b, 686, n. 1).

The transformation of Poincaré's second law (10) was less straightforward. It appears that instead of deriving a 4-vector expression as in the previous case, Sommerfeld followed Poincaré's lead by eliminating the Lorentz-invariant factor C from the denominator on the right-hand side of the first law (44), which results in the equation:

$$\frac{k_0\mathfrak{K}}{mm'} = -\frac{1}{B^3}\left(C\mathfrak{R} - \frac{1}{c}A\mathfrak{B}\right). \tag{45}$$

Sommerfeld expressed Poincaré's kinematic invariants A, B, and C as scalar products:

$$A = -\frac{1}{c}(\mathfrak{R}\mathfrak{B}_0), \quad B = -\frac{1}{c}(\mathfrak{R}\mathfrak{B}), \quad C = -\frac{1}{c^2}(\mathfrak{B}_0\mathfrak{B}). \tag{46}$$

He also replaced the mass term m' in (44) and (45) by the product of rest mass m_0 and the Lorentz factor k_0, i.e., $m' = m_0k_0$. At this point, he could express Poincaré's two laws exclusively in terms of constants, scalars, and 4-vectors:

$$mm_0c^3\frac{(\mathfrak{B}_0\mathfrak{B})\mathfrak{R} - (\mathfrak{B}_0\mathfrak{R})\mathfrak{B}}{(\mathfrak{R}\mathfrak{B})^3(\mathfrak{B}_0B)}, \tag{47}$$

$$-mm_0c\frac{(\mathfrak{B}_0\mathfrak{B})\mathfrak{R} - (\mathfrak{B}_0\mathfrak{R})\mathfrak{B}}{(\mathfrak{R}\mathfrak{B})^3}. \tag{48}$$

In the latter form, Sommerfeld's (approximate) version of Minkowski's second law (43) matches exactly his (exact) version of Poincaré's second law (48). Sommerfeld pointed out this equivalence, and noted again that the difference between (47) and (48) amounted to a single factor, in the scalar product of 4-velocities: $C = -(\mathfrak{B}_0\mathfrak{B})/c^2$. (All six Lorentz-invariant laws of gravitation of Poincaré, Minkowski, and Sommerfeld are presented in Table 1.) Sommerfeld summed up his result by saying that when the acceleration of the active mass is neglected, Minkowski's special formulation of Newton's law (41) is subsumed by Poincaré's indeterminate formulation. In other words, the approximate form of Minkowski's second law was captured by Poincaré's remark that his first law (9) could be multiplied by an unlimited number of Lorentz-invariant quantities (within certain constraints).

The message of the basic equivalence of Poincaré's pair of laws to Minkowski's pair echoes the latter's argument in his Cologne lecture, to the effect that spacetime mechanics removed the disharmonies of classical mechanics and electrodynamics (see above, p. 234). This message was reinforced by Sommerfeld's graphical representation of the 4-vector components of these laws in a spacetime diagram, reproduced in Fig. 4. The 4-vector relations in (47) and (48) are shown in the figure; the world line of the active mass m appears on the left-hand side of the diagram, and the line OL (which coincides with $\mathfrak{R}$) lies on the retrograde lightcone from the origin O on the world line of the passive mass m_0. All three 4-vectors in (47) and (48), $\mathfrak{R}$, $\mathfrak{B}_0$, and $\mathfrak{B}_0$ are represented in the diagram, along with an angle ψ corresponding to the Lorentz-invariant $C = \cos\psi$ distinguishing (47) and (48).[123]

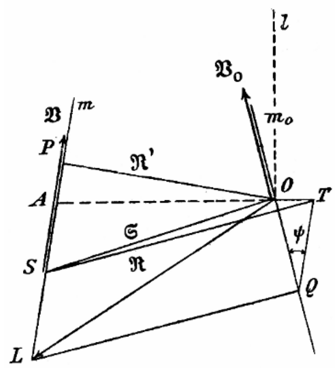


Figure 4: Sommerfeld's illustration of the two laws of gravitation (Sommerfeld 1910b, 687).

So far, Sommerfeld had dealt with three of the four laws of gravitation, leaving out only Minkowski's first law. Since Minkowski's presentation of his first law was a purely geometric affair, Sommerfeld had no choice but to reconstruct his argument with reference to a spacetime diagram describing the components of (27) in terms of the angle ψ and a fourth 4-vector, $\mathfrak{S}$. He showed the numerator in (47) and (48) to be equal to the product $(\mathfrak{B}_0\mathfrak{B})\mathfrak{S}$, and expressed the denominator of (48) in terms of the length R' of the 4-vector $\mathfrak{R}'$ in Fig. 4, to obtain the formula:

$$\mathfrak{K} = mm_0 \cos\psi \frac{\mathfrak{S}}{R'^3}, \tag{49}$$

which he showed to be equivalent to (47). Eliminating the factor $C = \cos\psi$ from the latter equation, Sommerfeld obtained an expression for (48) in terms of $\mathfrak{S}$:

$$\mathfrak{K} = \frac{mm_0\mathfrak{S}}{R'^3}. \tag{50}$$

The latter two driving force equations, (49) and (50), were thus rendered geometrically by Sommerfeld, facilitating the comprehension of their respective vector-symbolic expressions (47) and (48).

In general, the driving force of (49) is weaker, *ceteris paribus*, than that of (50) due to the cosine in the former, but Sommerfeld did not develop these results numerically, noting only that the four laws were equally valid from an empirical stand-

123 Sommerfeld explained Fig. 4 roughly as follows: two skew 4-velocities $\mathfrak{B}$ and $\mathfrak{B}_0$ determine a three-dimensional space, containing all the lines shown. Points $OLSAP$ are coplanar, while the triangles OQT and OTS, and the parallelogram $LQTS$ all generally lie in distinct planes. In particular, T lies outside the plane of $OLSAP$, and OT is orthogonal to $\mathfrak{B}_0$. The broken vertical line l represents the temporal axis of a frame with origin O; a space-like plane orthogonal to l at O intersects the world line of m at point A. The space-like 4-vector $\mathfrak{R}'$ is orthogonal to $\mathfrak{B}$, while $\mathfrak{S}$ is orthogonal to $\mathfrak{B}_0$; both $\mathfrak{R}'$ and $\mathfrak{S}$ intersect the origin, while $\mathfrak{B}$ and $\mathfrak{B}_0$ together form an angle ψ.

point.[124] He noted that Poincaré's analysis allowed for several other laws, but that in all cases, one sticking-point remained: there was no answer to the question of how to localize momentum in the gravitational field.

By rewriting Poincaré's and Minkowski's laws in his new 4-vector formalism, Sommerfeld effectively rationalized their contributions for physicists. The goal of his paper, announced at the outset, was to display the "remarkable simplification of electrodynamic concepts and calculations" resulting from "Minkowski's profound space-time conception."[125] Actually, Sommerfeld's comparison of Poincaré's and Minkowski's laws of gravitation was designed to show *his* formalism in an attractive light. In realizing this comparison in his own formalism, Sommerfeld smoothed out the idiosyncrasies of Poincaré's method, inappropriately lending him a 4-vector approach. He felt that Poincaré had "already employed 4-vectors" (Sommerfeld 1910b, 685) although as shown in the first section, Poincaré's use of four-dimensional entities was tightly circumscribed by the objective of formulating Lorentz-invariants. In Thomas Kuhn's optical metaphor (Kuhn 1970, 112), Sommerfeld read Poincaré's theory through a Minkowskian lens; in other words, he read it as a space-time theory. For Sommerfeld, no less than for Minkowski, the discussion of gravitation and relativity was modulated by the programmatic objective of promoting a four-dimensional formalism. Satisfying this objective without ignoring Poincaré's work, however, meant rationalizing Poincaré's contribution.[126]

Sommerfeld's reading of Minkowski's second law contrasts with its muted exposition in the original text (see above, pp. 234–234), in that he gave it pride of place with respect to the other three laws. This change in emphasis on Sommerfeld's part reflects his own research interests in electrodynamics, and his outlook on the future direction of physics.[127] But what originally motivated him to propose a 4-dimensional formalism? The inevitability of a 4-dimensional vector algebra as a standard tool of the physicist was probably a foregone conclusion for him by 1910, such that the promotion of the ordinary vector notation used in the *Encyklopädie* obliged him to propose essentially the same notation for 4-vectors. Sommerfeld referred modestly to his work as an "explanation of Minkowskian ideas" (Sommerfeld 1910a, 749) but as he explained to his friend Willy Wien, co-editor with Planck of the *Annalen der*

---

124 This view was confirmed independently by the Dutch astronomer W. de Sitter, who worked out the numbers for the one-body problem (de Sitter 1911). De Sitter found the second law to require a post-Newtonian centennial advance in Mercury's perihelion of $7''$, while the first law required no additional advance. His figure for the second law agrees with the one given by Poincaré (see above, p. 208).

125 "In dieser und einigen anschließenden Studien möchte ich darstellen, wie merkwürdig sich die elektrodynamischen Begriffe und Rechnungen vereinfachen, wenn man sich dabei von der tiefsinnigen Raum-Zeit-Auffassung Minkowskis leiten läßt" (Sommerfeld 1910a, 749).

126 Faced with a similar situation in his Cologne lecture of September, 1908, Minkowski simply neglected to mention Poincaré's contribution; see (Walter 1999a, 56).

127 Sommerfeld later preferred Gustav Mie's field theory of gravitation. Such an approach was more promising than that of Poincaré and Minkowski, which grasped gravitation "to some extent as action at a distance" (Sommerfeld 1913, 73).

Physik, Minkowski's original 4-vector scheme had evolved. "The geometrical systematics" Sommerfeld announced, "is now hyper-Minkowskian."[128] In the same letter to Wien, Sommerfeld confessed that his paper had required substantial effort, and he expressed doubt that it would prove worthwhile. Sommerfeld displayed either pessimism or modesty here, but in fact his effort was richly rewarded, as his streamlined four-dimensional algebra and analysis quickly won both Einstein's praise and the confidence of his contemporaries.[129]

Sommerfeld's work was eagerly read by young theoretical physicists raised in the heady atmosphere of German vectorial electrodynamics. One of the early adepts of Sommerfeld's formalism was Philipp Frank (1884–1966), who was then a Privatdozent in Vienna. By way of introduction to his 1911 study of the Lorentz-covariance of Maxwell's equations, Frank described the new four-dimensional algebra as a combination of "Sommerfeld's intuitiveness with Minkowski's mathematical elegance" (Frank 1911, 600). He recognized, however, that of late, physicists had been overloaded with outlandish symbolic systems and terminology, and promised to stay within the boundaries of Sommerfeld's system, at least as far as this was possible.

Physicists were indeed inundated in 1910–1911 with a bewildering array of new symbolic systems, including an ordinary vector algebra (Burali-Forti and Marcolongo 1910), and a quaternionic calculus (Conway 1911), in addition to the hyperbolic-coordinate system and three 4-vector formalisms already mentioned. By 1911, 4-vector and 6-vector operations featured prominently in the pages of the *Annalen der Physik*. Out of the nine theoretical papers concerning relativity theory published in the *Annalen* that year, five made use of a four-dimensional approach to physics, either in terms of 4-vector operations, or by referring to spacetime coordinates. Four out of five authors of "four-dimensional" papers cited Minkowski's or Sommerfeld's work; the fifth referred to Max Laue's new relativity textbook (Laue 1911). This timely and well-written little book went far in standardizing the terminology and notation of four-dimensional algebra, such that by January of 1912, Max Abraham preferred the Sommerfeld-Laue notation to his own for the exposition of his theory of gravitation (Abraham 1910, 1912a, 1912b).

While young theorists were quick to pick up on the Sommerfeld-Laue calculus, textbook writers did not follow the trend. Of the four textbooks to appear on relativity in 1913–1914, only the second edition of Laue's book (Laue 1913) employed this formalism. Ebenezer Cunningham presented a 4-dimensional approach based on Minkowski's work, but explicitly rejected Sommerfeld's "quasi-geometrical language", which conflicted with his own purely algebraic presentation (Cunningham

128 "Die geometrische Systematik ist jetzt hyper-minkowskisch" (Sommerfeld to Wien, 11 July, 1910, Sommerfeld 2001–2004, 1:388).

129 Einstein to Sommerfeld, July, 1910, (CPAE 5, 243–247; Sommerfeld 2001–2004, 1:386–388). In light of Einstein and Laub's earlier dismissal of Minkowski's formalism (see above, pp. 222–223), Sommerfeld naturally supposed that Einstein would disapprove of his system, prompting the protest: "Wie können Sie denken, dass ich die Schönheit einer solchen Untersuchung nicht zu schätzen wüsste?"

1914, 99). A third textbook by Ludwik Silberstein (Silberstein 1914), a former student of Planck, gave preference to a quaternionic presentation, while the fourth, by Max B. Weinstein (Weinstein 1913), opted for Cartesian coordinates. Curiously enough, Weinstein dedicated his work to the memory of Minkowski. Apparently disturbed by this profession of fidelity, Max Born, who had briefly served as Minkowski's assistant, deplored the form of Weinstein's approach to relativity:

> [Minkowski] put perhaps just as much value on his presentation as on its content. For this reason, I do not believe that entrance to his conceptual world is facilitated when it is overwhelmed by an enormous surfeit of formulas.[130]

By this time, Born himself had dropped Minkowski's formalism in favor of the Sommerfeld-Laue approach, such that the target of his criticism was Weinstein's disregard for 4-dimensional methods in general, and not the neglect of Minkowski's matrix calculus.[131] What Born was pointing out here was that it had become highly impractical to study the theory of relativity without recourse to a 4-dimensional formalism. This may explain why Laue's was the only one of the four textbooks on relativity to be reedited, reaching a sixth edition in 1955.

In summary, the language developed by Sommerfeld for the expression of the laws of gravitation of Poincaré and Minkowski endured, while the laws themselves remained tentative at best. This much was clear as early as 1912, when Jun Ishiwara reported from Japan on the state of relativity theory. This theory, Ishiwara felt, had shed no light on the problem of gravitation, with a single exception: Minkowski and Sommerfeld's "formal mathematical treatment" (Ishiwara 1912, 588). The trend from Poincaré to Sommerfeld was one of increasing reliance on formal techniques catering to Lorentz-invariance; in the space of five years, the physical content of the laws of gravitation remained stable, while their formal garb evolved from Cartesian to hyper-Minkowskian.

4. CONCLUSION: ON THE EMERGENCE OF THE FOUR-DIMENSIONAL VIEW

After a century-long process of accommodation to the use of tensor calculus and spacetime diagrams for analysis of physical interactions, the mathematical difficulties encountered by the pioneers of 4-dimensional physics are hard to come to terms with. Not only is the oft-encountered image of flat-spacetime physics as a trivial consequence of Einstein's special theory of relativity and Felix Klein's geometry consistent with such accommodation, it reflects Minkowski's own characterization of the back-

130 "[Minkowski] hat auf seine Darstellung vielleicht ebenso viel Wert gelegt, wie auf ihren Inhalt. Darum glaube ich nicht, daß der Zugang zu seiner Gedankenwelt erleichtert wird, wenn sie von einer ungeheuren [sic] Fülle von Formeln überschüttet wird" (Born 1914).

131 By the end of 1911 Born had already acknowledged that, despite its "formal simplicity and greater generality compared to the tradition of vectorial notation," Minkowski's calculus was "unable to hold its ground in mathematical physics" (Born 1912, 175).

ground of the four-dimensional approach (cf. p. 218). However, this description ought not be taken at face value, being better understood as a rhetorical ploy designed to induce mathematicians to enter the nascent field of relativistic physics (Walter 1999a). When the principle of relativity was formulated in 1905, even for one as adept as Henri Poincaré in the application of group methods, the path to a four-dimensional language for physics appeared strewn with obstacles. Much as Poincaré had predicted (above, p. 209), the construction of this language cost Minkowski and Sommerfeld considerable pain and effort.

Clear-sighted as he proved to be in this regard, Poincaré did not foresee the emergence of forces that would accelerate the construction and acquisition of a four-dimensional language. With hindsight, we can identify five factors favoring the use and development of a four-dimensional language for physics between 1905 and 1910: the elaboration of new concepts and definitions, the introduction of a graphic model of spacetime, the experimental confirmation of relativity theory, the vector-symbolic movement, and problem-solving performance.

In the beginning, the availability of workable four-dimensional concepts and definitions regulated the analytic reach of a four-dimensional approach to physics. Poincaré's discovery of the 4-vectors of velocity and force in the course of his elaboration of Lorentz-invariant quantities, and Minkowski's initial misreading of Poincaré's definitions underline how unintuitive these notions appeared to turn-of-the-century mathematicians. The lack of a 4-velocity definition visibly hindered Minkowski's elaboration of spacetime mechanics and theory of gravitation. It is remarkable that even after Minkowski presented the notions of proper time, world line, rest-mass density, and the energy-momentum tensor, putting the spacetime electrodynamics and mechanics on the same four-dimensional footing, his approach failed to convince physicists. Nevertheless, all of these discoveries extended the reach of the four-dimensional approach, in the end making it a viable candidate for the theorist's toolbox.

Next, Minkowski's visually-intuitive spacetime diagram played a decisive role in the emergence of the four-dimensional view. While the spacetime diagram reflects some of the concepts mentioned above, its utility as a cognitive tool exceeded by far that of the sum of its parts. In Minkowski's hands, the spacetime diagram was more than a tool, it was a model used to present both of his laws of gravitation. Beyond their practical function in problem-solving, spacetime diagrams favored the diffusion in wider circles of both the theory of relativity and the four-dimensional view of this theory, in particular among non-mathematicians, by providing a visually intuitive means of grasping certain consequences of the theory of relativity, such as time dilation and Lorentz contraction. Minkowski's graphic model of spacetime thus enhanced both formal and intuitive approaches to special relativity.

In the third place, the ultimate success of the four-dimensional view hinged on the empirical adequacy of the theory of relativity. It is remarkable that the conceptual groundwork, and much of the formal elaboration of the four-dimensional view was accomplished during a time when the theory of relativity was less well corroborated by experiment than its rivals. The reversal of this situation in favor of relativity theory in late 1908 favored the reception of the existing four-dimensional methods, and provided new impetus both for their application and extension, and for the development of alternatives, such as that of Sommerfeld.

The fourth major factor influencing the elaboration of a four-dimensional view of physics was the vector-symbolic movement in physics and mathematics at the turn of the twentieth century (McCormmach 1976, xxxi). The participants in this movement, in which Sommerfeld was a leading figure, believed in the efficacy of vector-symbolic methods in physics and geometry, and sought to unify the plethora of notations employed by various writers. The movement's strength varied from country to country; it was largely ignored in France, for example, in favor of the coordinate-based notation favored by Poincaré and others. Poincaré's pronounced disinterest in the application and development of a four-dimensional calculus for physics was typical of contemporary French attitudes toward vector-symbolic methods. In Germany, on the other hand, electrodynamicists learned Maxwell's theory from the mid-1890s in terms of curl $\mathfrak{h}$ and div $\mathfrak{E}$. In Zürich and Göttingen during this period, Minkowski instructed students (including Einstein) in the ways of the vector calculus. Unlike Poincaré, Minkowski was convinced that a four-dimensional language for physics would be worth the effort spent on its elaboration, yet he ultimately abandoned the vector-symbolic model in favor of an elegant and sophisticated matrix calculus. This choice was deplored by physicists (including Einstein), and mooted by Sommerfeld's conservative extension of the standard vector formalism into an immediately successful 4-vector algebra and analysis. In sum, the vector-symbolic movement functioned alternatively as an accelerator of the elaboration of four-dimensional calculi (existing systems served as templates), and as a regulator (penalizing Minkowski's neglect of standard vector operations).

The fifth and final parameter affecting the emergence of the four-dimensional view of physics was problem-solving performance. From the standpoint of ease of calculation, any four-dimensional vector formalism at all compared well to a Cartesian-coordinate approach, as Weinstein's textbook demonstrated; the advantage of ordinary vector methods over Cartesian coordinates was less pronounced. As we have seen, Poincaré applied his approach to the problem of constructing a Lorentz-invariant law of gravitational attraction, and was followed in turn by Minkowski and Sommerfeld, both of whom also provided examples of problem-solving. In virtue of the clarity and order of Sommerfeld's detailed, coordinate-free comparison of the laws of gravitation of Poincaré and Minkowski, his 4-vector algebra appeared to be the superior four-dimensional approach, just when physicists and mathematicians were turning to relativity in greater numbers.

ACKNOWLEDGMENTS

This study was inaugurated with the encouragement and support of Jürgen Renn, during a stay at the Max Planck Institute for the History of Science (Berlin) in 1998. I am especially grateful to Urs Schoepflin and the Institute's library staff for their expert assistance. The themes explored here were presented at the Mathematisches Forschunginstitut Oberwolfach (January 2000), the University of Heidelberg (June 2000), the University of Paris 7–REHSEIS (April, 2001), the joint AMS/SMF meeting in Lyons (July, 2001), and the Eighth International Conference on the History of General Relativity (Amsterdam, June 2002). Several scholars shared their ideas with me, shaping the final form of the paper. In particular, I am indebted to Olivier Darrigol for insightful comments on an early draft. The paper has been improved thanks to a careful reading by Shaul Katzir, and discussions with John Norton and Philippe Lombard. The responsibility for any remaining infelicities is my own.

Table 1: Lorentz-Invariant Laws of Gravitation, 1906–1910

Poincaré (1906)[a]	**Minkowski (1908)**[b]	**Sommerfeld (1910)**[c]
$X_1 = \frac{x}{k_0 B^3} - \xi_1 \frac{k_1}{k_0}\frac{A}{B^3 C}$ $Y_1 = \frac{y}{k_0 B^3} - \eta_1 \frac{k_1}{k_0}\frac{A}{B^3 C}$ $Z_1 = \frac{z}{k_0 B^3} - \zeta_1 \frac{k_1}{k_0}\frac{A}{B^3 C}$ $T_1 = -\frac{r}{k_0 B^3} - \frac{k_1}{k_0}\frac{A}{B^3 C}$	$mm^*\left(\frac{OA'}{B^*D^*}\right)^3 BD^*$	$mm_0 c^3 \frac{(\mathfrak{B}_0\mathfrak{B})\mathfrak{R} - (\mathfrak{B}_0\mathfrak{R})\mathfrak{B}}{(\mathfrak{R}\mathfrak{B})^3(\mathfrak{B}_0\mathfrak{B})}$
$X_1 = \frac{\lambda}{B^3} - \frac{\eta\nu' - \zeta\mu'}{B^3}$ $Y_1 = \frac{\mu}{B^3} - \frac{\zeta\lambda' - \xi\nu'}{B^3}$ $Z_1 = \frac{\nu}{B^3} - \frac{\xi\mu' - \eta\lambda'}{B^3}$	$mm_1\left(\dot{t}_1 - \frac{\dot{x}_1}{c}\right)\mathfrak{K}$	$-mm_0 c \frac{(\mathfrak{B}_0\mathfrak{B})\mathfrak{R} - (\mathfrak{B}_0\mathfrak{R})\mathfrak{B}}{(\mathfrak{R}\mathfrak{B})^3}$

[a] Mass terms are neglected, such that the right-hand side of each equation is implicitly multiplied by the product of the two masses. When both sides of the four equations are multiplied by the factor k_0 they express components of a 4-vector, $k_0(X_1, Y_1, Z_1, iT_1)$. The constants k_0 and k_1 are defined as: $k_0 = 1/\sqrt{1-\Sigma\xi^2}$ and $k_1 = 1/\sqrt{1-\Sigma\xi_1^2}$. A, B, and C denote the last three Lorentz-invariants in (1): $A = \frac{t - \Sigma x\xi}{\sqrt{1-\Sigma\xi^2}}$, $B = \frac{t - \Sigma x\xi_1}{\sqrt{1-\Sigma\xi_1^2}}$, $C = \frac{1 - \Sigma\xi\xi_1}{\sqrt{(1-\Sigma\xi^2)(1-\Sigma\xi_1^2)}}$, where $\Sigma\xi$ and $\Sigma\xi_1$ designate the

ordinary velocities of the passive and active mass points, with components ξ, η, ζ, and ξ_1, η_1, ζ_1. The time t is set equal to the negative distance between the passive mass point and the retarded position of the active mass point, $t = -\sqrt{\Sigma x^2} = -r$. Poincaré's second law is shown in the bottom row; he neglected to write the fourth component T_1, determined from the first three by the orthogonality condition $T_1 = \Sigma X_1 \xi$. The new variables in the bottom row are:

$$\lambda = k_1(x + r\xi_1), \qquad \mu = k_1(y + r\eta_1), \qquad \nu = k_1(z + r\zeta_1),$$
$$\lambda' = k_1(\eta_1 z - \zeta_1 y), \quad \mu' = k_1(\zeta_1 x - \xi_1 z), \quad \nu' = k_1(\xi_1 y - x\eta_1).$$

[b] The formula in the top row describes the first three components of the driving force: the fourth component is obtained analytically. The constants m and m^* designate the passive and active proper mass, respectively, while the remaining letters stand for spacetime points, as reconstructed in Fig. 1 (p. 227). The formula in the bottom row represents the driving force of gravitation as described, but not formally expressed, in (Minkowski 1909). The constants m and m_1 designate the passive and active proper mass, $\dot{t}_1$ and $\dot{x}_1$ are 4-velocity components of the passive mass, c is the speed of light and $\mathfrak{K}$ is a 4-vector, for the definition of which see note 111.

[c] The constants m_0 and m designate the passive and active proper mass, respectively, c denotes the speed of light, $\mathfrak{B}_0$ and $\mathfrak{B}$ represent the corresponding 4-velocities, and $\mathfrak{R}$ stands for the light-like interval between the mass points.

REFERENCES

Abraham, Max. 1905. *Elektromagnetische Theorie der Strahlung*. (*Theorie der Elektrizität*, vol. 2.) Leipzig: Teubner.

———. 1909. "Zur elektromagnetischen Mechanik." *Physikalische Zeitschrift* 10: 737–741.

———. 1910. "Sull' elettrodinamica di Minkowski." *Rendiconti del Circolo Matematico di Palermo* 30: 33–46.

———. 1912a. "Das Elementargesetz der Gravitation." *Physikalische Zeitschrift* 13: 4–5.

———. 1912b. "Zur Theorie der Gravitation." *Physikalische Zeitschrift* 13: 1–4. (English translation in this volume.)

———. 1914. "Die neue Mechanik." *Scientia* (*Rivista di Scienza*) 15: 8–27.

Andrade Martins, Roberto de. 1999. "The search for gravitational absorption in the early 20th century." In (Goenner et al. 1999), 3–44.

Arzeliès, Henri, and J. Henry. 1959. *Milieux conducteurs ou polarisables en mouvement*. Paris: Gauthier-Villars.

Barrow-Green, June E. 1997. *Poincaré and the Three Body Problem*. (*History of Mathematics*, vol. 11.) Providence: AMS and LMS.

Bork, Alfred M. 1966. "'Vectors versus quaternions'—the letters in *Nature*." *American Journal of Physics* 34: 202–211.

Born, Max. 1909. "Die Theorie des starren Elektrons in der Kinematik des Relativitätsprinzips." *Annalen der Physik* 30: 1–56.

———. 1912. "Besprechung von Max Laue, Das Relativitätsprinzip." *Physikalische Zeitschrift* 13: 175–176.

———. 1914. "Besprechung von Max Weinstein, Die Physik der bewegten Materie und die Relativitätstheorie." *Physikalische Zeitschrift* 15: 676.

Buchwald, Jed Z. 1985. *From Maxwell to Microphysics*. Chicago: University of Chicago Press.

Buchwald, Jed Z., and Andrew Warwick. 2001. *Histories of the Electron: The Birth of Microphysics*. Dibner Institute Studies in the History of Science and Technology. Cambridge MA: MIT Press.

Burali-Forti, Cesare, and Roberto Marcolongo. 1910. *Éléments de calcul vectoriel*. Paris: Hermann.

Cantor, Geoffrey N., and Michael J. S. Hodge. 1981. *Conceptions of Ether: Studies in the History of Ether Theories 1740–1900*. Cambridge: Cambridge University Press.

Carazza, Bruno, and Helge Kragh. 1990. "Augusto Righi's magnetic rays: a failed research program in early 20th-century physics." *Historical Studies in the Physical and Biological Sciences* 21: 1–28.

Cayley, Arthur. 1869. “On the six coordinates of a line.” *Transactions of the Cambridge Philosophical Society* 11: 290–323.

Conway, Arthur W. 1911. “On the application of quaternions to some recent developments of electrical theory.” *Proceedings of the Royal Irish Academy* 29: 1–9.

Corry, Leo. 1997. “Hermann Minkowski and the postulate of relativity.” *Archive for History of Exact Sciences* 51: 273–314.

———. 2004. *David Hilbert and the Axiomatization of Physics, 1898–1918: From “Grundlagen der Geometrie” to “Grundlagen der Physik”*. Dordrecht: Kluwer.

Coster, H. G. L. and J. R. Shepanski. 1969. “Gravito-inertial fields and relativity.” *Journal of Physics A* 2: 22–27.

CPAE 2. 1989. John Stachel, David C. Cassidy, Jürgen Renn, and Robert Schulmann (eds.), *The Collected Papers of Albert Einstein*. Vol. 2. *The Swiss Years: Writings, 1900–1909*. Princeton: Princeton University Press.

CPAE 5. 1993. Martin J. Klein, A. J. Kox, and Robert Schulmann (eds.), *The Collected Papers of Albert Einstein*. Vol. 5. *The Swiss Years: Correspondence, 1902–1914*. Princeton: Princeton University Press.

Cunningham, Ebenezer. 1914. *The Principle of Relativity.* Cambridge: Cambridge University Press.

Cuvaj, Camillo. 1968. “Henri Poincaré’s mathematical contributions to relativity and the Poincaré stresses.” *American Journal of Physics* 36: 1102–1113.

———. 1970. *A History of Relativity: The Role of Henri Poincaré and Paul Langevin*. Ph.D. dissertation, Yeshiva University.

Darrigol, Olivier. 1993. “The electrodynamic revolution in Germany as documented by early German expositions of ‘Maxwell’s theory’.” *Archive for History of Exact Sciences* 45: 189–280.

———. 1994. “The electron theories of Larmor and Lorentz: a comparative study.” *Historical Studies in the Physical and Biological Sciences* 24: 265–336.

———. 2000. *Electrodynamics from Ampère to Einstein*. Oxford: Oxford University Press.

Dugac, Pierre. 1986. “La correspondance d’Henri Poincaré avec des mathématiciens de A à H.” *Cahiers du séminaire d’histoire des mathématiques* 7: 59–219.

———. 1989. “La correspondance d’Henri Poincaré avec des mathématiciens de J à Z.” *Cahiers du séminaire d’histoire des mathématiques* 10: 83–229.

Dziobek, Otto F. 1888. *Die mathematischen Theorien der Planeten-Bewegungen*. Leipzig: Barth.

Einstein, Albert. 1905. “Zur Elektrodynamik bewegter Körper.” *Annalen der Physik* 17: 891–921, (CPAE 2, Doc. 23).

———. 1907. “Relativitätsprinzip und die aus demselben gezogenen Folgerungen.” *Jahrbuch der Radioaktivität und Elektronik* 4: 411–462, (CPAE 2, Doc. 47).

Einstein, Albert, and Jakob J. Laub. 1908. “Über die elektromagnetischen Grundgleichungen für bewegte Körper.” *Annalen der Physik* 26: 532–540, (CPAE 2, Doc. 51).

Frank, Philipp. 1911. “Das Verhalten der elektromagnetischen Feldgleichungen gegenüber linearen Transformationen.” *Annalen der Physik* 35: 599–607.

Föppl, August O. 1894. *Einführung in die Maxwell’sche Theorie der Elektricität*. Leipzig: Teubner.

Galison, Peter. 1979. “Minkowski’s spacetime: from visual thinking to the absolute world.” *Historical Studies in the Physical Sciences* 10: 85–121.

Gans, Richard. 1905. “Gravitation und Elektromagnetismus.” *Jahresbericht der deutschen Mathematiker-Vereinigung* 14: 578–581.

Gibbs, Josiah W. and Edwin B. Wilson. 1901. *Vector Analysis*. New York: Charles Scribner’s Sons.

Gispert, Hélène. 2001. “The German and French editions of the Klein-Molk Encyclopedia: contrasted images.” In Umberto Bottazzini and Amy Dahan Dalmedico (eds.), *Changing Images in Mathematics: From the French Revolution to the New Millennium*, 93–112. *Studies in the History of Science, Technology and Medicine* 13. London: Routledge.

Goenner, Hubert, Jürgen Renn, Tilman Sauer, and Jim Ritter (eds.). 1999. *The Expanding Worlds of General Relativity.* (*Einstein Studies* vol. 7.) Boston/Basel/Berlin: Birkhäuser.

Gray, Jeremy. 1992. “Poincaré and the solar system.” In Peter M. Harman and Alan E. Shapiro (eds.), *The Investigation of Difficult Things*, 503–524. Cambridge: Cambridge University Press.

Hargreaves, Richard. 1908. “Integral forms and their connexion with physical equations.” *Transactions of the Cambridge Philosophical Society* 21: 107–122.

Harvey, A. L. 1965. “A brief review of Lorentz-covariant theories of gravitation.” *American Journal of Physics* 33: 449–460.

Havas, Peter 1979. “Equations of motion and radiation reaction in the special and general theory of relativity.” In Jürgen Ehlers (ed.), *Isolated Gravitating Systems in General Relativity*, 74–155. Proceedings of the International School of Physics “Enrico Fermi” 67. Amsterdam: North-Holland.

Heaviside, Oliver. 1893. “A gravitational and electromagnetic analogy.” In *Electromagnetic Theory,* 3 vols., 1: 455–464. London: The Electrician.

Hon, Giora. 1995. "The case of Kaufmann's experiment and its varied reception." In Jed Z. Buchwald (ed.), *Scientific Practice: Theories and Stories of Doing Physics*, 170–223. Chicago: University of Chicago Press.

Ishiwara, Jun. 1912. "Bericht über die Relativitätstheorie." *Jahrbuch der Radioaktivität und Elektronik* 9: 560–648.

Jackson, J. David. 1975. *Classical Electrodynamics*. New York: Wiley, 2nd edition.

Jungnickel, Christa and Russell McCormmach. 1986. *Intellectual Mastery of Nature: Theoretical Physics from Ohm to Einstein*. Chicago: University of Chicago Press.

Klein, Felix. 1907. *Vorträge über den mathematischen Unterricht an den höheren Schulen*, Rudolf Schimmack (ed.). (*Mathematische Vorlesungen an der Universität Göttingen,* vol. 1.) Leipzig: Teubner.

Klein, Felix, and Arnold Sommerfeld. 1897–1910. *Über die Theorie des Kreisels*. 4 vols. Leipzig: Teubner.

Kottler, Felix. 1922. "Gravitation und Relativitätstheorie." In Karl Schwarzschild, Samuel Oppenheim, and Walther von Dyck (eds.), *Astronomie*, 2 vols, 2: 159–237. *Encyklopädie der mathematischen Wissenschaften mit Einschluss ihrer Anwendungen* 6. Leipzig: Teubner.

Kretschmann, Erich. 1914. *Eine Theorie der Schwerkraft im Rahmen der ursprünglichen Einsteinschen Relativitätstheorie*. Ph.D. dissertation, University of Berlin.

Krätzel, Ekkehard. 1989. "Kommentierender Anhang: zur Geometrie der Zahlen." In Ekkehard Krätzel and Bernulf Weissbach (eds.), *Ausgewählte Arbeiten zur Zahlentheorie und zur Geometrie*, 233–246. *Teubner-Archiv zur Mathematik* 12. Leipzig: Teubner.

Kuhn, Thomas S. 1970. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press, 2nd edition.

Laue, Max von. 1911. *Das Relativitätsprinzip*. (*Die Wissenschaft*, vol. 38.) Braunschweig: Vieweg.

———. 1913. *Das Relativitätsprinzip*. (*Die Wissenschaft*, vol. 38.) Braunschweig: Vieweg, 2nd edition.

———. 1951. "Sommerfelds Lebenswerk." *Naturwissenschaften* 38: 513–518.

Lewis, Gilbert N. 1910a. "On four-dimensional vector analysis, and its application in electrical theory." *Proceedings of the American Academy of Arts and Science* 46: 165–181.

———. 1910b. "Über vierdimensionale Vektoranalysis und deren Anwendung auf die Elektrizitätstheorie." *Jahrbuch der Radioaktivität und Elektronik* 7: 329–347.

Liu, Chuang. 1991. *Relativistic Thermodynamics: Its History and Foundations*. Ph.D. dissertation, University of Pittsburgh.

Lorentz, Hendrik A. 1900. "Considerations on Gravitation." *Proceedings of the Section of Sciences. Verslag Koninklijke Akademie van Wetenschapen* 2: 559–574. (Printed in this volume.)

———. 1904a. "Electromagnetic phenomena in a system moving with any velocity less than that of light." *Proceedings of the Section of Sciences. Verslag Koninklijke Akademie van Wetenschapen* 6: 809–831.

———. 1904b. "Maxwells elektromagnetische Theorie." In (Sommerfeld 1903–1926), 2: 63–144.

———. 1904c. "Weiterbildung der Maxwellschen Theorie; Elektronentheorie." In (Sommerfeld 1903–1926), 2: 145–280.

———. 1910. "Alte und neue Fragen der Physik." *Physikalische Zeitschrift* 11: 1234–1257. (English translation in this volume.)

———. 1914. "La gravitation." *Scientia* (*Rivista di Scienza*) 16: 28–59.

Lützen, Jesper. 1999. "Geometrising configurations: Heinrich Hertz and his mathematical precursors." In Jeremy Gray (ed.), *The Symbolic Universe: Geometry and Physics, 1890–1930*, 25–46. Oxford: Oxford University Press.

Maltese, Giulio. 2000. "The late entrance of relativity into Italian scientific community (1906–1930)." *Historical Studies in the Physical and Biological Sciences* 31: 125–173.

Manegold, Karl-Heinz. 1970. *Universität, Technische Hochschule und Industrie*. Berlin: Duncker & Humblot.

Marcolongo, Roberto. 1906. "Sugli integrali delle equazioni dell'elettrodinamica." *Rendiconti della Reale Accademia dei Lincei* 15: 344–349.

McCormmach, Russell. 1976. "Editor's foreword." *Historical Studies in the Physical Sciences* 7: xi–xxxv.

Miller, Arthur I. 1973. "A Study of Henri Poincaré's 'Sur la dynamique de l'électron'." *Archive for History of Exact Sciences* 10: 207–328.

———. 1981. *Albert Einstein's Special Theory of Relativity: Emergence (1905) and Early Interpretation*. Reading, MA: Addison-Wesley.

Minkowski, Hermann. 1888. "Über die Bewegung eines festen Körpers in einer Flüssigkeit." *Sitzungsberichte der königliche preußischen Akademie der Wissenschaften* 40: 1095–1110.

———. 1890–1893. "H. Poincaré, Sur le problème des trois corps et les équations de la dynamique." *Jahrbuch über die Fortschritte der Mathematik* 22: 907–914.

———. 1896. *Geometrie der Zahlen*. Leipzig: Teubner.

———. 1907. "Kapillarität." In (Sommerfeld 1903–1926), 1: 558–613.

———. 1908. "Die Grundgleichungen für die electromagnetischen Vorgänge in bewegten Körpern." *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen*, 53–111. (English translation of the appendix "Mechanics and the Relativity Postulate" in this volume.)

———. 1909. "Raum und Zeit." *Jahresbericht der deutschen Mathematiker-Vereinigung* 18: 75–88.

———. 1915. "Das Relativitätsprinzip." *Jahresbericht der deutschen Mathematiker-Vereinigung* 24: 372–382.

———. 1973. *Briefe an David Hilbert*. Lily Rüdenberg and Hans Zassenhaus (eds.). Berlin: Springer-Verlag.

Minkowski, Hermann, and Max Born. 1910. "Eine Ableitung der Grundgleichungen für die elektromagnetischen Vorgänge in bewegten Körpern." *Mathematische Annalen* 68: 526–550.

Møller, Christian. 1972. *The Theory of Relativity.* Oxford: Oxford University Press, 2nd edition.

North, John D. 1965. *The Measure of the Universe: A History of Modern Cosmology.* Oxford: Oxford University Press.

Norton, John D. 1992. "Einstein, Nordström and the early demise of Lorentz-covariant theories of gravitation." *Archive for History of Exact Sciences* 45: 17–94.

Pauli, Wolfgang. 1921. "Relativitätstheorie." In (Sommerfeld 1903–1926), 2: 539–775.

Planck, Max. 1906. "Das Prinzip der Relativität und die Grundgleichungen der Mechanik." *Verhandlungen der Deutschen Physikalischen Gesellschaft* 8: 136–141.

———. 1907. "Zur Dynamik bewegter Systeme." *Sitzungsberichte der königliche preußischen Akad. der Wiss.*: 542–570.

Poincaré, Henri. 1885. "Sur l'équilibre d'une masse fluide animée d'un mouvement de rotation." *Acta mathematica* 7: 259–380.

———. 1895. *Capillarité*. J. Blondin (ed.). Paris: Georges Carré.

———. 1898–1905. "Préface." In Charles Hermite, Henri Poincaré, and Eugène Rouché (eds.), *Œuvres de Laguerre* 1: v–xv. Paris: Gauthier-Villars.

———. 1901. *Électricité et optique: la lumière et les théories électrodynamiques*. Jules Blondin and Eugène Néculcéa (eds.). Paris: Carré et Naud.

———. 1902a. *Figures d'équilibre d'une masse fluide*. Léon Dreyfus (ed.). Paris: C. Naud.

———. 1902b. "Sur la stabilité de l'équilibre des figures piriformes affectées par une masse fluide en rotation." *Philosophical Transactions of the Royal Society* A 198: 333–373.

———. 1904. "L'état actuel et l'avenir de la physique mathématique." *Bulletin des sciences mathématiques* 28: 302–324.

———. 1906. "Sur la dynamique de l'électron." *Rendiconti del Circolo Matematico di Palermo* 21: 129–176. (English translation of excerpt in this volume.)

———. 1907. "La relativité de l'espace". *Année psychologique* 13: 1–17.

———. 1908. "La dynamique de l'électron." *Revue générale des sciences pures et appliquées* 19: 386–402.

———. 1909. "La mécanique nouvelle." *Revue scientifique* 12: 170–177.

———. 1910. *Sechs Vorträge über ausgewählte Gegenstände aus der reinen Mathematik und mathematischen Physik*. (*Mathematische Vorlesungen an der Universität Göttingen*, vol. 4.) Leipzig/Berlin: Teubner.

———. 1912. "L'espace et le temps." *Scientia* (*Rivista di Scienza*) 12: 159–170.

———. 1953. "Les limites de la loi de Newton." *Bulletin astronomique* 17: 121–269.

Pomey, Jean-Baptiste. 1914–1931. *Cours d'électricité théorique*, 3 vols. Bibliothèque des Annales des Postes, Télégraphes et Téléphones. Paris: Gauthier-Villars.

Pyenson, Lewis. 1973. *The Goettingen Reception of Einstein's General Theory of Relativity.* Ph.D. dissertation, Johns Hopkins University.

———. 1985. *The Young Einstein: The Advent of Relativity.* Bristol: Hilger.

Reich, Karin. 1994. *Die Entwicklung des Tensorkalküls: vom absoluten Differentialkalkül zur Relativitätstheorie*. (*Science Networks Historical Studies*, vol. 11.) Basel/Boston: Birkhäuser.

———. 1996. "The emergence of vector calculus in physics: the early decades." In Gert Schubring (ed.), *Hermann Günther Graßmann (1809–1877): Visionary Mathematician, Scientist and Neohumanist Scholar*, 197–210. Dordrecht: Kluwer.

Reiff, Richard. and Arnold Sommerfeld. 1904. "Standpunkt der Fernwirkung: Die Elementargesetze." In (Sommerfeld 1903–1926), 2: 3–62.

Ritz, Walter. 1908. "Recherches critiques sur l'électrodynamique générale." *Annales de chimie et de physique* 13: 145–275.

Roseveare, N. T. 1982. *Mercury's Perihelion: From Le Verrier to Einstein*. Oxford: Oxford University Press.

Rowe, David E. 1989. "Klein, Hilbert, and the Göttingen Mathematical Tradition." In Katherina M. Olesko (ed.), *Science in Germany*, 186–213. *Osiris* 5. Philadelphia: History of Science Society.

———. 1992. *Felix Klein, David Hilbert, and the Göttingen Mathematical Tradition*. Ph.D. dissertation, City University of New York.

———. 1997. "In Search of Steiner's ghosts: imaginary elements in 19th-century geometry." In Dominique Flament (ed.), *Le nombre, un hydre à n visages: entre nombres complexes et vecteurs*, 193–208. Paris: Éditions de la Maison des Sciences de l'Homme.

Rüdenberg, Lily. 1973. "Einleitung: Erinnerungen an H. Minkowski." In (Minkowski 1973), 9–16.

Schwartz, Hermann M. 1972. "Poincaré's Rendiconti paper on relativity, III." *American Journal of Physics* 40: 1282–1287.

Schwermer, Joachim. 1991. "Räumliche Anschauung und Minima positiv definiter quadratischer Formen." *Jahresbericht der deutschen Mathematiker-Vereinigung* 93: 49–105.

Serre, Jean-Pierre. 1993. "Smith, Minkowski et l'Académie des Sciences." *Gazette des mathématiciens* 56: 3–9.

Silberstein, Ludwik. 1914. *The Theory of Relativity*. London: Macmillan.

de Sitter, Willem. 1911. "On the bearing of the principle of relativity on gravitational astronomy." *Monthly Notices of the Royal Astronomical Society* 71: 388–415.

Sommerfeld, Arnold. (ed). 1903–1926. *Physik*, 3 vols. (*Encyklopädie der mathematischen Wissenschaften mit Einschluss ihrer Anwendungen,* vol. 5.) Leipzig: Teubner.

Sommerfeld, Arnold. 1904. "Bezeichnung und Benennung der elektromagnetischen Grössen in der Enzyklopädie der mathematischen Wissenschaften V." *Physikalische Zeitschrift* 5: 467–470.

———. 1910a. "Zur Relativitätstheorie, I: Vierdimensionale Vektoralgebra." *Annalen der Physik* 32: 749–776.

———. 1910b. "Zur Relativitätstheorie, II: Vierdimensionale Vektoranalysis." *Annalen der Physik* 33: 649–689.

———. 1913. "Anmerkungen zu Minkowski, Raum und Zeit." In Otto Blumenthal (ed.), *Das Relativitätsprinzip; Eine Sammlung von Abhandlungen*, 69–73. (*Fortschritte der mathematischen Wissenschaften in Monographien*, vol. 2.) Leipzig: Teubner.

———. 2001–2004. *Wissenschaftlicher Briefwechsel*. Michael Eckert and Karl Märker (eds.). Diepholz: GNT-Verlag.

Staley, Richard. 1998. "On the histories of relativity: propagation and elaboration of relativity theory in participant histories in Germany, 1905–1911." *Isis* 89: 263–299.

Stein, Howard. 1987. "After the Baltimore Lectures: some philosophical reflections on the subsequent development of physics." In Robert Kargon and Peter Achinstein (eds.), *Kelvin's Baltimore Lectures and Modern Theoretical Physics: Historical and Philosophical Perspectives*, 375–398. Cambridge, MA: MIT Press.

Strobl, Walter. 1985. "Aus den wissenschaftlichen Anfängen Hermann Minkowskis." *Historia Mathematica* 12: 142–156.

Tait, Peter G. 1882–1884. *Traité élémentaire des quaternions*. Paris: Gauthier-Villars.

Tisserand, Francois-Félix. 1889–1896. *Traité de mécanique céleste*. Paris: Gauthier-Villars.

Torretti, Roberto. 1996. *Relativity and Geometry*. New York: Dover Publications, 2nd edition.

Varičak, Vladimir. 1910. "Anwendung der Lobatschefskijschen Geometrie in der Relativtheorie." *Physikalische Zeitschrift* 11: 93–96.

Vizgin, Vladimir P. 1994. *Unified Field Theories in the First Third of the 20th Century*. (*Science Networks Historical Studies*, vol. 3.) Basel: Birkhäuser.

Wacker, Fritz. 1906. "Über Gravitation und Elektromagnetismus." *Physikalische Zeitschrift* 7: 300–302.

Walter, Scott. 1997. "La vérité en géométrie: sur le rejet mathématique de la doctrine conventionnaliste." *Philosophia Scientiæ* 2: 103–135.

———. 1999a. "Minkowski, Mathematicians, and the Mathematical Theory of Relativity." In (Goenner et al. 1999), 45–86.

———. 1999b. "The Non-Euclidean Style of Minkowskian Relativity." In J. Gray (ed.), *The Symbolic Universe: Geometry and Physics, 1890–1930*, 91–127. Oxford: Oxford University Press.

Weinstein, Max B. 1913. *Die Physik der bewegten Materie und die Relativitätstheorie*. Leipzig: Barth.

———. 1914. *Kräfte und Spannungen: Das Gravitations- und Strahlenfeld*. (*Sammlung Vieweg*, vol. 8.) Braunschweig: Vieweg.

Whitrow, Gerald J., and George E. Morduch. 1965. "Relativistic theories of gravitation: a comparative analysis with particular reference to astronomical tests." *Vistas in Astronomy* 6: 1–67.

Whittaker, Edmund T. 1951–1953. *A History of the Theories of Aether and Electricity*. London: T. Nelson.

de Wisniewski, Felix J. 1913a. "Zur Minkowskischen Mechanik I." *Annalen der Physik* 40: 387–390.

———. 1913b. "Zur Minkowskischen Mechanik II." *Annalen der Physik* 40: 668–676.

Yaghjian, Arthur D. 1992. *Relativistic Dynamics of a Charged Sphere: Updating the Lorentz-Abraham Model*. (*Lecture Notes in Physics: Monographs*, vol. 11.) Berlin: Springer.

Zahar, Elie. 1989. *Einstein's Revolution: A Study in Heuristic*. La Salle, Ill.: Open Court.

Zassenhaus, Hans J. 1975. "On the Minkowski-Hilbert dialogue on mathematization." *Canadian Mathematical Bulletin* 18: 443–461.

Zenneck, Jonathan. 1903. "Gravitation." In (Sommerfeld 1903–1926), 1: 25–67. (Printed in this volume.)

HENRI POINCARÉ

ON THE DYNAMICS OF THE ELECTRON (EXCERPTS)

Originally published as "Sur la dynamique de l'électron" in Rendiconti del Circolo Matematico di Palermo 21 (1906), pp. 129–175. Author's date: Paris, July 1905. The Introduction, §1, and §9 are translated here.[1]

INTRODUCTION

It seems at first that the aberration of light and related optical and electrical phenomena will provide us with a means of determining the absolute motion of the Earth, or rather its motion with respect to the aether, as opposed to its motion with respect to other celestial bodies. Fresnel pursued this idea, but soon recognized that the Earth's motion does not alter the laws of refraction and reflection. Analogous experiments, like that of the water-filled telescope, and all those considering terms no higher than first order relative to the aberration, yielded only negative results; the explanation was soon discovered. But Michelson, who conceived an experiment sensitive to terms depending on the square of the aberration, failed in turn.

It appears that this impossibility to detect the absolute motion of the Earth by experiment may be a general law of nature; we are naturally inclined to admit this law, which we will call the *Postulate of Relativity* and admit without restriction. Whether or not this postulate, which up to now agrees with experiment, may later be corroborated or disproved by experiments of greater precision, it is interesting in any case to ascertain its consequences.

An explanation was proposed by Lorentz and FitzGerald, who introduced the hypothesis of a contraction of all bodies in the direction of the Earth's motion and proportional to the square of the aberration. This contraction, which we will call the *Lorentzian contraction*, would explain Michelson's experiment and all others performed up to now. The hypothesis would become insufficient, however, if we were to admit the postulate of relativity in full generality.

Lorentz then sought to extend his hypothesis and to modify it in order to obtain perfect agreement with this postulate. This is what he succeeded in doing in his article entitled *Electromagnetic phenomena in a system moving with any velocity smaller than that of light* (*Proceedings of the Amsterdam Academy*, 27 May, 1904).

The importance of the question persuaded me to take it up in turn; the results I | obtained agree with those of Mr. Lorentz on all the significant points. I was led [130]

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 3
Gravitation in the Twilight of Classical Physics: Between Mechanics, Field Theory, and Astronomy.

merely to modify and extend them only in a few details; further on we will see the points of divergence, which are of secondary importance.

Lorentz's idea may be summed up like this: if we are able to impress a translation upon an entire system without modifying any observable phenomena, it is because the equations of an electromagnetic medium are unaltered by certain transformations, which we will call *Lorentz transformations*. Two systems, one of which is at rest, the other in translation, become thereby exact images of each other.

Langevin*) sought to modify Lorentz's idea; for both authors, the moving electron takes the form of a flattened ellipsoid. For Lorentz, two axes of the ellipsoid remain constant, while for Langevin, ellipsoid volume remains constant. The two scientists also showed that these two hypotheses are corroborated by Kaufmann's experiments to the same extent as the original hypothesis of Abraham (rigid-sphere electron).

The advantage of Langevin's theory is that it requires only electromagnetic forces, and bonds; it is, however, incompatible with the postulate of relativity. This is what Lorentz showed, and this is what I found in turn using a different method, which calls on principles of group theory.

We must return therefore to Lorentz's theory, but if we want to do this and avoid intolerable contradictions, we must posit the existence of a special force that explains both the contraction, and the constancy of two of the axes. I sought to determine this force, and found that *it may be assimilated to a constant external pressure on the deformable and compressible electron, whose work is proportional to the electron's change in volume*.

If the inertia of matter is exclusively of electromagnetic origin, as generally admitted in the wake of Kaufmann's experiment, and all forces are of electromagnetic origin (apart from this constant pressure that I just mentioned), the postulate of relativity may be established with perfect rigor. This is what I show by a very simple calculation based on the principle of least action.

But that is not all. In the article cited above, Lorentz judged it necessary to extend his hypothesis in such a way that the postulate remains valid in case there are forces of non-electromagnetic origin. According to Lorentz, all forces are affected by the Lorentz transformation (and consequently by a translation) in the same way as electromagnetic forces.

It was important to examine this hypothesis closely, and in particular to ascertain the modifications we would have to apply to the laws of gravitation.

[131] We find first of all that it requires us to assume that gravitational propagation | is not instantaneous, but occurs with the speed of light. One might think that this is reason enough to reject the hypothesis, since Laplace demonstrated that this cannot be the case. In reality, however, the effect of this propagation is compensated in large

* Langevin was anticipated by Mr. Bucherer of Bonn, who earlier advanced the same idea. (See: Bucherer, *Mathematische Einführung in die Elektronentheorie*, August, 1904. Teubner, Leipzig). [Poincaré's footnote.]

part by a different cause, in such a way that no contradiction arises between the proposed law and astronomical observations.

Is it possible to find a law satisfying Lorentz's condition, and reducing to Newton's law whenever the speeds of celestial bodies are small enough to allow us to neglect their squares (as well as the product of acceleration and distance) with respect to the square of the speed of light?

To this question we must respond in the affirmative, as we will see later.

Modified in this way, is the law compatible with astronomical observations?

It seems so on first sight, but the question will be settled only after an extended discussion.

Suppose, then, that this discussion is settled in favor of the new hypothesis, what should we conclude? If propagation of attraction occurs with the speed of light, it could not be a fortuitous accident. Rather, it must be because it is a function of the aether, and then we would have to try to penetrate the nature of this function, and to relate it to other fluid functions.

We cannot be content with a simple juxtaposition of formulas that agree with each other by good fortune alone; these formulas must, in a manner of speaking, interpenetrate. The mind will be satisfied only when it believes it has perceived the reason for this agreement, and the belief is strong enough to entertain the illusion that it could have been predicted.

But the question may be viewed from a different perspective, better shown via an analogy. Let us imagine a pre-Copernican astronomer who reflects on Ptolemy's system; he will notice that for all the planets, one of two circles—epicycle or deferent—is traversed in the same time. This fact cannot be due to chance, and consequently between all the planets there is a mysterious link we can only guess at.

Copernicus, however, destroys this apparent link by a simple change in the coordinate axes that were considered fixed. Each planet now describes a single circle, and orbital periods become independent (until Kepler reestablishes the link that was believed to have been destroyed).

It is possible that something analogous is taking place here. If we were to admit the postulate of relativity, we would find the same number in the law of gravitation and the laws of electromagnetism—the speed of light—and we would find it again in all other forces of any origin whatsoever. This state of affairs may be explained in one of two ways: either everything in the universe would be of electromagnetic origin, or this aspect—shared, as it were, by all physical phenomena—would be a mere epiphenomenon, something due to our methods of | measurement. How do we go about [132]
measuring? The first response will be: we transport objects considered to be invariable solids, one on top of the other. But that is no longer true in the current theory if we admit the Lorentzian contraction. In this theory, two lengths are equal, by definition, if they are traversed by light in equal times.

Perhaps if we were to abandon this definition Lorentz's theory would be as fully overthrown as was Ptolemy's system by Copernicus's intervention. Should that happen some day, it would not prove that Lorentz's efforts were in vain, because regardless of what one may think, Ptolemy was useful to Copernicus.

I, too, have not hesitated to publish these few partial results, even if at this very moment the discovery of magneto-cathode rays seems to threaten the entire theory.

1. LORENTZ TRANSFORMATION

Lorentz adopted a certain system of units in order to do away with 4π factors in formulas. I will do the same, and in addition, select units of length and time in such a way that the speed of light equals 1. Under these conditions, and denoting electric displacement f, g, h, magnetic intensity α, β, γ, vector potential F, G, H, scalar potential ψ, charge density ρ, electron velocity ξ, η, ζ, and current u, v, w, the fundamental formulas become:

$$\left.\begin{array}{llll} u = \dfrac{df}{dt} + \rho\xi = \dfrac{d\gamma}{dy} - \dfrac{d\beta}{dz}, & \alpha = \dfrac{dH}{dy} - \dfrac{dG}{dz}, & f = -\dfrac{dF}{dt} - \dfrac{d\psi}{dx}, & \\ \dfrac{d\alpha}{dt} = \dfrac{dg}{dz} - \dfrac{dh}{dy}, \quad \dfrac{d\rho}{dt} + \sum \dfrac{d\rho\xi}{dx} = 0, & \sum \dfrac{df}{dx} = \rho, & \dfrac{d\psi}{dt} + \sum \dfrac{dF}{dx} = 0, & \\ \square = \Delta - \dfrac{d^2}{dt^2} = \sum \dfrac{d^2}{dx^2} - \dfrac{d^2}{dt^2}, & \square\psi = -\rho, & \square F = -\rho\xi. & \end{array}\right\} \tag{1}$$

An elementary particle of matter of volume $dxdydz$ is acted upon by a mechanical force, the components of which are derived from the formula:

$$X = \rho f + \rho(\eta\gamma - \zeta\beta). \tag{2}$$

These equations admit a remarkable transformation discovered by Lorentz, which owes its interest to the fact that it explains why no experiment can inform us of the absolute motion of the universe. Let us put:

$$x' = kl(x + \varepsilon t), \quad t' = kl(t + \varepsilon x), \quad y' = ly, \quad z' = lz, \tag{3}$$

where l and ε are two arbitrary constants, such that

$$k = \frac{1}{\sqrt{1 - \varepsilon^2}}. \; |$$

[133] Now if we put:

$$\square' = \sum \frac{d^2}{dx'^2} - \frac{d^2}{dt'^2},$$

we will have:

$$\square' = \square l^{-2}.$$

Let a sphere be carried along with the electron in uniform translation, and let the equation of this mobile sphere be:

$$(x-\xi t)^2+(y-\eta t)^2+(z-\zeta t)^2 = r^2,$$

and the volume of the sphere be $\frac{4}{3}\pi r^3$. [2]

The transformation will change the sphere into an ellipsoid, the equation of which is easy to find. We thus deduce easily from (3):

$$x=\frac{k}{l}(x'-\varepsilon t'),\quad t=\frac{k}{l}(t'-\varepsilon x'),\quad y=\frac{y'}{l},\quad z=\frac{z'}{l}. \tag{3'}$$

The equation of the ellipsoid then becomes:

$$k^2(x'-\varepsilon t'-\xi t'+\varepsilon\xi x')^2+(y'-\eta kt'+\eta k\varepsilon x')^2+(z'-\zeta kt'+\zeta k\varepsilon x')^2 = l^2r^2.$$

This ellipsoid is in uniform motion; for $t' = 0$, it reduces to

$$k^2x'^2(1+\xi\varepsilon)^2+(y'+\eta k\varepsilon x')^2+(z'+\zeta k\varepsilon x')^2 = l^2r^2.$$

and has a volume:

$$\frac{4}{3}\pi r^3\frac{l^3}{k(1+\xi\varepsilon)}.$$

If we want electron charge to be unaltered by the transformation, and if we designate the new charge density ρ' we will find:

$$\rho' = \frac{k}{l^3}(\rho+\varepsilon\rho\xi). \tag{4}$$

What will be the new velocity components ξ', η' and ζ'? We should have:

$$\xi' = \frac{dx'}{dt'} = \frac{d(x+\varepsilon t)}{d(t+\varepsilon x)} = \frac{\xi+\varepsilon}{1+\varepsilon\xi},$$
$$\eta' = \frac{dy'}{dt'} = \frac{dy}{kd(t+\varepsilon x)} = \frac{\eta}{k(1+\varepsilon\xi)},\qquad \zeta' = \frac{\zeta}{k(1+\varepsilon\xi)},$$

whence:

$$\rho'\xi' = \frac{k}{l^3}(\rho\xi+\varepsilon\rho),\qquad \rho'\eta' = \frac{1}{l^3}\rho\eta,\qquad \rho'\zeta' = \frac{1}{l^3}\rho\zeta. \tag{4'}$$

Here is where I must point out for the first time a difference with Lorentz. In my notation, Lorentz put (loc. cit., page 813, formulas 7 and 8):

$$\rho'=\frac{1}{kl^3}\rho,\quad \xi'=k^2(\xi+\varepsilon),\quad \eta'=k\eta,\quad \zeta'=k\zeta. \;|$$

In this way we recover the formulas: [134]

$$\rho'\xi'=\frac{k}{l^3}(\rho\xi+\varepsilon\rho),\quad \rho'\eta'=\frac{1}{l^3}\rho\eta,\quad \rho'\zeta'=\frac{1}{l^3}\rho\zeta;$$

although the value of ρ' differs.

It is important to notice that the formulas (4) and (4') satisfy the condition of continuity

$$\frac{d\rho'}{dt'} + \sum \frac{d\rho'\xi'}{dx'} = 0.$$

To see this, let λ be an undetermined coefficient and D the Jacobian of

$$t + \lambda\rho, \quad x + \lambda\rho\xi, \quad y + \lambda\rho\eta, \quad z + \lambda\rho\zeta \tag{5}$$

with respect to t, x, y, z. It follows that:

$$D = D_0 + D_1\lambda + D_2\lambda^2 + D_3\lambda^3 + D_4\lambda^4,$$

with $D_0 = 1, D_1 = \dfrac{d\rho}{dt} + \sum \dfrac{d\rho\xi}{dx} = 0$.

Let $\lambda' = l^4\lambda$,[3] then the 4 functions

$$t' + \lambda'\rho', \quad x' + \lambda'\rho'\xi', \quad y' + \lambda'\rho'\eta', \quad z' + \lambda'\rho'\zeta' \tag{5'}$$

are related to the functions (5) by the same linear relationships as the old variables to the new ones. Therefore, if we denote D' the Jacobian of the functions (5') with respect to the new variables, it follows that:

$$D' = D, \quad D' = D'_0 + D'_1\lambda' + \ldots + D'_4\lambda'^4,$$

and thereby:[4]

$$D'_0 = D_0 = 1, \quad D'_1 = l^{-4}D_1 = 0 = \frac{d\rho'}{dt'} + \sum \frac{d\rho'\xi'}{dx'}. \quad \text{Q. E. D.}$$

Under Lorentz's hypothesis, this condition would not be met since ρ' has a different value.

We will define the new vector and scalar potentials in such a way as to satisfy the conditions

$$\Box'\psi' = -\rho', \qquad \Box'F' = -\rho'\xi'. \tag{6}$$

From this we deduce:

$$\psi' = \frac{k}{l}(\psi + \varepsilon F), \qquad F' = \frac{k}{l}(F + \varepsilon\psi), \qquad G' = \frac{1}{l}G, \qquad H' = \frac{1}{l}H. \tag{7}$$

These formulas differ noticeably from those of Lorentz, although the divergence stems ultimately from the definitions employed.

New electric and magnetic fields are now chosen in order to satisfy the equations:

$$f' = -\frac{dF'}{dt'} - \frac{d\psi'}{dx'}, \qquad \alpha' = \frac{dH'}{dy'} - \frac{dG'}{dz'}. \tag{8}$$

[135] | It is easy to see that:

$$\frac{d}{dt'} = \frac{k}{l}\left(\frac{d}{dt} - \varepsilon\frac{d}{dx}\right), \qquad \frac{d}{dx'} = \frac{k}{l}\left(\frac{d}{dx} - \varepsilon\frac{d}{dt}\right), \qquad \frac{d}{dy'} = \frac{1}{l}\frac{d}{dy}, \qquad \frac{d}{dz'} = \frac{1}{l}\frac{d}{dz},$$

and we deduce thereby:

$$\left.\begin{aligned} f' &= \frac{1}{l^2}f, \qquad & g' &= \frac{k}{l^2}(g + \varepsilon\gamma), \qquad & h' &= \frac{k}{l^2}(h - \varepsilon\beta), \\ \alpha' &= \frac{1}{l^2}\alpha, \qquad & \beta' &= \frac{k}{l^2}(\beta - \varepsilon h), \qquad & \gamma' &= \frac{k}{l^2}(\gamma + \varepsilon g). \end{aligned}\right\} \tag{9}$$

These formulas are identical to those of Lorentz.

Our transformation does not alter (1). In fact, the condition of continuity, as well as (6) and (8) were already featured in (1) (neglecting the primes).

Combining (6) with the condition of continuity, we obtain:

$$\frac{d\psi'}{dt'} + \sum \frac{dF'}{dx'} = 0. \tag{10}$$

It remains for us to establish:

$$\frac{df'}{dt'} + \rho'\xi' = \frac{d\gamma'}{dy'} - \frac{d\beta'}{dz'}, \qquad \frac{d\alpha'}{dt'} = \frac{dg'}{dz'} - \frac{dh'}{dy'}, \qquad \sum \frac{df'}{dx'} = \rho',$$

and it is easy to see that these are necessary consequences of (6), (8) and (10).

We must now compare forces before and after the transformation.

Let X, Y, Z be the force prior to the transformation, and X', Y', Z' the force after the transformation, both forces being per unit volume. In order for X' to satisfy the same equations as before the transformation, we must have:

$$\begin{aligned} X' &= \rho' f' + \rho'(\eta'\gamma' - \zeta'\beta'), \\ Y' &= \rho' g' + \rho'(\zeta'\alpha' - \xi'\gamma'), \\ Z' &= \rho' h' + \rho'(\xi'\beta' - \eta'\alpha'), \end{aligned}$$

or, replacing all quantities by their values (4), (4') and (9), and in light of (2):

$$\left.\begin{aligned} X' &= \frac{k}{l^5}\left(X + \varepsilon \sum X\xi\right), \\ Y' &= \frac{1}{l^5}Y, \\ Z' &= \frac{1}{l^5}Z. \end{aligned}\right\} \tag{11}$$

Instead of representing the components of force per unit volume by X_1, Y_1, Z_1, we now let these terms represent the force per unit electron charge, and we let X'_1, Y'_1, Z'_1 represent the latter force after transformation. It follows that:

$$X_1 = f + \eta\gamma - \zeta\beta, \quad X'_1 = f' + \eta'\gamma' - \zeta'\beta', \quad X = \rho X_1, \quad X' = \rho' X'_1$$

[136] | and we obtain the equations:

$$\left.\begin{aligned} X'_1 &= \frac{k}{l^5}\frac{\rho}{\rho'}\left(X_1 + \varepsilon\sum X_1\xi\right), \\ Y'_1 &= \frac{1}{l^5}\frac{\rho}{\rho'}Y_1, \\ Z'_1 &= \frac{1}{l^5}\frac{\rho}{\rho'}Z_1. \end{aligned}\right\} \qquad (11')$$

Lorentz found (page 813, equation (10) with different notation):

$$\left.\begin{aligned} X_1 &= l^2 X'_1 - l^2\varepsilon(\eta' g' + \zeta' h'), \\ Y_1 &= \frac{l^2}{k}Y'_1 + \frac{l^2\varepsilon}{k}\xi' g', \\ Z_1 &= \frac{l^2}{k}Z'_1 + \frac{l^2\varepsilon}{k}\xi' h'. \end{aligned}\right\} \qquad (11'')$$

Before going any further, it is important to locate the source of this significant divergence. It obviously springs from the fact that the formulas for ξ', η' and ζ' are not the same, while the formulas for the electric and magnetic fields are the same.

If electron inertia is exclusively of electromagnetic origin, and if electrons are subject only to forces of electromagnetic origin, then the conditions of equilibrium require that:

$$X = Y = Z = 0$$

inside the electrons.

According to (11), these relationships are equivalent to

$$X' = Y' = Z' = 0.$$

The electron's equilibrium conditions are therefore unaltered by the transformation.

Unfortunately, such a simple hypothesis is inadmissible. In fact, if we assume $\xi = \eta = \zeta = 0$, the condition $X = Y = Z = 0$ leads necessarily to $f = g = h = 0$, and consequently, to $\sum \frac{df}{dx} = 0$, i.e., $\rho = 0$. Similar results obtain for the most general case. We must then admit that in addition to electromagnetic forces there are either non-electromagnetic forces or bonds. Therefore, we need to identify the conditions that these forces or these bonds must satisfy for electron equilibrium to be undisturbed by the transformation. This will be the object of an upcoming section.

[…]

9. HYPOTHESES CONCERNING GRAVITATION. [166]

In this way Lorentz's theory would fully explain the impossibility of detecting absolute motion, if all forces were of electromagnetic origin.

But there exist other forces to which an electromagnetic origin cannot be attributed, such as gravitation, for example. It may in fact happen, that two systems of bodies produce equivalent electromagnetic fields, i.e., exert the same action on electrified bodies and on currents, and at the same time, these two systems do not exert the same gravitational action on Newtonian masses. The gravitational field is therefore distinct from the electromagnetic field. Lorentz was obliged thereby to extend his hypothesis with the assumption that *forces of any origin whatsoever, and gravitation in particular, are affected by a translation* (or, if one prefers, by the Lorentz transformation) *in the same manner as electromagnetic forces*.

It is now appropriate to enter into the details of this hypothesis, and to examine it more closely. If we want the Newtonian force to be affected by the Lorentz transformation in this fashion, we can no longer suppose that it depends only on the relative position of the attracting and attracted bodies at the instant considered. The force should also depend on the velocities of the two bodies. And that is not all: it will be natural to suppose that the force acting on the attracted body at the instant t depends on the position and velocity of this body at this same instant t, but it will also depend on the position and velocity of the *attracting* body, not at the instant t, but at *an earlier instant*, as if gravitation had taken a certain time to propagate.

Let us now consider the position of the attracted body at the instant t_0 and let x_0, y_0, z_0 be its coordinates, and ξ, η, ζ its velocity components at this instant; let us consider also the attracting body at the corresponding instant $t_0 + t$ and let its coordinates be $x_0 + x, y_0 + y, z_0 + z$, and its velocity components be ξ_1, η_1, ζ_1 at this instant.

First we should have a relationship

$$\varphi(t, x, y, z, \xi, \eta, \zeta, \xi_1, \eta_1, \zeta_1) = 0 \tag{1}$$

in order to define the time t. This relationship will define the law of propagation of gravitational action (I do not constrain myself by any means to a propagation velocity equal in all directions).

Now let X_1, Y_1, Z_1 be the three components of the action exerted on the attracted body at the instant t_0;[5] we want to express X_1, Y_1, Z_1 as functions of

$$t, x, y, z, \xi, \eta, \zeta, \xi_1, \eta_1, \zeta_1. \tag{2}$$

What conditions must be satisfied? |

1° The condition (1) should not be altered by transformations of the Lorentz group. [167]

2° The components X_1, Y_1, Z_1 should be affected by transformations of the Lorentz group in the same manner as the electromagnetic forces designated by the same letters, i.e., in accordance with (11') of section 1.

3° When the two bodies are at rest, the ordinary law of attraction will be recovered.

It is important to note that in the latter case, the relationship (1) vanishes, because if the two bodies are at rest the time t plays no role.

Posed in this fashion the problem is obviously indeterminate. We will therefore seek to satisfy to the utmost other, complementary conditions.

4° Since astronomical observations do not seem to show a sensible deviation from Newton's law, we will choose the solution that differs the least with this law for small velocities of the two bodies.

5° We will make an effort to arrange matters in such a way that t is always negative. Although we can imagine that the effect of gravitation requires a certain time in order to propagate, it would be difficult to understand how this effect could depend on the position *not yet attained* by the attracting body.

There is one case where the indeterminacy of the problem vanishes; it is the one where the two bodies are in mutual *relative* rest, i.e., where:

$$\xi = \xi_1, \qquad \eta = \eta_1, \qquad \zeta = \zeta_1;$$

this is then the case we will examine first, by supposing that these velocities are constant, such that the two bodies are engaged in a common uniform rectilinear translation.

We may suppose that the $x-$axis is parallel to this translation, such that $\eta = \zeta = 0$, and we will let $\varepsilon = -\xi$.

If we apply the Lorentz transformation under these conditions, after the transformation the two bodies will be at rest, and it follows that:

$$\xi' = \eta' = \zeta' = 0.$$

The components X'_1, Y'_1, Z'_1 should then agree with Newton's law and we will have, apart from a constant factor:

$$\left.\begin{aligned} &X'_1 = -\frac{x'}{r'^3}, \qquad Y'_1 = -\frac{y'}{r'^3}, \qquad Z'_1 = -\frac{z'}{r'^3}, \\ &r'^2 = x'^2 + y'^2 + z'^2. \end{aligned}\right\} \tag{3}$$

But according to section 1 we have:

$$x' = k(x + \varepsilon t), \qquad y' = y, \qquad z' = z, \qquad t' = k(t + \varepsilon x),$$

$$\frac{\rho'}{\rho} = k(1 + \xi\varepsilon) = k(1 - \varepsilon^2) = \frac{1}{k}, \qquad \sum X_1 \xi = -X_1 \varepsilon,$$

[168] $$X'_1 = k\frac{\rho}{\rho'}(X_1 + \varepsilon \sum X_1 \xi) = k^2 X_1 (1 - \varepsilon^2) = X_1,$$

$$Y'_1 = \frac{\rho}{\rho'} Y_1 = k Y_1, \qquad Z'_1 = k Z_1.$$

We have in addition:

$$x + \varepsilon t = x - \xi t, \qquad r'^2 = k^2(x - \xi t)^2 + y^2 + z^2$$

and

$$X_1 = \frac{-k(x - \xi t)}{r'^3}, \qquad Y_1 = \frac{-y}{kr'^3}, \qquad Z_1 = \frac{-z}{r'^3}; \tag{4}$$

which may be written:

$$X_1 = \frac{dV}{dx}, \qquad Y_1 = \frac{dV}{dy}, \qquad Z_1 = \frac{dV}{dz}; \qquad V = \frac{1}{kr'}. \tag{4'}$$

It seems at first that the indeterminacy remains, since we made no hypotheses concerning the value of t, i.e., the transmission speed; and that besides, x is a function of t. It is easy to see, however, that the terms appearing in our formulas, $x - \xi t, y, z,$ do not depend on t.

We see that if the two bodies translate together, the force acting on the attracted body is perpendicular to an ellipsoid, at the center of which lies the attracting body.

To advance further, we need to look for the *invariants of the Lorentz group*.

We know that the substitutions of this group (assuming $l = 1$) are linear substitutions that leave unaltered the quadratic form

$$x^2 + y^2 + z^2 - t^2.$$

Let us also put:

$$\xi = \frac{\delta x}{\delta t}, \qquad \eta = \frac{\delta y}{\delta t}, \qquad \zeta = \frac{\delta z}{\delta t};$$

$$\xi_1 = \frac{\delta_1 x}{\delta_1 t}, \qquad \eta_1 = \frac{\delta_1 y}{\delta_1 t}, \qquad \zeta_1 = \frac{\delta_1 z}{\delta_1 t};$$

we see that the Lorentz transformation will make $\delta x, \delta y, \delta z, \delta t,$ and $\delta_1 x, \delta_1 y, \delta_1 z, \delta_1 t$ undergo the same linear substitutions as x, y, z, t.

Let us regard

$$\begin{array}{llll} x, & y, & z, & t\sqrt{-1}, \\ \delta x, & \delta y, & \delta z, & \delta t\sqrt{-1}, \\ \delta_1 x, & \delta_1 y, & \delta_1 z, & \delta_1 t\sqrt{-1}, \end{array}$$

as the coordinates of 3 points P, P', P'' in space of 4 dimensions. We see that the Lorentz transformation is merely a rotation in this space about the origin, assumed fixed. Consequently, we will have no distinct invariants apart from the 6 distances between the 3 points P, P', P'', considered separately and with the origin, or, if one prefers, | apart from the two expressions [169]

$$x^2 + y^2 + z^2 - t^2, \qquad x\delta x + y\delta y + z\delta z - t\delta t,$$

or the 4 expressions of like form deduced from an arbitrary permutation of the 3 points P, P', P''.

But what we seek are invariants that are functions of the 10 variables (2). Therefore, among the combinations of our 6 invariants we must find those depending only on these 10 variables, i.e., those that are 0th degree homogeneous with respect both to $\delta x, \delta y, \delta z, \delta t$, and to $\delta_1 x, \delta_1 y, \delta_1 z, \delta_1 t$. We will then be left with 4 distinct invariants:

$$\sum x^2 - t^2, \qquad \frac{t - \sum x\xi}{\sqrt{1 - \sum \xi^2}}, \qquad \frac{t - \sum x\xi_1}{\sqrt{1 - \sum \xi_1^2}}, \qquad \frac{1 - \sum \xi\xi_1}{\sqrt{(1 - \sum \xi^2)(1 - \sum \xi_1^2)}}. \tag{5}$$

Next let us see how the force components are transformed; we recall the equations (11) of section 1, that refer not to the force X_1, Y_1, Z_1, considered at present, but to the force per unit volume: X, Y, Z.

We designate moreover

$$T = \sum X\xi;$$

we will see that (11) can be written $(l = 1)$:

$$\left.\begin{array}{ll} X' = k(X + \varepsilon T), & T' = k(T + \varepsilon X), \\ Y' = Y, & Z' = Z; \end{array}\right\} \tag{6}$$

in such a way that X, Y, Z, T undergo the same transformation as x, y, z, t. Consequently, the group invariants will be

$$\sum X^2 - T^2, \qquad \sum Xx - Tt, \qquad \sum X\delta x - T\delta t, \qquad \sum X\delta_1 x - T\delta_1 t.$$

However, it is not X, Y, Z, that we need, but X_1, Y_1, Z_1, with

$$T_1 = \sum X_1 \xi.$$

We see that

$$\frac{X_1}{X} = \frac{Y_1}{Y} = \frac{Z_1}{Z} = \frac{T_1}{T} = \frac{1}{\rho}.$$

Therefore, the Lorentz transformation will act in the same manner on X_1, Y_1, Z_1, T_1 as on X, Y, Z, T, except that these expressions will be multiplied moreover by

$$\frac{\rho}{\rho'} = \frac{1}{k(1 + \xi\varepsilon)} = \frac{\delta t}{\delta t'}.$$

Likewise, the Lorentz transformation will act in the same way on $\xi, \eta, \zeta, 1$ as on $\delta x, \delta y, \delta z, \delta t$, except that these expressions will be multiplied moreover by the *same* factor:

$$\frac{\delta t}{\delta t'} = \frac{1}{k(1+\xi\varepsilon)}.$$

Next we consider $X, Y, Z, T\sqrt{-1}$ as the coordinates of a fourth point | Q; the [170]
invariants will then be functions of the mutual distances of the five points

$$O, \quad P, \quad P', \quad P'', \quad Q$$

and among these functions we must retain only those that are 0th degree homogeneous with respect, on one hand, to

$$X, \quad Y, \quad Z, \quad T, \quad \delta x, \quad \delta y, \quad \delta z, \quad \delta t$$

(variables that can be replaced further by $X_1, Y_1, Z_1, T_1, \xi, \eta, \zeta, 1$), and on the other hand, with respect to[6]

$$\delta_1 x, \quad \delta_1 y, \quad \delta_1 z, \quad \delta_1 t$$

(variables that can be replaced further by $\xi_1, \eta_1, \zeta_1, 1$).

In this way we find, beyond the four invariants (5), four distinct new invariants:

$$\frac{\sum X_1^2 - T_1^2}{1-\sum\xi^2}, \quad \frac{\sum X_1 x - T_1 t}{\sqrt{1-\sum\xi^2}}, \quad \frac{\sum X_1\xi_1 - T_1}{\sqrt{1-\sum\xi^2}\sqrt{1-\sum\xi_1^2}}, \quad \frac{\sum X_1\xi - T_1}{1-\sum\xi^2}. \tag{7}$$

The latter invariant is always null according to the definition of T_1.

These terms being settled, what conditions must be satisfied?

1° The first term of (1), defining the velocity of propagation, has to be a function of the 4 invariants (5).

A wealth of hypotheses can obviously be entertained, of which we will examine only two:

A) We can have

$$\sum x^2 - t^2 = r^2 - t^2 = 0,$$

from whence $t = \pm r$, and, since t has to be negative, $t = -r$. This means that the velocity of propagation is equal to that of light. It seems at first that this hypothesis ought to be rejected outright. Laplace showed in effect that the propagation is either instantaneous or much faster than that of light. However, Laplace examined the hypothesis of finite propagation velocity *ceteris non mutatis*; here, on the contrary, this hypothesis is conjoined with many others, and it may be that between them a more or less perfect compensation takes place. The application of the Lorentz transformation has already provided us with numerous examples of this.

B) We can have

$$\frac{t-\sum x\xi_1}{\sqrt{1-\sum\xi_1^2}} = 0, \qquad t = \sum x\xi_1.$$

The propagation velocity is therefore much faster than that of light, but in certain cases t could be positive, which, as we mentioned, seems hardly admissible.[7] *We will therefore stick with hypothesis* (A).

2° The four invariants (7) ought to be functions of the invariants (5).

[171] 3° When the two bodies are at absolute rest, X_1, Y_1, Z_1 ought to have the | values given by Newton's law, and when they are at relative rest, the values given by (4).

For the case of absolute rest, the first two invariants (7) ought to reduce to

$$\sum X_1^2, \qquad \sum X_1 x,$$

or, by Newton's law, to

$$\frac{1}{r^4}, \qquad -\frac{1}{r};$$

in addition, according to hypothesis (A), the 2nd and 3rd invariants in (5) become:

$$\frac{-r-\sum x\xi}{\sqrt{1-\sum \xi^2}}, \qquad \frac{-r-\sum x\xi_1}{\sqrt{1-\sum \xi_1^2}},$$

that is, for absolute rest,

$$-r, \qquad -r.$$

We may therefore admit, *for example*, that the first two invariants in (7) reduce to[8]

$$\frac{(1-\sum \xi_1^2)^2}{(r+\sum x\xi_1)^4}, \qquad -\frac{\sqrt{1-\sum \xi_1^2}}{r+\sum x\xi_1};$$

although other combinations are possible.

A choice must be made among these combinations, and furthermore, we need a 3rd equation in order to define X_1, Y_1, Z_1. In making such a choice, we should try to come as close as possible to Newton's law. Let us see what happens when we neglect the squares of the velocities ξ, η, etc. (still letting $t = -r$). The 4 invariants (5) then become:

$$0, \qquad -r-\sum x\xi, \qquad -r-\sum x\xi_1, \qquad 1$$

and the 4 invariants (7) become:

$$\sum X_1^2, \qquad \sum X_1(x+\xi r), \qquad \sum X_1(\xi_1-\xi), \qquad 0.$$

Before we can make a comparison with Newton's law, another transformation is required. In the case under consideration, $x_0+x, y_0+y, z_0+z,$ represent the coordinates of the attracting body at the instant t_0+t, and $r = \sqrt{\Sigma x^2}$. With Newton's law we have to consider the coordinates of the attracting body $x_0+x_1, y_0+y_1, z_0+z_1$ at the instant t_0, and the distance $r_1 = \sqrt{\sum x_1^2}$.

We may neglect the square of the time t required for propagation, and proceed, consequently, as if the motion were uniform; we then have:

$$x = x_1 + \xi_1 t, \quad y = y_1 + \eta_1 t, \quad z = z_1 + \zeta_1 t, \quad r(r - r_1) = \sum x \xi_1 t;$$

or, since $t = -r$, [172]

$$x = x_1 - \xi_1 r, \quad y = y_1 - \eta_1 r, \quad z = z_1 - \zeta_1 r, \quad r = r_1 - \sum x \xi_1;$$

such that our 4 invariants (5) become:

$$0, \quad -r_1 + \sum x(\xi_1 - \xi), \quad -r_1, \quad 1$$

and our 4 invariants (7) become:

$$\sum X_1^2, \quad \sum X_1[x_1 + (\xi - \xi_1)r_1], \quad \sum X_1(\xi_1 - \xi), \quad 0.$$

In the second of these expressions I wrote r_1 instead of r, because r is multiplied by $\xi - \xi_1$ and because I neglect the square of ξ.

For these 4 invariants (7), Newton's law would yield

$$\frac{1}{r_1^4}, \quad -\frac{1}{r_1} - \frac{\sum x_1(\xi - \xi_1)}{r_1^2}, \quad \frac{\sum x_1(\xi - \xi_1)}{r_1^3}, \quad 0.$$

Therefore, if we designate the 2nd and 3rd of the invariants (5) as A and B, and the first 3 invariants of (7) as M, N, P we will satisfy Newton's law to first-order terms in the square of velocity by setting:

$$M = \frac{1}{B^4}, \qquad N = \frac{+A}{B^2}, \qquad P = \frac{A - B}{B^3}. \tag{8}$$

This solution is not unique. Let C be the 4th invariant in (5); $C - 1$ is of the order of the square of ξ, and it is the same with $(A - B)^2$.

The solution (8) appears at first to be the simplest, nevertheless, it may not be adopted. In fact, since M, N, P are functions of X_1, Y_1, Z_1, and $T_1 = \sum X_1 \xi$, the values of X_1, Y_1, Z_1 can be drawn from these three equations (8), but in certain cases these values would become imaginary.

To avoid this difficulty we will proceed in a different manner. Let us put:

$$k_0 = \frac{1}{\sqrt{1 - \sum \xi^2}}, \qquad k_1 = \frac{1}{\sqrt{1 - \sum \xi_1^2}},$$

which is justified by analogy with the notation

$$k = \frac{1}{\sqrt{1 - \varepsilon^2}},$$

featured in the Lorentz substitution.

In this case, and in light of the condition $-r = t$, the invariants (5) become:

$$0, \quad A = -k_0(r + \sum x\xi), \quad B = -k_1(r + \sum x\xi_1), \quad C = k_0 k_1(1 - \sum \xi\xi_1).$$

[173] Moreover, we notice that the following systems of quantities:

$$\begin{array}{llll} x, & y, & z, & -r = t \\ k_0 X_1, & k_0 Y_1, & k_0 Z_1, & k_0 T_1 \\ k_0\xi, & k_0\eta, & k_0\zeta, & k_0 \\ k_1\xi_1, & k_1\eta_1, & k_1\zeta_1, & k_1 \end{array}$$

undergo the *same* linear substitutions when the transformations of the Lorentz group are applied to them. We are led thereby to put:

$$\left.\begin{aligned} X_1 &= x\frac{\alpha}{k_0} + \xi\beta + \xi_1\frac{k_1}{k_0}\gamma, \\ Y_1 &= y\frac{\alpha}{k_0} + \eta\beta + \eta_1\frac{k_1}{k_0}\gamma, \\ Z_1 &= z\frac{\alpha}{k_0} + \zeta\beta + \zeta_1\frac{k_1}{k_0}\gamma, \\ T_1 &= -r\frac{\alpha}{k_0} + \beta + \frac{k_1}{k_0}\gamma. \end{aligned}\right\} \tag{9}$$

It is clear that if α, β, γ are invariants, X_1, Y_1, Z_1, T_1 will satisfy the fundamental condition, i.e., the Lorentz transformations will make them undergo an appropriate linear substitution.

However, for equations (9) to be compatible we must have

$$\sum X_1\xi - T_1 = 0,$$

which becomes, replacing X_1, Y_1, Z_1, T_1 with their values in (9) and multiplying by k_0^2,

$$-A\alpha - \beta - C\gamma = 0. \tag{10}$$

What we would like is that the values of X_1, Y_1, Z_1 remain in line with Newton's law when we neglect (as above) the squares of velocities ξ, etc. with respect to the square of the velocity of light, and the products of acceleration and distance.

We could select

$$\beta = 0, \qquad \gamma = -\frac{A\alpha}{C}.$$

To the adopted order of approximation, we obtain

$$k_0 = k_1 = 1, \qquad C = 1, \qquad A = -r_1 + \sum x(\xi_1 - \xi), \qquad B = -r_1,$$
$$x = x_1 + \xi_1 t = x_1 - \xi_1 r.$$

The 1st equation in (9) then becomes

$$X_1 = \alpha(x - A\xi_1).$$

But if the square of ξ is neglected, $A\xi_1$ can be replaced by $-r_1\xi_1$ or | by $-r\xi_1$, [174] which yields:

$$X_1 = \alpha(x + \xi_1 r) = \alpha x_1.$$

Newton's law would yield

$$X_1 = -\frac{x_1}{r_1^3}.$$

Consequently, we must select a value for the invariant α, which reduces to $-\frac{1}{r_1^3}$ in the adopted order of approximation, that is, $\frac{1}{B^3}$. Equations (9) will become:

$$\left.\begin{aligned} X_1 &= \frac{x}{k_0 B^3} - \xi_1 \frac{k_1}{k_0}\frac{A}{B^3 C}, \\ Y_1 &= \frac{y}{k_0 B^3} - \eta_1 \frac{k_1}{k_0}\frac{A}{B^3 C}, \\ Z_1 &= \frac{z}{k_0 B^3} - \zeta_1 \frac{k_1}{k_0}\frac{A}{B^3 C}, \\ T_1 &= -\frac{r}{k_0 B^3} - \frac{k_1}{k_0}\frac{A}{B^3 C}. \end{aligned}\right\} \tag{11}$$

We notice first that the corrected attraction is composed of two components: one parallel to the vector joining the positions of the two bodies, the other parallel to the velocity of the attracting body.

Remember that when we speak of the position or velocity of the attracting body, this refers to its position or velocity at the instant the gravitational wave takes off; for the attracted body, on the contrary, this refers to the position or velocity at the instant the gravitational wave arrives, assuming that this wave propagates with the velocity of light.

I believe it would be premature to seek to push the discussion of these formulas further; I will therefore confine myself to a few remarks.

1° The solutions (11) are not unique; we may, in fact, replace the global factor $\frac{1}{B^3}$, by

$$\frac{1}{B^3} + (C-1) f_1(A, B, C) + (A - B)^2 f_2(A, B, C),$$

where f_1 and f_2 are arbitrary functions of A, B, C. Alternatively, we may forgo setting β to zero, but add any complementary terms to α, β, γ that satisfy condition (10) and are of second order with respect to the ξ for α and of first order for β and γ.

2° The first equation in (11) may be written:

$$X_1 = \frac{k_1}{B^3C}[x(1-\sum\xi\xi_1)+\xi_1(r+\sum x\xi)] \tag{11'}$$

and the quantity in brackets itself may be written:

$$(x+r\xi_1)+\eta(\xi_1 y-x\eta_1)+\zeta(\xi_1 z-x\zeta_1), \tag{12}$$

[175] | such that the total force may be separated into three components corresponding to the three parentheses of expression (12); the first component is vaguely analogous to the mechanical force due to the electric field, the two others to the mechanical force due to the magnetic field; to extend the analogy I may, in light of the first remark, replace $1/B^3$ in (11) by $C/(B^3,)$ in such a way that X_1, Y_1, Z_1 are linear functions of the attracted body's velocity ξ, η, ζ, since C has vanished from the denominator of (11').

Next we put:

$$\left.\begin{array}{lll} k_1(x+r\xi_1) = \lambda, & k_1(y+r\eta_1) = \mu, & k_1(z+r\zeta_1) = \nu, \\ k_1(\eta_1 z-\zeta_1 y) = \lambda', & k_1(\zeta_1 x-\xi_1 z) = \mu', & k_1(\xi_1 y-x\eta_1) = \nu'; \end{array}\right\} \tag{13}$$

and since C has vanished from the denominator of (11'), it will follow that:

$$\left.\begin{aligned} X_1 &= \frac{\lambda}{B^3}+\frac{\eta\nu'-\zeta\mu'}{B^3}, \\ Y_1 &= \frac{\mu}{B^3}+\frac{\zeta\lambda'-\xi\nu'}{B^3}, \\ Z_1 &= \frac{\nu}{B^3}+\frac{\xi\mu'-\eta\lambda'}{B^3}; \end{aligned}\right\} \tag{14}$$

and we will have moreover:

$$B^2 = \sum\lambda^2-\sum\lambda'^2. \tag{15}$$

Now λ, μ, ν, or $\frac{\lambda}{B^3}, \frac{\mu}{B^3}, \frac{\nu}{B^3}$, is an electric field of sorts, while λ', μ', ν', or rather $\frac{\lambda'}{B^3}, \frac{\mu'}{B^3}, \frac{\nu'}{B^3}$, is a magnetic field of sorts.

3° The postulate of relativity would compel us to adopt solution (11), or solution (14), or any solution at all among those derived on the basis of the first remark. However, the first question to ask is whether or not these solutions are compatible with astronomical observations. The deviation from Newton's law is of the order of ξ^2, i.e., 10000 times smaller than if it were of the order of ξ, i.e., if the propagation were to take place with the velocity of light, *ceteris non mutatis*; consequently, it is legitimate to hope that it will not be too large. To settle this question, however, would require an extended discussion.

EDITORIAL NOTES

[1] Translated by Scott Walter from *Rendiconti del Circolo Matematico di Palermo* 21, 1906, 129–176. The original notation is faithfully reproduced, including the use of "*d*" for both ordinary and partial differentiation. The translator's endnote calls are bracketed. For alternative translations of Poincaré's memoir, see C. W. Kilmister (*Special Theory of Relativity*, Oxford: Pergamon, 1970, 145–185), and by H. M. Schwartz (*American Journal of Physics* 39:1287–1294; 40:862–872, 1282–1287).

[2] The original reads: "$\frac{4}{3}\pi r^2$".

[3] The original reads: "$\lambda' = l^2\rho'$".

[4] The original reads: "$D'_1 = l^{-2}D_1$".

[5] The original reads: "à l'instant t".

[6] The original reads: "$\delta_1 x, \quad \delta_1 y, \quad \delta_1 z, \quad 1.$"

[7] The original reads: "t pourrait être négatif."

[8] The original has (4) instead of (7).

HERMANN MINKOWSKI

MECHANICS AND THE RELATIVITY POSTULATE

Originally published as "Mechanik und Relativitätspostulat," appended to "Die Grundgleichungen für elektromagnetischen Vorgänge in bewegten Körpern" in Nachrichten der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse, 1908, pp. 53–111.

It would be highly unsatisfactory if the new conception of the notion of time, which is [98]
characterized by the freedom of Lorentz transformations, could be accepted as valid for only a subfield of physics.

Now, many authors state that classical mechanics is opposed to the relativity postulate, which has been taken here to be the foundation of electrodynamics.

In order to assess this we focus on a special Lorentz transformation as represented by eqs. (10), (11), (12),[1] with a non-zero vector v of arbitrary direction, and magnitude q that is <1. But for the moment we will pretend that the ratio of the unit of length and the unit of time has not yet been established, and accordingly we will write in these equations $ct, ct', \frac{q}{c}$ instead of t, t', q, where c is a certain positive constant, and we must have $q < c$. The equations referred to then become

$$\mathfrak{r}_{\bar{\mathrm{v}}}{}' = \mathfrak{r}_{\bar{\mathrm{v}}}, \qquad \mathfrak{r}_{\mathrm{v}}{}' = \frac{c(\mathfrak{r}_{\mathrm{v}} - qt)}{\sqrt{c^2 - q^2}}, \qquad t' = \frac{-q\mathfrak{r}_{\mathrm{v}} + c^2 t}{c\sqrt{c^2 - q^2}};$$

we recall that $\mathfrak{r}$ means the spatial vector x, y, z, and $\mathfrak{r}'$ means the spatial vector x', y', z'.
In these equations we fix v and pass to the limit $c = \infty$, with the result

$$\mathfrak{r}_{\bar{\mathrm{v}}}{}' = \mathfrak{r}_{\bar{\mathrm{v}}}, \qquad \mathfrak{r}_{\mathrm{v}}{}' = \mathfrak{r}_{\mathrm{v}} - qt, \qquad t' = t.$$

| These new equations would mean a transition from the system of spatial coordinates [99]
x, y, z to another spatial coordinate system x', y', z' with parallel axes, whose origin moves with respect to the first in a straight line with constant velocity, while the time parameter would be totally unaffected.

On the basis of this remark one may state:

Classical mechanics postulates covariance of the laws of physics for the group of homogeneous linear transformations of the expression

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 3
Gravitation in the Twilight of Classical Physics: Between Mechanics, Field Theory, and Astronomy.

$$-x^2 - y^2 - z^2 + c^2 t^2 \quad (1)$$

into itself, with the specification $c = \infty$.

Now it would be downright confusing to find in one subfield of physics a covariance of the laws for transformations of expression (1) into itself with a certain finite c, but with $c = \infty$ in another subfield. That Newtonian mechanics could only claim this covariance for $c = \infty$, and could not devise it with c equal to the speed of light, needs no explanation. But is it not legitimate today to try and regard that traditional covariance for $c = \infty$ as only an approximation, gained from preliminary experience, to a more exact covariance of the laws of nature for a certain finite c?

I want to explain in detail that by *reforming mechanics, replacing Newton's relativity postulate with* $c = \infty$ *by one with a finite* c, the axiomatic construction of mechanics even seems to attain considerable perfection.

The ratio of the unit of time to the unit of length shall be normalized so that the relevant relativity postulate has $c = 1$.

When I now want to transfer geometrical pictures to the manifold of the four variables x, y, z, t, it may be convenient for ease of understanding what follows to exclude totally y, z from consideration at first, and to interpret x and t as arbitrary oblique-angled rectilinear coordinates in a plane.

A spacetime origin $O(x, y, z, t = 0, 0, 0, 0)$ remains fixed under the Lorentz transformations. The object

$$-x^2 - y^2 - z^2 + t^2 = 1, \qquad t > 0, \quad (2)$$

[100] | a *hyperbolic shell*, includes the spacetime point $A(x, y, z, t = 0, 0, 0, 1)$ and all spacetime points A' that appear as $(x', y', z', t' = 0, 0, 0, 1)$ in the new components x', y', z', t' after some Lorentz transformation.

The direction of a radius vector OA' from O to a point A' of (2) and the directions of the tangents to (2) at A' shall be called *normal* to each other.

Let us follow a definite point in the matter on its orbit for all times t. I call the totality of the spacetime points x, y, z, t, that correspond to this point at different times t a *spacetime line* [world line].[2]

The problem of determining the motion of matter is to be understood in this way: *For every spacetime point the direction of the spacetime line running through it is to be determined.*

To *transform* a spacetime point $P(x, y, z, t)$ *to rest* means to introduce by a Lorentz-transformation a system of reference x', y', z', t' such that the t'-axis OA' acquires that direction which is exhibited by the spacetime line running through P. The space $t' = \text{const.}$, which is to be laid down through P, shall be called the *normal* space to the spacetime line at P. To the increment dt of time t starting at P there corresponds the increment[1]

1 We resume the earlier conventions (see sections 3 and 4) for the notation with indices and the symbols $\mathfrak{w}, w$. [3]

$$d\tau = \sqrt{dt^2 - dx^2 - dy^2 - dz^2} = dt\sqrt{1 - \mathfrak{w}^2} = \frac{dx_4}{w_4} \tag{3}$$

of the parameter t' to be introduced in this construction. The value of the integral

$$\int d\tau = \int \sqrt{-(dx_1^2 + dx_2^2 + dx_3^2 + dx_4^2)},$$

calculated over the spacetime line from any fixed initial point P^0 to an endpoint P, taken to be variable, shall be called the *proper time* of the corresponding point of the matter at the spacetime point P. (This is a generalization of the concept of *local time* introduced by Lorentz for uniform motion.)

If we take a spatially extended body R^0 at a particular time t^0, then the region of all the spacetime lines passing through the spacetime points R^0, t^0 shall be called a *spacetime thread* [pencil].

If we have an analytic expression $\Theta(x, y, z, t)$, such that $\Theta(x, y, z, t) = 0$ is met at one point by every spacetime line of the thread, where |

$$-\left(\frac{\partial\Theta}{\partial x}\right)^2 - \left(\frac{\partial\Theta}{\partial y}\right)^2 - \left(\frac{\partial\Theta}{\partial z}\right)^2 + \left(\frac{\partial\Theta}{\partial t}\right)^2 > 0, \qquad \frac{\partial\Theta}{\partial t} > 0$$ [101]

then we will call the set Q of the meeting points involved a *cross section* of the tread. At every point $P(x, y, z, t)$ of such a cross section we can introduce a system of reference x', y', z', t' by means of a Lorentz-transformation such that thereafter we have

$$\frac{\partial\Theta}{\partial x'} = 0, \qquad \frac{\partial\Theta}{\partial y'} = 0, \qquad \frac{\partial\Theta}{\partial z'} = 0, \qquad \frac{\partial\Theta}{\partial t'} > 0$$

The direction of the corresponding unique t'-axis shall be called the *upper normal* [future-pointing normal] of the cross section Q at the point P, and the value $dJ = \iiint dx'dy'dz'$ for a neighborhood of P within the cross section shall be called an *element of volume content* [*Inhaltselement*] of the cross section. In this sense R^0, t^0 can be called the thread's cross section $t = t^0$ normal to the t-axis, and the volume of the body R^0 can be called the *volume content* [*Inhalt*] of this cross section.

By allowing the space R^0 to converge to a point we arrive at the concept of an *infinitely thin* spacetime thread. In such a thread we will always think of *one* spacetime line as a somehow distinguished *central line*. Also, we understand by the *proper time of the thread* the proper time measured on this central line; by the *normal section* of the thread its intersections with the spaces that are normal to the central line at its points.

Now we formulate the *principle of conservation of mass*.

To every region R at a time t there belongs a positive quantity, the *mass within R at time t*. If R converges to a point x, y, z, t, then the quotient of this mass and the volume of R shall approach a limit $\mu(x, y, z, t)$, the *mass density* at the spacetime point x, y, z, t.

The principle of mass conservation states: *For an infinitely thin spacetime thread the product* μdJ *of the mass density* μ *at one point* x, y, z, t *of the thread* (i.e., of the central line of the thread) *and the volume content* dJ *of the normal section to the* t*-axis passing through the point is always constant along the entire thread.*

Now the volume content dJ_n of the thread's normal section passing through x, y, z, t is calculated as |

[102]

$$dJ_n = \frac{1}{\sqrt{1-\mathfrak{w}^2}} dJ = -iw_4 dJ = \frac{dt}{d\tau} dJ, \tag{4}$$

and therefore let us define

$$\nu = \frac{\mu}{-iw_4} = \mu\sqrt{1-\mathfrak{w}^2} = \mu\frac{d\tau}{dt} \tag{5}$$

as the *rest-mass density* at the point x, y, z, t. Then the principle of mass conservation can also be formulated as follows:

For an infinitely thin spacetime thread the product of the rest-mass density and the volume content of the normal section at a point of the thread is always constant along the entire thread.

In an arbitrary spacetime thread, fix an initial cross section Q^0 and then a second cross section Q^1, which has those and only those points in common with Q^0 that lie on the thread's boundary, and let the spacetime lines within the thread assume larger values t on Q^1 than on Q^0. The finite region bounded by Q^0 and Q^1 together shall be called a *spacetime sickle* [lens], with Q^0 the *lower* and Q^1 the *upper* boundary of the sickle.

Think of the thread decomposed into many very thin spacetime threads; then to every entrance of a thin thread in the lower boundary of the sickle there corresponds an exit through the upper boundary, where the product νdJ_n formed in the sense of (4) and (5) takes the same value each time. Therefore the difference of the two integrals $\int \nu dJ_n$ vanishes, when the first extends over the upper, the second over the lower boundary of the sickle. According to a well-known theorem of integral calculus this difference equals the integral[4]

$$\iiiint \mathrm{lor}\, \nu\overline{w}\, dx dy dz dt,$$

taken over the whole region of the sickle, where we have (cf. (67) in section 12)[5]

$$\mathrm{lor}\, \nu\overline{w} = \frac{\partial \nu w_1}{\partial x_1} + \frac{\partial \nu w_2}{\partial x_2} + \frac{\partial \nu w_3}{\partial x_3} + \frac{\partial \nu w_4}{\partial x_4}.$$

If the sickle is contracted to one spacetime point x, y, z, t, the consequence of this is the differential equation

$$\mathrm{lor}\,\nu\overline{w} = 0, \tag{6}$$

that is, the *continuity condition* |

$$\frac{\partial\mu\mathfrak{w}_x}{\partial x} + \frac{\partial\mu\mathfrak{w}_y}{\partial y} + \frac{\partial\mu\mathfrak{w}_z}{\partial z} + \frac{\partial\mu}{\partial t} = 0.$$ [103]

Further we form the integral

$$\boldsymbol{N} = \iiiint \nu\, dx dy dz dt. \tag{7}$$

taken over the whole region of a spacetime sickle. We divide the sickle into thin spacetime threads, and further divide each thread into small elements $d\tau$ of its proper time, which are however still large compared to the linear dimension of the normal section; we put the mass of such a thread $\nu dJ_n = dm$ and write τ^0 and τ^1 for the proper time of the thread on the lower resp. upper boundary of the sickle; then the integral (7) can also be interpreted as

$$\iint \nu dJ_n d\tau = \int (\tau^1 - \tau^0) dm$$

over all the threads in the sickle.

Now I consider the spacetime lines within a spacetime sickle as curves made of a substance, consisting of substance-points, and I imagine them subjected to a continuous change in position within the sickle of the following type: The whole curves shall be displaced arbitrarily but with *fixed end points on the lower and upper boundary* of the sickle, and each substance-point on them shall be guided so that it is always displaced *normal to its curve*. The whole process shall be capable of analytic representation by means of a parameter ϑ and the value $\vartheta = 0$ shall correspond to the curves that actually run as spacetime lines inside the sickle. Such a process shall be called a *virtual displacement within the sickle*.

Let the point x, y, z, t in the sickle for $\vartheta = 0$ arrive at $x + \delta x, y + \delta y, z + \delta z, t + \delta t$ for the parameter value ϑ; then the latter quantities are functions of x, y, z, t, ϑ. Let us again consider an infinitely thin spacetime thread at the location x, y, z, t with a normal section of volume content dJ_n, and let $dJ_n + \delta dJ_n$ be the volume content of the normal section at the corresponding location of the varied thread; then we will take the *principle of conservation of mass* into account by assuming at this varied location a rest mass density $\nu + \delta\nu$ according to

$$(\nu + \delta\nu)(dJ_n + \delta dJ_n) = \nu dJ_n = dm, \tag{8}$$

| where ν is understood to be the actual rest mass density at x, y, z, t. According to [104]
this convention the integral (7), extended over the region of the sickle, then varies upon the virtual displacement as a definite function $\boldsymbol{N} + \delta\boldsymbol{N}$ of ϑ, and we will call this function $\boldsymbol{N} + \delta\boldsymbol{N}$ the *mass action* associated with the virtual displacement.

Using index notation we will have:

$$d(x_h + \delta x_h) = dx_h + \sum_k \frac{\partial \delta x_h}{\partial x_k} dx_k + \frac{\partial \delta x_h}{\partial \vartheta} d\vartheta \qquad \begin{pmatrix} k = 1, 2, 3, 4 \\ h = 1, 2, 3, 4 \end{pmatrix}. \tag{9}$$

Now it soon becomes evident by reason of the remarks already made above that the value of $\boldsymbol{N} + \delta\boldsymbol{N}$ for parameter value ϑ will be:

$$\boldsymbol{N} + \delta\boldsymbol{N} = \iiiint \nu \frac{d(\tau + \delta\tau)}{d\tau} dx dy dz dt, \tag{10}$$

taken over the sickle, where $d(\tau + \delta\tau)$ means the quantity derived from

$$\sqrt{-(dx_1 + d\delta x_1)^2 - (dx_2 + d\delta x_2)^2 - (dx_3 + d\delta x_3)^2 - (dx_4 + d\delta x_4)^2}$$

by means of (9) and

$$dx_1 = w_1 d\tau, \quad dx_2 = w_2 d\tau, \quad dx_3 = w_3 d\tau, \quad dx_4 = w_4 d\tau, \quad d\vartheta = 0$$

thus we have

$$\frac{d(\tau + \delta\tau)}{d\tau} = \sqrt{-\sum_h \left(w_h + \sum_k \frac{\partial \delta x_h}{\partial x_k} w_k \right)^2} \qquad \begin{pmatrix} k = 1, 2, 3, 4 \\ h = 1, 2, 3, 4 \end{pmatrix}. \tag{11}$$

Now we want to subject the value of the differential quotient

$$\left(\frac{d(\boldsymbol{N} + \delta\boldsymbol{N})}{d\vartheta} \right)_{(\vartheta = 0)} \tag{12}$$

to a transformation. Since every δx_h as function of the arguments $x_1, x_2, x_3, x_4, \vartheta$ vanishes in general for $\vartheta = 0$, we also have in general $\frac{\partial \delta x_h}{\partial x_k} = 0$ for $\vartheta = 0$. If we now put

$$\left(\frac{\partial \delta x_h}{\partial \vartheta} \right)_{\vartheta = 0} = \xi_h \qquad (h = 1, 2, 3, 4), \tag{13}$$

then it follows by reason of (10) and (11) for the expression (12):

$$-\iiiint \nu \sum_h w_h \left(\frac{\partial \xi_h}{\partial x_1} w_1 + \frac{\partial \xi_h}{\partial x_2} w_2 + \frac{\partial \xi_h}{\partial x_3} w_3 + \frac{\partial \xi_h}{\partial x_4} w_4 \right) dx dy dz dt.$$

For the systems x_1, x_2, x_3, x_4 on the boundary of the sickle, $\delta x_1, \delta x_2, \delta x_3, \delta x_4$ are to
[105] vanish for all values of ϑ, and | therefore also $\xi_1, \xi_2, \xi_3, \xi_4$ are everywhere zero. Accordingly the last integral changes by partial integration into

$$\iiiint \sum_h \xi_h \left(\frac{\partial \nu w_h w_1}{\partial x_1} + \frac{\partial \nu w_h w_2}{\partial x_2} + \frac{\partial \nu w_h w_3}{\partial x_3} + \frac{\partial \nu w_h w_4}{\partial x_4} \right) dx\ dy\ dz\ dt.$$

Herein the expression in parentheses is

$$= w_h \sum_k \frac{\partial \nu w_k}{\partial x_k} + \nu \sum_k w_k \frac{\partial w_h}{\partial x_k}.$$

Here the first sum vanishes due to the continuity condition (6), the second one can be represented as

$$\frac{\partial w_h}{\partial x_1}\frac{dx_1}{d\tau} + \frac{\partial w_h}{\partial x_2}\frac{dx_2}{d\tau} + \frac{\partial w_h}{\partial x_3}\frac{dx_3}{d\tau} + \frac{\partial w_h}{\partial x_4}\frac{dx_4}{d\tau} = \frac{dw_h}{d\tau} = \frac{d}{d\tau}\left(\frac{dx_h}{d\tau}\right)$$

where $d/d\tau$ indicates differential quotients in the direction of the spacetime line of one location. Hence this finally results in *the expression for the differential quotient* (12)

$$\iiiint \nu \left(\frac{dw_1}{d\tau}\xi_1 + \frac{dw_2}{d\tau}\xi_2 + \frac{dw_3}{d\tau}\xi_3 + \frac{dw_4}{d\tau}\xi_4 \right) dx\ dy\ dz\ dt. \tag{14}$$

For a virtual displacement in the sickle we had demanded in addition that the points, considered as points of a substance, should proceed normal to the curves that are formed from them; this means for $\vartheta = 0$ that the ξ_h have to satisfy the *condition*

$$w_1\xi_1 + w_2\xi_2 + w_3\xi_3 + w_4\xi_4 = 0. \tag{15}$$

If we recall the Maxwell stresses in the electrodynamics of bodies at rest and consider on the other hand our results in the sections 12 and 13, then a certain *adaptation of Hamilton's principle* for continuous extended elastic media to the *relativity postulate* suggests itself.

Let there be specified at every spacetime point (as in section 13) a spacetime matrix of the second kind

$$S = \left| \begin{matrix} S_{11}, S_{12}, S_{13}, S_{14}, \\ S_{21}, S_{22}, S_{23}, S_{24}, \\ S_{31}, S_{32}, S_{33}, S_{34}, \\ S_{41}, S_{42}, S_{43}, S_{44}, \end{matrix} \right| = \left| \begin{matrix} X_x, & Y_x, & Z_x, & -iT_x \\ X_y, & Y_y, & Z_y, & -iT_y \\ X_z, & Y_z, & Z_z, & -iT_z \\ -iX_t, & -iY_t, & -iZ_t, & T_t \end{matrix} \right| \tag{16}$$

where $X_x, Y_x, \ldots Z_z, T_x, \ldots X_t, \ldots T_t$ are real quantities. |
For a virtual displacement in a spacetime sickle with the notation as above, let the value of the integral [106]

$$W + \delta W = \iiiint \Big(\sum_{h,k} S_{hk} \frac{\partial(x_k + \delta x_k)}{\partial x_h} \Big) dxdydzdt, \tag{17}$$

over the region of the sickle be called the *stress action* associated with the virtual displacement.

The sum that occurs here is, written out in more detail and with real quantities

$$\begin{aligned} & X_x + Y_y + Z_z + T_t \\ & + X_x \frac{\partial \delta x}{\partial x} + X_y \frac{\partial \delta x}{\partial y} + \ldots + Z_z \frac{\partial \delta z}{\partial z} \\ & - X_t \frac{\partial \delta x}{\partial t} - \ldots + T_x \frac{\partial \delta t}{\partial x} + \ldots + T_t \frac{\partial \delta t}{\partial t}. \end{aligned}$$

Now we will postulate the following *principle of least action* [*Minimalprinzip*] *for mechanics:*

Whenever any spacetime sickle is delimited, then for each virtual displacement in the sickle the sum of the mass action and of the stress action shall always be an extremum for the actually occurring behavior of the spacetime lines in the sickle.

The point of this statement is that for each virtual displacement we shall have (using the notation explained above)

$$\Big(\frac{d(\delta \boldsymbol{N} + \delta W)}{d\vartheta} \Big)_{\vartheta = 0} = 0. \tag{18}$$

By the methods of variational calculus the following four differential equations follow immediately by transformation (14) from this principle of least action, taking into account the condition (15)

$$\nu \frac{dw_h}{d\tau} = K_h + \kappa w_h \qquad (h = 1, 2, 3, 4), \tag{19}$$

where

$$K_h = \frac{\partial S_{1h}}{\partial x_1} + \frac{\partial S_{2h}}{\partial x_2} + \frac{\partial S_{3h}}{\partial x_3} + \frac{\partial S_{4h}}{\partial x_4} \tag{20}$$

are the components of the spacetime vector of the first kind $K = \mathrm{lor} S$ and κ is a factor to be determined from $w\bar{w} = -1$. By multiplying (19) by w_h and then | summing over $h = 1, 2, 3, 4$ one finds $\kappa = K\bar{w}$ and $K + (K\bar{w})w$ clearly becomes a spacetime vector of the first kind *normal* to w. Writing the components of this vector as
[107]

$$X, Y, Z, iT,$$

we arrive at the following *laws for the motion of matter:*

$$
\begin{aligned}
\nu\frac{d}{d\tau}\frac{dx}{d\tau} &= X, \\
\nu\frac{d}{d\tau}\frac{dy}{d\tau} &= Y, \\
\nu\frac{d}{d\tau}\frac{dz}{d\tau} &= Z, \\
\nu\frac{d}{d\tau}\frac{dt}{d\tau} &= T.
\end{aligned}
\tag{21}
$$

Here we have

$$\left(\frac{dx}{d\tau}\right)^2 + \left(\frac{dy}{d\tau}\right)^2 + \left(\frac{dz}{d\tau}\right)^2 = \left(\frac{dt}{d\tau}\right)^2 - 1$$

and

$$X\frac{dx}{d\tau} + Y\frac{dy}{d\tau} + Z\frac{dZ}{d\tau} = T\frac{dt}{d\tau},$$

and by reason of these relations the fourth of the equations (21) could be viewed as a consequence of the first three of them.

From (21) we further derive the laws for the motion of a *point mass*, that is for the course of an infinitely thin spacetime thread.

Let x, y, z, t denote a point of the center line, defined arbitrarily within the thread. We form the equations (21) for the points of the thread's *normal section* through x, y, z, t and integrate their product by the volume element [*Inhaltselement*] of the section over the entire region of the normal section. Let the integrals of the right-hand sides be R_x, R_y, R_z, R_t and m be the constant mass of the thread, yielding

$$
\begin{aligned}
m\frac{d}{d\tau}\frac{dx}{d\tau} &= R_x, \\
m\frac{d}{d\tau}\frac{dy}{d\tau} &= R_y, \\
m\frac{d}{d\tau}\frac{dz}{d\tau} &= R_z, \\
m\frac{d}{d\tau}\frac{dt}{d\tau} &= R_t.
\end{aligned}
\tag{22}
$$

| Here R, with the components R_x, R_y, R_z, iR_t, is again a spacetime vector of the first kind, which is *normal* to spacetime vector of the first kind w, the velocity of the mass point, with components [108]

$$\frac{dx}{d\tau}, \qquad \frac{dy}{d\tau}, \qquad \frac{dz}{d\tau}, \qquad i\frac{dt}{d\tau}.$$

We will call this vector R the *accelerating force* [*bewegende Kraft*] of the mass point.

If, on the other hand, one integrates the equations not over the thread's normal section but correspondingly over a cross section *normal to the t-axis* and passing through x, y, z, t then one obtains (see (4)) the equations (22) multiplied by $d\tau/dt$; in particular the last equations is

$$m\frac{d}{dt}\Big(\frac{dt}{d\tau}\Big) = \mathfrak{w}_x R_x \frac{d\tau}{dt} + \mathfrak{w}_y R_y \frac{d\tau}{dt} + \mathfrak{w}_z R_z \frac{d\tau}{dt}.$$

The right-hand side will have to be interpreted as the *work done* on the point mass per unit time. The equation itself will then be viewed as the *energy theorem* for the motion of the point mass, and the expression

$$m\Big(\frac{dt}{d\tau} - 1\Big) = m\Big(\frac{1}{\sqrt{1-\mathrm{w}^2}} - 1\Big) = m\Big(\frac{1}{2}|\mathfrak{w}|^2 + \frac{3}{8}|\mathfrak{w}|^4 + \cdots\Big)$$

will be identified with the *kinetic energy* of the point mass.

Since one always has $dt > d\tau$, one could characterize the quotient $(dt - d\tau)/(d\tau)$ as the advancement of time with respect to proper time of the point mass, and then put it as follows: the kinetic energy of the point mass is the product of its mass and the advancement of time with respect to its proper time.

The *quadruplet* of equations (22) again shows the full symmetry in x, y, z, it as demanded by the relativity postulate, *where the fourth equation*, analogous to what we have encountered in electrodynamics, *can be said, as it were, to be more highly evident physically.* On the basis of this symmetry the triplet of the first three equations is to be constructed according to the pattern of the fourth equation, and in view of this circumstance the claim is justified: *If the relativity postulate is put at the head*
[109] *of mechanics, then the complete | laws of motion follow solely from the energy theorem.*

I would not like omit making it plausible that a contradiction to the assumptions of the relativity postulate is not to be expected from the phenomena of *gravitation*.[2]

For $B^*(x^*, y^*, z^*, t^*)$ a fixed spacetime point, the region of all spacetime points $B(x, y, z, t)$ satisfying

$$(x-x^*)^2 + (x-y^*)^2 + (z-z^*)^2 = (t-t^*)^2, \qquad t - t^* \geqq 0 \tag{23}$$

shall be called the *ray object* [future lightcone] of the spacetime point B^*.

This object intersects any spacetime line only in a single spacetime point B, as follows on the one hand from the *convexity* of the object, and on the other hand from the circumstance that all possible directions of the spacetime line are directed from B^* only toward the concave side of the object. B^* shall then be called a *light point* of B.

2 H. Poincaré *Rend. Circ. Matem. Palermo*, Vol. XXI (1906), p. 129 [in this volume] has attempted to make the Newtonian law of attraction compatible with the relativity postulatem, along quite different lines than what I present here.

If the point $B(x, y, z, t)$ in the condition (23) is considered as fixed, and the point $B^*(x^*, y^*, z^*, t^*)$ as variable, then the same relation represents the region of all spacetime points B^* that are light points of B, and one shows in an analogous way that on any spacetime line there always occurs just a single point B^* that is a light point of B.

Let a mass point F of mass m experience an accelerating force, in the presence of another mass point F^* of mass m^* according to the following law. Consider the spacetime threads of F and F^* with center lines contained in them. Let BC be an infinitesimal element of the center line of F, further let B^* and C^* be the light points of B and C, respectively, on the center line of F^*, and let OA' be the radius vector parallel to B^*C^* of the fundamental hyperbolic shell (2), and finally let D^* be the intersection point of the line B^*C^* with its normal space that passes through B. *The accelerating force of the point mass F at the spacetime point B shall now be that spacetime vector of the first kind, normal to BC, which is formed additively from the vector*

$$mm^*\left(\frac{OA'}{B^*D^*}\right)^3 BD^* \tag{24}$$

in the direction BD and a suitable vector* | *in the direction B*C**. Here OA'/B^*D^* [110]
is understood to be the ratio of the two parallel vectors concerned.

It is clear that this definition is to be characterized as covariant with respect to the Lorentz group.

Now we ask how the spacetime thread of F behaves according to these considerations if the point mass F^* executes uniform translational motion, so that the center line of the thread of F^* is a straight line. We shift the spacetime origin O to it and can then introduce this straight line as the t-axis by means of a Lorentz transformation. Now let x, y, z, t mean the point B and τ^* the proper time of the point B^*, with origin at O. Our definition then leads to the equations

$$\frac{d^2x}{d\tau^2} = -\frac{m^*x}{(t-\tau^*)^3}, \qquad \frac{d^2y}{d\tau^2} = -\frac{m^*y}{(t-\tau^*)^3}, \qquad \frac{d^2z}{d\tau^2} = -\frac{m^*z}{(t-\tau^*)^3} \tag{25}$$

and

$$\frac{d^2t}{d\tau^2} = -\frac{m^*}{(t-\tau^*)^2}\frac{d(t-\tau^*)}{dt}, \tag{26}$$

where we have

$$x^2+y^2+z^2 = (t-\tau^*)^2 \tag{27}$$

and

$$\left(\frac{dx}{d\tau}\right)^2+\left(\frac{dy}{d\tau}\right)^2+\left(\frac{dz}{d\tau}\right)^2 = \left(\frac{dt}{d\tau}\right)^2-1. \tag{28}$$

In view of (27) the three equations (25) read the same as the equations for the motion of a mass point under attraction by a fixed center according to Newton's law, except that the time t is replaced by the proper time τ of the mass point. The fourth equation (26) then gives the connection between proper time and time for the point mass.

Now let the orbit of the space point x, y, z for different τ be an ellipse with semi-major axis a and eccentricity e, and let E be its eccentric anomaly, $\boldsymbol{T}$ the increase in proper time for a full execution of an orbit, and finally $n\boldsymbol{T} = 2\pi$, so that for a suitable origin of τ Kepler's equation

$$n\tau = E - e\sin E \tag{29}$$

holds. If we also change the unit of time and denote the speed of light by c, then (28) becomes

$$\left(\frac{dt}{d\tau}\right)^2 - 1 = \left(\frac{m^*}{ac^2}\frac{1 + e\cos E}{1 - e\cos E}\right). \tag{30}$$

[111] | By neglecting c^{-4} compared to 1 it follows that

$$ndt = nd\tau\left(1 + \frac{1}{2}\frac{m^*}{ac^2}\frac{1 + e\cos E}{1 - e\cos E}\right),$$

which by using (29) results in[6]

$$nt + \text{const.} = \left(1 + \frac{1}{2}\frac{m^*}{ac^2}\right)n\tau + \frac{m^*}{ac^2}e\sin E. \tag{31}$$

Here the factor $\dfrac{m^*}{ac^2}$ is the square of the ratio of a certain mean speed of F in its orbit to the speed of light. If we substitute for m^* the mass of the Sun and for a the semi-major axis of the Earth's orbit, then this factor amounts to 10^{-8}.

A law of attraction for masses according to the formulation exhibited above in connection with the relativity postulate would also imply *propagation of gravitation with the speed of light*. In view of the smallness of the periodic term in (31) a decision *against* such a law and the suggested modified mechanics in favor of Newton's law of attraction with Newtonian mechanics should not be derivable from astronomical observations.

EDITORIAL NOTES

[1] The equations (10) to (12) in the actual body of the text (not reproduced here) represent the Lorentz transformations and read

$$\mathfrak{r}'_{\mathfrak{v}} = \frac{\mathfrak{r}_{\mathrm{v}} - qt}{\sqrt{1-q^2}} \text{ for the direction of } \mathfrak{v}, \tag{10}$$

$$\mathfrak{r}'_{\bar{\mathfrak{v}}} = \mathfrak{r}_{\bar{\mathfrak{v}}} \text{ for any direction } \bar{\mathfrak{v}} \text{ perpendicular to } \mathfrak{v}, \tag{11}$$

$$\text{and further } t' = \frac{-q\mathfrak{r}_{\mathfrak{v}} + t}{\sqrt{1-q^2}}. \tag{12}$$

[2] For some of Minkowski's technical terms, the English equivalent according to present-day physics is given in square brackets.

[3] The references are to sections in the actual body of the text. $\mathfrak{w}$ denotes the three-dimensional velocity vector of matter at the given spacetime point, w denotes the corresponding four-vector:

$$\left(\frac{\mathfrak{w}}{\sqrt{1-\mathfrak{w}^2}}, \frac{i}{\sqrt{1-\mathfrak{w}^2}}\right).$$

[4] Minkowski defines lor as the 1×4 matrix

$$\left(\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \frac{\partial}{\partial x_3}, \frac{\partial}{\partial x_4}\right).$$

[5] Equation (67) in the body of the text reads

$$\mathrm{lor}\bar{s} = \frac{\partial s_1}{\partial x_1} + \frac{\partial s_2}{\partial x_2} + \frac{\partial s_3}{\partial x_3} + \frac{\partial s_4}{\partial x_4}. \tag{67}$$

[6] The eccentricity e in the second term on the right-hand side of eq. (32) has been inserted, correcting an obvious omission.

HENDRIK A. LORENTZ

OLD AND NEW QUESTIONS IN PHYSICS (EXCERPT)

Originally published as "Alte und neue Fragen der Physik" in Physikalische Zeitschrift, 11, 1910, pp. 1234–1257. Based on the Wolfskehl lectures given in Göttingen, 24–29 October 1910. The second and third lectures are translated here (pp. 1236–1244). Lorentz made use of Born's notes from the lecture and submitted the paper on 2 November 1910.

SECOND LECTURE

To discuss Einstein's *principle of relativity* here at Göttingen, where Minkowski was active, seems to me a particularly welcome task.

The significance of this principle can be illuminated from several different angles. We will not speak here of the mathematical aspect of the question, which was given such a splendid treatment by Minkowski, and which was further developed by Abraham, Sommerfeld and others. Rather, after some epistemological remarks about the concepts of space and time, the physical phenomena that may contribute to an experimental test of the principle shall be discussed.

The principle of relativity claims the following: If a physical phenomenon is described by certain equations in the system of reference x, y, z, t, then a phenomenon will also exist that can be described by the same equations in another system of reference x', y', z', t'. Here the two systems of reference are connected by relations containing the speed of light c and expressing the motion of one system with a uniform velocity relative to the other.

If observer A is located in the first, and B in the second system of reference, and each is supplied with measuring rods and clocks at rest in his system, then A will measure the values of x, y, z, t, and B the values of x', y', z', t', where it should be noted that A and B can also use one and the same measuring rod and the same clock. We have to assume that when the first observer somehow hands his rod and clock over to the second observer, they automatically assume the proper length and the proper rate so that B arrives at the values x', y', z', t' from his measurements. Either one will then find the same value for the speed of light, and will quite generally be able to make the same observations.

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 3
Gravitation in the Twilight of Classical Physics: Between Mechanics, Field Theory, and Astronomy.

Assume there is an aether; then among all systems x, y, z, t a single one would be distinguished by the state of rest of its coordinate axes as well as its clock in the aether. If one associates with this the idea (also held tenaciously by the speaker) that space and time are totally different from each other and that there is a "true time" (simultaneity would then exist independent of location, corresponding to the circumstance that we are able to imagine infinitely large velocities) then it is easily seen that this true time should be shown precisely by clocks that are at rest in the aether. Now, if the principle of relativity were generally valid in nature, then we would of course not be in a position to determine whether the system of reference being used at the moment is the distinguished one. Thus one arrives at the same results as those found when one denies the existence of the aether and of the true time, and regards all systems of reference as equivalent, following Einstein and Minkowski. It is surely up to each individual which of the two schools of thought he wishes to identify with.

In order to discuss the physical aspect of the question, we first have to establish
[1237] the transformation formulas, limiting ourselves to a special form | already used in the year 1887 by W. Voigt in his treatment of the Doppler principle; namely,

$$x' = x, \qquad y' = y, \qquad z' = az - bct, \qquad t' = at - \frac{b}{c}z;$$

where the constants $a > 0$, b satisfy the relation

$$a^2 - b^2 = 1,$$

which entails the identity $x'^2 + y'^2 + z'^2 - c^2t'^2 = x^2 + y^2 + z^2 - c^2t^2$. The origin of the system x', y', z' moves with respect to the system x, y, z in the z- direction with speed $(b/a)c$, which is always less than c. In general we have to assume that every velocity is less than c.

All state variables of any phenomenon, measured in one or the other system are connected by certain transformation formulas. For example, for the speed of a point these are

$$\mathfrak{v}_x' = \frac{\mathfrak{v}_x}{\omega}, \qquad \mathfrak{v}_y' = \frac{\mathfrak{v}_y}{\omega}, \qquad \mathfrak{v}_z' = \frac{a\mathfrak{v}_z - bc}{\omega},$$

where

$$\omega = a - \frac{b\mathfrak{v}_z}{c}.$$

Further we consider a system of points whose velocity is a continuous function of the coordinates. Let dS be a volume element surrounding the point $P(x, y, z)$ at time t; to this value of t there corresponds according to the transformation equations a point P in time t' in the other system of reference, and every point lying in dS at time t has certain [coordinates] x', y', z' for this fixed value of t'. The points x', y', z' fill a volume element dS', which is related to dS as follows[1]

$$dS' = \frac{dS}{\omega}.$$

Let us imagine some agent (matter, electricity etc.) connected with the points, and let us assume that the observer B has occasion to associate with each point the same amount of the agent as the observer A, then the spatial densities must obviously be in the inverse ratio as the volume elements, that is

$$\rho' = \omega\rho.$$

All these relations are reciprocal, that is, the primed and unprimed letters may be interchanged if at the same time b is replaced by $-b$.

The basic equations of the electromagnetic field retain their form under the transformation if the following quantities are introduced:

$$\mathfrak{d}_x' = a\mathfrak{d}_x - b\mathfrak{h}_y, \qquad \mathfrak{d}_y' = a\mathfrak{d}_y + b\mathfrak{h}_x, \qquad \mathfrak{d}_z' = \mathfrak{d}_z,$$
$$\mathfrak{h}_x' = a\mathfrak{h}_x + b\mathfrak{d}_y, \qquad \mathfrak{h}_y' = a\mathfrak{h}_y - b\mathfrak{d}_x, \qquad \mathfrak{h}_z' = \mathfrak{h}_z.$$

Thus in the system x', y', z', t' the following equations hold between these quantities, the transformed space density ρ' and the transformed velocity $\mathfrak{v}'$:

$$\begin{aligned}
\operatorname{div} \mathfrak{d}' &= \rho', \\
\operatorname{div} \mathfrak{h}' &= 0, \\
\operatorname{curl} \mathfrak{h}' &= \frac{1}{c}(\dot{\mathfrak{d}}' + \rho'\mathfrak{v}'), \\
\operatorname{curl} \mathfrak{d}' &= -\frac{1}{c}\dot{\mathfrak{h}}'.
\end{aligned}$$

With this the field equations of the electron theory satisfy the principle of relativity; but there is still the matter of harmonizing the equations of motion of the electrons themselves with this principle.

We will consider somewhat more generally the motion of an arbitrary material point. Here it is useful to introduce the concept of “proper time,” Minkowski’s beautiful invention. According to this there belongs to each point a time of its own, as it were, which is independent of the system of reference chosen; its differential is defined by the equation

$$d\tau = \sqrt{1 - \frac{\mathfrak{v}^2}{c^2}}dt.$$

The expressions formed with the aid of the proper time τ,

$$\frac{d}{d\tau}\frac{dx}{d\tau}, \qquad \frac{d}{d\tau}\frac{dy}{d\tau}, \qquad \frac{d}{d\tau}\frac{dz}{d\tau},$$

which are linear homogeneous functions of the components of the ordinary acceleration, will be called the components of the “Minkowskian acceleration.” We describe the motion of a point by the equations:

$$m\frac{d}{d\tau}\frac{dx}{d\tau} = \mathfrak{K}_x, \text{ etc.},$$

where m is a constant, which we call the “Minkowskian mass.” We designate the vector $\mathfrak{K}$ as the “Minkowskian force.”

It is then easy to derive the transformation formulas for this acceleration and this force; we leave m unchanged. Thus we have

$$\mathfrak{K}_x' = \mathfrak{K}_x, \qquad \mathfrak{K}_y' = \mathfrak{K}_y, \qquad \mathfrak{K}_z' = a\mathfrak{K}_z - \frac{b}{c}(\mathfrak{v}\cdot\mathfrak{K}).$$

The essential point is the following. The principle of relativity demands that if for an actual phenomenon the Minkowskian forces depend in a certain way on the coordinates, velocities, etc. in one system of reference, then the transformed Minkowskian
[1238] forces | depend in the same way on the transformed coordinates, velocities etc. in the other system of reference. This is a special property that must be shared by all forces in nature if the principle of relativity is to be valid. Presupposing this we can calculate the forces acting on moving bodies if we know them for the case of rest. For example, if an electron of charge e is in motion, we consider a system of reference in which it is momentarily at rest. Then the electron in this system is under the influence of the Minkowskian force

$$\mathfrak{K} = c\mathfrak{d};$$

from this it follows by application of the transformation equations for $\mathfrak{K}$ and $\mathfrak{d}$ that the Minkowskian force acting on the electron that moves with velocity $\mathfrak{v}$ in an arbitrary coordinate system amounts to

$$\mathfrak{K} = c\frac{\mathfrak{d} + \frac{1}{c}[\mathfrak{v}\cdot\mathfrak{h}]}{\sqrt{1-\frac{\mathfrak{v}^2}{c^2}}}.$$

This formula does not agree with the usual ansatz of the electron theory, because of the presence of the denominator. The difference is due to the fact that usually one does not operate with our Minkowskian force, but with the “Newtonian force” $\mathfrak{F}$, and we see that these two forces are related as follows for an electron:

$$\mathfrak{F} = \mathfrak{K}\sqrt{1-\frac{\mathfrak{v}^2}{c^2}}.$$

It is to be assumed that this relation is valid for arbitrary material points.

Thus the phenomena of motion can be treated in two different ways, using either the Minkowskian or the Newtonian force. In the latter case the equations of motion take the form

$$\mathfrak{F} = m_1 \mathfrak{j}_1 + m_2 \mathfrak{j}_2,$$

and here $\mathfrak{j}_1$ means the ordinary acceleration in the direction of motion, $\mathfrak{j}_2$ the ordinary normal acceleration, and the factors

$$m_1 = \frac{m}{\sqrt{\left(1 - \frac{\mathfrak{v}^2}{c^2}\right)^3}}, \qquad m_2 = \frac{m}{\sqrt{1 - \frac{\mathfrak{v}^2}{c^2}}},$$

are called the "longitudinal" and the "transverse mass."

Just like the Minkowskian forces, the Newtonian forces that occur in nature must also fulfill certain conditions in order to satisfy the relativity principle. This is the case if, for example, a normal pressure of a constant magnitude p per unit area acts on a surface regardless of the state of motion; then in the transformed system a normal pressure of the same magnitude acts on the corresponding moving surface element. Since we have already recognized the invariance of the field equations, the question of whether the motions in an electron system satisfy the relativity principle amounts merely to an experimental test of the formulas for the longitudinal and transverse masses m_1, m_2; although the experiments of Bucherer and Hupka seem to confirm these formulas, one has not yet arrived at a definitive decision.

Concerning the mass of the electron, one should remember that this is electromagnetic in nature; so it will depend on the distribution of charge within the electron. Therefore the formulas for the mass can be correct only if the charge distribution, and hence also the shape of the electron, vary in a definite way with the velocity. One must assume that an electron, which is a sphere when at rest, becomes an ellipsoid that is flattened in the direction of motion as a result of translation; the amount of flattening is

$$\sqrt{1 - \frac{\mathfrak{v}^2}{c^2}}.$$

If we assume that the shape and size of the electron are regulated by internal forces, then to agree with the relativity principle these forces must have properties such that this flattening occurs by itself when in motion. Regarding this Poincaré has made the following hypothesis. The electron is a charged, expandible skin, and the electrical repulsion of the different points of the electron is balanced by an inner normal tension of unchangeable magnitude. Indeed, according to the above such normal tensions satisfy the relativity principle.

In the same way all molecular forces acting within ponderable matter, as well as the quasi-elastic and resistive forces acting on the electron, have to satisfy certain conditions in order to be in accord with the relativity principle. Then every moving

body will be unchanged for a co-moving observer, but for an observer at rest it will experience a change in dimensions, which is a consequence of the change in molecular forces demanded by these conditions. This also leads automatically to the contraction of bodies, which was already
[1239] | devised earlier to explain the negative outcome of Michelson's interferometer experiment and of all similar experiments that were to determine an influence of the Earth's motion on optical phenomena.

Concerning rigid bodies, as investigated by Born, Herglotz, Noether, and Levi-Civita, the difficulties occurring in the consideration of rotation can surely be relieved by ascribing their rigidity to the action of particularly intense molecular forces.

Finally let us turn to *gravitation*. The relativity principle demands a modification of Newton's law, foremost a propagation of the effect with the speed of light. The possibility of a finite speed of propagation of gravity was already discussed by Laplace, who imagined as the cause of gravity a fluid streaming toward the Sun, which pushes the planets toward the Sun. He found that the speed c of this fluid must be assumed to be at least 100 million times larger than that of light, so that the calculations remain in agreement with the astronomical observations. The necessity of such a large value of c is due to the occurrence of v/c to the first power in his final formulas, where v is the speed of a planet. But if the propagation speed c of gravity is to have the value of the speed of light, as demanded by the relativity principle, then a contradiction with the observations can only be avoided if only quantities of second (and higher) order in v/c occur in the expression for the modified law of gravitation.

Restricting oneself to quantities of second order, one can, on the basis of a suggestive electron-theoretic analogy, easily give a condition that determines the modified law in a unique way. Namely, if one considers the force acting on an electron that moves with a velocity $\mathfrak{v}$,

$$e\left(\mathfrak{d} + \frac{1}{c}[\mathfrak{v}\cdot\mathfrak{h}]\right),$$

then the vectors $\mathfrak{d}$ and $\mathfrak{h}$ depend, in addition, on the velocities $\mathfrak{v}'$ of the electrons that produce the field; in the vector product $[\mathfrak{v}\cdot \mathrm{h}]$, products of the form $\mathfrak{v}\mathfrak{v}'$ do occur, but not the square $\mathfrak{v}^2$ of the speed of the electron under consideration. Accordingly let us assume that in the expression for the attraction acting on the point 1 due to point 2 there is no term in the square of the velocity $\mathfrak{v}_1^2$ of point 1. Then all velocities whatsoever must drop out in a system of reference in which point 2 is at rest $(\mathfrak{v}_2 = 0)$; therefore the law will reduce to the usual Newtonian one in this system. Now making the transition by transforming to an arbitrary coordinate system, one finds that the force acting on point 1 is composed of two parts, the first, an attraction in the direction of the line connecting them of magnitude

$$R + \frac{1}{c^2}\left\{\frac{1}{2}\mathfrak{v}_2^{\,2}R + \frac{1}{2}\mathfrak{v}_{2r}^{\,2}\left(r\frac{dR}{dr} - R\right) - (\mathfrak{v}_1\cdot\mathfrak{v}_2)R\right\},$$

the second, a force in the direction of $\mathfrak{v}_2$ of magnitude

$$\frac{1}{c^2}\mathfrak{v}_{1r}\, R\mathfrak{v}_2\,;$$

here r means the distance between two simultaneous positions of the two points, $\mathfrak{v}_r$ the component of $\mathfrak{v}$ along the connecting line drawn from 1 to 2, and R that function of r which represents the law of attraction in the case of rest ($R = k/r^2$ for Newtonian attraction, $R = kr$ for quasielastic forces). Note that "force" is always understood to be the "Newtonian force," not the "Minkowskian" one. Minkowski, by the way, has given a somewhat different expression for the law of gravity. The latter as well as the one described above can be found in Poincaré.

THIRD LECTURE

At the end of the previous lecture a modified law of gravitation was given, which is in agreement with the relativity principle. Concerning this one should note that the principle of equality of action and reaction is not satisfied.

Now the perturbations that can arise due to those additional second order terms will be discussed. Besides many short-period perturbations, which have no significance, there is a secular motion of the planets' perihelia. Prof. de Sitter computes $6.69''$ per century for Mercury's perturbations. Since Laplace, it has been known that Mercury has an anomalous perihelion motion of $44''$ per century; although this anomaly has the right sign, it is much too large to be explained by those additional terms. Instead, Seeliger attributes it to a perturbation by the carrier of the zodiacal light, | whose mass one can suitably determine in a plausible way. So, from this one [1240] can arrive at no decision, as long as the accuracy of astronomical measurements is not significantly increased. To be absolutely accurate one would also have to take into account the difference between the Earth's "proper time" and the time of the solar system.

A different method to test the validity of the modified law of gravitation can be based on a procedure suggested by Maxwell to decide whether the solar system moves through the aether. If this were the case, then the *eclipses of Jupiter's moons* should be advanced or delayed depending on Jupiter's position with respect to the Earth.

For if the Jupiter-Earth distance is a and the component of the solar system's velocity in the aether in the direction of the line connecting Jupiter to Earth is v, then the time required to cover the distance a in the case of rest, a/c, would be changed to $a/(c+v)$; thus the motion brings about an advance or delay, which amounts to av/c^2 up to terms of second order, and which takes on different values according to the value of the velocity component v, which of course depends on the position of the two planets. Now it is clear that such a dependence of the phenomena on the motion through the aether contradicts the relativity principle.

In order to clear up this contradiction let us simplify the situation schematically. We suppose that the Sun S has a mass that is infinitely large compared to that of the planet. Let the velocity of the solar system coincide with the z- axis, which we lay through the Sun. The intersection points of the planet's orbit with the z- axis are denoted as the upper resp. lower transit, A resp. B. (Fig. 1)

We place the observer at the Sun. At each transit of the planet through the z- axis a signal will propagate towards the Sun. The period of revolution shall be T. When the Sun is at rest the time between an upper and lower transit will be $(1/2)T$ for the assumed circular motion; the same is true for the time between the arrivals of the two light signals. By contrast, if the Sun is in motion in the z- direction, the signal from the upper transit must suffer an advance av/c^2, that from the lower transit a delay of the same amount; if the uniform orbital motion (assumed as self-evident by Maxwell) is preserved without perturbation, the time interval between the arrivals of the light signals of two successive passes would appear alternately increased and decreased by $2av/c^2$. Preservation of the uniform circular motion during a translation through the aether, as is assumed above, is, however, impossible according to the relativity principle. For if we describe the process in a coordinate system that does not take part in the motion, the modified law of gravitation will have to be applied, and this results in a non-uniform planetary motion, due to which the difference in time intervals between the arrivals of the light signals exactly cancels.

Therefore the determination of whether an advance or delay of the eclipses in fact occurs can be used to decide in favor or against the relativity principle. However, the numerical situation is again rather unfavorable. Thus Mr. Burton, who has access to 330 photometric observations of eclipses of Jupiter's first moon made at the Harvard observatory, estimates the probable error of the final result for v as 50 km/sec; on the other hand, one has observed speeds of stars of 70 km/sec, and the speed of the solar system with respect to the fixed stars is estimated at 20 km/sec. The relativity principle is therefore hardly supported by Burton's calculations; at best they could invalidate it, namely if, for example, the final result were a value exceeding 100 km/sec.

Let us leave it undecided whether or not the new mechanics will receive confirmation by astronomical observations. But we will not fail to familiarize ourselves with some of its basic formulas.

[1241] If one defines work as the scalar product of "Newtonian force" and | displacement, then the equations of motion yield the *energy principle* in its usual form, so that the work done per unit of time equals the increase in energy ε:

$$\mathfrak{F}_x \frac{dx}{dt} + \mathfrak{F}_y \frac{dy}{dt} + \mathfrak{F}_z \frac{dz}{dt} = \frac{d\varepsilon}{dt}.$$

Here *energy* is expressed by

$$\varepsilon = mc^2\left(\frac{1}{\sqrt{1-\frac{v^2}{c^2}}}-1\right);$$

this agrees up to second order terms with the value of the kinetic energy in customary mechanics:

$$\varepsilon = \frac{1}{2}\mathrm{mv}^2.$$

Furthermore, from the equations of motion one can derive *Hamilton's principle*

$$\int_{t_1}^{t_2}(\delta L + \delta A)dt = 0;$$

here δA is the work of the "Newtonian force" upon a virtual displacement and L is the *Lagrangian*, which takes the form

$$L = -mc^2\left(\sqrt{1-\frac{\mathfrak{v}^2}{c^2}}-1\right).$$

From Hamilton's principle one can conversely obtain the equations of motion. The quantities

$$\frac{\partial L}{\partial \dot{x}}, \qquad \frac{\partial L}{\partial \dot{y}}, \qquad \frac{\partial L}{\partial \dot{z}}$$

are to be identified as the *components of the momentum*.

All these formulas can be verified by the electromagnetic equations of motion for an electron. One then has to take the following value for the "Minkowskian mass" m,

$$m = \frac{e^2}{6\pi Rc^2},$$

and to add to the electric and magnetic energy the energy of those internal stresses which determine the shape of the electron, as we saw above. Thus from the general principle of least action for arbitrary electromagnetic systems, discussed in the first lecture, one can obtain Hamilton's principle for a point mass as given above by specialization to an electron, but the work of those internal stresses must be taken into account.

We now go over to a discussion of the *equations of the electromagnetic field for ponderable bodies*. These have been written down purely phenomenologically by Minkowski, then M. Born and Ph. Frank showed that they can also be derived from the ideas of electron theory; by the latter procedure Lorentz himself also found these equations, in a slightly different technical form.

To obtain relations between observable quantities one must smear out the details of the phenomena due to individual electrons by averaging over a large number of them. One is lead to the following equations (which are identical to those of the usual Maxwell theory):

$$\begin{aligned} \operatorname{div}\mathfrak{D} &= \rho_l, \\ \operatorname{div}\mathfrak{B} &= 0, \\ \operatorname{curl}\mathfrak{H} &= \frac{1}{c}(\mathfrak{C} + \dot{\mathfrak{D}}), \\ \operatorname{curl}\mathfrak{E} &= -\frac{1}{c}\dot{\mathfrak{B}}. \end{aligned}$$

Here $\mathfrak{D}$ is the dielectric displacement, $\mathfrak{B}$ the magnetic induction, $\mathfrak{H}$ the magnetic force, $\mathfrak{E}$ the electric force, $\mathfrak{C}$ the electric current, and ρ_l the density of the observable electric charges. Denoting mean values by an overbar we have, for example,

$$\mathfrak{E} = \bar{\mathfrak{d}}, \qquad \mathfrak{B} = \bar{\mathfrak{h}},$$

where $\mathfrak{d}$, $\mathfrak{h}$ have their former meaning; further we have

$$\begin{aligned} \mathfrak{D} &= \mathfrak{E} + \mathfrak{P}, \\ \mathfrak{H} &= \mathfrak{B} - \mathfrak{M} - \frac{1}{c}[\mathfrak{P} \cdot \mathfrak{w}], \end{aligned}$$

where $\mathfrak{P}$ is the electric moment and $\mathfrak{M}$ the magnetization per unit volume, and $\mathfrak{w}$ denotes the velocity of matter. When deriving these formulas one divides the electrons into three types. The first type, the polarization electrons, generate the electric moment $\mathfrak{P}$ by their displacement; the second type, the magnetization electrons, generate the magnetic state $\mathfrak{M}$ by their orbital motion; the third type, the conduction electrons, move freely within the matter and generate the observable charge density ρ_l and the current $\mathfrak{C}$. The latter is additionally to be divided into two parts; for if $\mathfrak{u}$ is the relative velocity of the electrons with respect to the matter, then the total velocity of the electrons is $\mathfrak{v} = \mathfrak{w} + \mathfrak{u}$, hence the current carried by them is

$$\mathfrak{C} = \overline{\rho\mathfrak{v}} = \bar{\rho}\mathfrak{w} + \overline{\rho\mathfrak{u}};$$

$\bar{\rho}$ is the observable charge ρ_l, $\bar{\rho}\mathfrak{w}$ is the convection current, and $\overline{\rho\mathfrak{u}}$ the conduction current proper.

[1242] There are transformation formulas for all these quantities, | and we give a few of them below:

$$\mathfrak{C}_x' = \mathfrak{C}_x, \qquad \mathfrak{C}_y' = \mathfrak{C}_y, \qquad \mathfrak{C}_z' = a\mathfrak{C}_z - bc\rho_l,$$

$$\rho_l' = a\rho_l - \frac{b}{c}\mathfrak{C}_z,$$

$$\mathfrak{P}_x' = a\mathfrak{P}_x - \frac{b}{c}(\mathfrak{w}_z\mathfrak{P}_x - \mathfrak{w}_x\mathfrak{P}_z) + b\mathfrak{M}_y,$$

$$\mathfrak{P}_y' = a\mathfrak{P}_y - \frac{b}{c}(\mathfrak{w}_z\mathfrak{P}_y - \mathfrak{w}_y\mathfrak{P}_z) + b\mathfrak{M}_x,$$

$$\mathfrak{P}_z' = \mathfrak{P}_z.$$

Further, the following auxiliary vectors are useful:

$$\mathfrak{H}_1 = \mathfrak{H} - \frac{1}{c}[\mathfrak{w} \cdot \mathfrak{D}], \qquad \mathfrak{B}_1 = \mathfrak{B} - \frac{1}{c}[\mathfrak{w} \cdot \mathfrak{E}],$$

$$\mathfrak{E}_1 = \mathfrak{E} + \frac{1}{c}[\mathfrak{w} \cdot \mathfrak{B}], \qquad \mathfrak{D}_1 = \mathfrak{D} + \frac{1}{c}[\mathfrak{w} \cdot \mathfrak{H}].$$

Now the field equations given above must still be completed by establishing the relations that exist between the vectors $\mathfrak{E}, \mathfrak{H}$ and $\mathfrak{D}, \mathfrak{B}$. These relations can be obtained in two ways.

The first, phenomenological method proceeds as follows: One considers an arbitrarily moving point of matter and introduces a system of reference in which it is at rest; if the element of volume surrounding the point is isotropic in the rest system, the equation appropriate for bodies at rest (between $\mathfrak{E}$ and $\mathfrak{D}$, for example)

$$\mathfrak{D} = \varepsilon\mathfrak{E}$$

holds; or equally well

$$\mathfrak{D}_1 = \varepsilon\mathfrak{E}_1,$$

because the auxiliary vectors $\mathfrak{D}_1, \mathfrak{E}_1$ are identical with $\mathfrak{D}, \mathfrak{E}$ when $\mathfrak{w} = 0$ But $\mathfrak{D}_1$ and $\mathfrak{E}_1$ transform in the same way, and this implies that also in the original system the equation

$$\mathfrak{D}_1 = \varepsilon\mathfrak{E}_1,$$

and correspondingly

$$\mathfrak{B}_1 = \mu\mathfrak{H}_1,$$

remains valid. Concerning the conduction current we remark only that it depends on $\mathfrak{E}_1$.

The second method has its roots in the mechanics of electrons. Just as the equation $\mathfrak{D} = \varepsilon\mathfrak{E}$ for bodies at rest turns out to be a consequence of the assumption of quasielastic forces, which restore the electrons to their rest positions, so one will obtain the equation $\mathfrak{D}_1 = \varepsilon\mathfrak{E}_1$ for moving bodies if one ascribes to the quasi-elastic forces the properties demanded by the relativity principle. The latter will be satisfied if one takes for these forces the expression of the generalized law of attraction, where R must be taken proportional to r.

The explanation of the resistance to conduction proceeds similarly. A satisfactory electron-theoretic explanation of the magnetic properties of matter is presently not at hand.

Finally the significance of the above equations shall be elucidated in three remarkable cases.

The *first remark* is connected with the equation

$$\rho_l' = a\rho_l - \frac{b}{c}\mathfrak{C}_z.$$

According to this, ρ_l' can vanish without having $\rho_l = 0$ if a current $\mathfrak{C}$ is present; that is, an observer A will declare a body to be charged that must be treated as uncharged by B moving relative to him. One can understand this by noting that every body contains an equal number of positive and negative electrons, which compensate in uncharged bodies. When the body moves with velocity $\mathfrak{w}$ in the presence of a conduction current, the two types of electrons will attain different total velocities, therefore the quantity $\omega = a - b(\mathfrak{v}_z/c)$ will also have different values for the two types. When an observer B moving with the body calculates the mean charge density $\bar{\rho}' = \overline{\omega\rho}$ for both types of electrons he can obtain zero for the sum, even when for an observer A in whose reference frame the body is moving the mean values $\bar{\rho}$ of the positive and negative electrons do not compensate.

This circumstance calls forth the memory of an old question. Around the year 1880 there was a great discussion among physicists about *Clausius' fundamental law* of electrodynamics. One attempted to derive a contradiction between this law and observations by concluding that according to the law a current-carrying conductor on the Earth would have to exert an influence on a co-moving charge e due to the motion of the Earth, which could have been observed. That the law actually does not demand this influence was noted by Budde; this is because the current due to the Earth's motion acts on itself and causes a "compensating charge" in the current-carrying conductor, which exactly cancels the first influence. The electron theory leads
[1243] to similar conclusions, and Lorentz finds |

$$\frac{1}{c^2}\mathfrak{w}_z\,\mathfrak{C}_z\,;$$

for the density of the compensating charge, if the velocity is in the direction of the z- axis; this must be assumed by an observer A who does not take part in the motion of the Earth, whereas it does not exist for a co-moving observer B. The value given above agrees exactly with the formula derived from the relativity principle; for if $\rho_l' = 0$, one finds from this formula

$$\rho_l = \frac{b}{ac}\mathfrak{C}_z,$$

and since, according to what was said in the second lecture on p. 288 [p. 1237 in the original], $\mathfrak{w}_z = bc/a$ is the speed of the two systems of reference with respect to each other, one indeed finds

$$\rho_l = \frac{1}{c^2}\mathfrak{w}_z\mathfrak{E}_z.$$

The *second remark* starts from the transformation equations for the electric moment $\mathfrak{P}$ p. 296 [p. 1241 in the original] in which the presence of the magnetization $\mathfrak{M}$ lets us recognize the impossibility of differentiating precisely between polarization- and magnetization electrons. Rather, in a magnetized body $(\mathfrak{M} \neq 0)$ $\mathfrak{P} = 0$, can vanish when judged from one system of reference, whereas in another $\mathfrak{P}'$ differs from zero. This will now be applied to a special case, where we confine attention to quantities of first order. The body we consider (such as a steel magnet) shall contain only conduction electrons and electrons that produce an $\mathfrak{M}$ but no $\mathfrak{P}$ when the body is at rest; it shall have the shape of an infinitely extended plate, bounded by two planes a, b; the middle plane shall be the yz-plane. (Fig. 2) When it is at rest a constant magnetization $\mathfrak{M}_y$ shall be present, whereas $\mathfrak{P} = 0$. When the body is given a speed v in the z-direction an observer who does not take part in the motion will observe the electric polarization

$$\mathfrak{P}_x = -\frac{v}{c}\mathfrak{M}_y.$$

Now we imagine at either side of the body two conductors c, d, which form together with it two equal condensers, and these shall be shorted out by a wire (from c to d). When there is motion, charges will be created on c and d, which can be calculated as follows. Since a current is clearly impossible in the x-direction, we have $\mathfrak{E}_{lx} = 0$ or $\mathfrak{E}_x = (v/c)\mathfrak{B}_y$. Since the process is stationary we have $\dot{\mathfrak{B}} = 0$; then the existence of a potential φ follows from $\mathrm{curl}\,\mathfrak{E} = 0$. If Δ is the thickness of the slab one has

$$\varphi_a - \varphi_b = \frac{v}{c}\Delta\mathfrak{B}_y.$$

From the symmetry of the arrangement it clearly follows that

$$\varphi_d - \varphi_a = \varphi_b - \varphi_c\ ,$$

and because the plates c, d are shorted out, we must have

$$\varphi_d = \varphi_c\ ;$$

this implies

$$\varphi_d - \varphi_a = -\frac{v}{2c}\Delta\mathfrak{B}_y.$$

If γ is the capacity of one of the two condensers, the charge of plate d becomes

$$-\frac{v}{2c}\gamma\Delta\mathfrak{B}_y,$$

and c receives the equal and opposite amount.

Now we compare this process with the inverse case, that the magnet a, b is at rest and the plates c, d move with the opposite velocity. According to the relativity principle everything would have to be the same as in the first case. Indeed one finds at once from the usual law of induction exactly the amount of charge on plate d given above. But this charge on d must now induce an equal and opposite one on the plane a of the magnet at rest, and corresponding statements must hold for b and c. Since no current can flow $(\mathfrak{C} = 0)$, there must be the same charges on the magnet, whether the magnet is moving and the plates are at rest or conversely. So we have to think how it happens that in the first case the opposite charge appears on the plane a of the moving magnet as on the plate d; this becomes possible only due to the polar-
[1244] ization $\mathfrak{P}_x = -(v/c)\mathfrak{M}_y$ | produced by the motion. For one has

$$\mathfrak{D}_x = \mathfrak{E}_x + \mathfrak{P}_x = \frac{v}{c}\mathfrak{B}_y - \frac{v}{c}\mathfrak{M}_y;$$

since here $\mathfrak{P}$ is to be neglected to first order in the velocity, that is the term $[\mathfrak{P}\cdot\mathfrak{w}]$, we have

$$\mathfrak{B} - \mathfrak{M} = \mathfrak{H},$$

But $\mathfrak{H}$ is zero because we assume the plate to be infinitely extended. This implies

$$\mathfrak{D}_x = 0,$$

i.e., in the moving plate there is no dielectric displacement, so the charge on a corresponds to that on d, as the relativity principle demands.

The *last remark* concerns again the circumstance that according to the relativity principle the motion of the Earth cannot have any influence on electromagnetic processes. But Liénard has pointed out a phenomenon where such an influence is to be expected, an influence of first order in magnitude; Poincaré has also discussed this case in his book *Electricité et Optique*. It concerns the ponderomotive force on a conductor. To determine this force one may make the suggestive ansatz for the force acting on the conduction electrons per unit charge:

$$\mathfrak{E}_1 = \mathfrak{E} + \frac{1}{c}[\mathfrak{v}\cdot\mathfrak{B}];$$

then this results in the force caused by the Earth's motion on the conductor in the direction of the motion by an amount

$$\frac{1}{c^2}(\mathfrak{C}_l\cdot\mathfrak{E})\mathfrak{w}_z;$$

since $(\mathfrak{C}_l \cdot \mathfrak{E})$ is the heat generated by the conduction current $\mathfrak{C}$ this expression is easily calculated numerically (which admittedly results in a value inaccessible to observation).

If one now asks oneself, how this result that contradicts the relativity principle can come about, one finds that indeed one has not calculated the force acting on the matter of the conductor, but on the electrons moving inside the conductor. The latter force must first be transferred to the matter by forces, which are unknown to us in detail, and that happens without change of magnitude only if action equals reaction for the forces between matter and electrons. But for moving bodies action does not equal reaction in this case according to the relativity principle, and this circumstance exactly compensates Liénard's force.

In summary, one can say that there is little prospect for an experimental confirmation of the principle of relativity; except for a few astronomical observations, only measurements of the electron mass are worth considering. But one must not forget that the outcome of the negative experiments, such as Michelson's interference experiment and the experiments to find a double refraction caused by the Earth's motion, can only be explained by the relativity principle.

EDITORIAL NOTE

[1] In the original, the denominator is missing from the right-hand side.

THE PROBLEM OF GRAVITATION AS A CHALLENGE FOR THE MINKOWSKI FORMALISM

JÜRGEN RENN

THE SUMMIT ALMOST SCALED: MAX ABRAHAM AS A PIONEER OF A RELATIVISTIC THEORY OF GRAVITATION

1. THE FRAGILE LADDER OF THE MINKOWSKI FORMALISM

1.1 Abraham's Bold Step

Today Max Abraham is known mainly for his achievements in the field of electrodynamics and, in particular, for the successful series of textbooks associated with his name.[1] He is, however, largely forgotten as a pioneer of a relativistic theory of gravitation. The papers he dedicated to the subject between 1911 and 1915 are mainly remembered for the controversy with Einstein that they document.[2] In hindsight it is clear that Abraham's approach to a relativistic theory of gravitation—an attempt to formulate a field theory of gravitation in the framework of Minkowski's formalism—would lead to a dead end. However, only by exploring the consequences of this approach did it eventually become clear that it did not lead anywhere. And it was, after all, the failure of Abraham's bold step which encouraged others to either pursue his endeavor through more appropriate means, as was the case with Gunnar Nordström, or to take even bolder steps than Abraham and attempt even higher summits, as was the case with Einstein.[3] In fact, it was largely due to Abraham's efforts that Einstein became familiar with the limits of Minkowski's formalism and also learned how to overcome them.

In the following, we will first review Abraham's attempt to take up the challenge of using Minkowski's formalism as the framework for a relativistic theory of gravitation and then show how this bold step led to an ardent controversy with Einstein which, for Abraham, eventually led to a rejection of relativity theory altogether. We will then consider some of the insights and achievements that Abraham attained in the course of his research, which have been largely forgotten because his approach turned out to lead to a dead end. In short, we will portray Abraham as someone who

1 See (Abraham and Föppl 1904–1908) and the subsequent editions of this work.
2 See, in particular (Abraham 1912e, 1912f) and for a historical discussion (Cattani and De Maria 1989), to which the following account is much indebted.
3 See the papers by Nordström in this volume.

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 3
Gravitation in the Twilight of Classical Physics: Between Mechanics, Field Theory, and Astronomy.

almost scaled the summit, who aimed high but failed to reach the goal he envisioned. As a consequence of his failure and of criticism by others, he eventually gave up mountain climbing altogether but, at the same time, encouraged others to attempt the summit he had failed to reach, not least because of the magnificent vistas about which, on the basis of his experience, he could report.

The starting point for Abraham's work on a relativistic theory of gravitation were some of the insights that Einstein had attained on the basis of his equivalence principle, in particular the idea of a variable speed of light.[4] Precisely because of this insight, Einstein did not consider Minkowski's formalism to be a useful tool for building up a relativistic theory of gravitation, since he took the constancy of the speed of light to be one of its fundamental principles. Moreover, Einstein was skeptical of such a path for other reasons as well.[5] Abraham, on the other hand, took the bold step of modifying Minkowski's framework by accommodating it to the assumption of a variable speed of light. In this way, he succeeded in overcoming the unacceptably restrictive conditions imposed by the constancy of the speed of light on a straightforward, special relativistic theory of gravitation. Abraham thus became the first to exploit the mathematical potential of Minkowski's four-dimensional formalism for a theory of gravitation.

In the introduction to a paper submitted in December 1911 and published in January 1912, Abraham first of all acknowledges his debt to Einstein's idea concerning the relation between a variable speed of light and the gravitational potential:

> In a recently published paper A. Einstein proposed the hypothesis that the speed of light (c) depends on the gravitational potential (Φ). In the following note, I develop a theory of the gravitational force which satisfies the principle of relativity and derive from it a relation between c and Φ, which in first approximation is equivalent to Einstein's.[6]

Abraham then adopts Minkowski's formalism for his purposes:

> Following Minkowski's presentation we regard
>
> $$x, y, z \qquad \text{and} \qquad u = il = ict \tag{1}$$
>
> as the coordinates of a four-dimensional space.[7]

---

4 See "The First Two Acts" and "The Third Way to General Relativity" (in vol. 1 and vol. 3 of this series respectively).

5 See, for instance, Einstein's comments in his contemporary letters, e.g., to Wilhelm Wien, 11 March 1912 and 17 May 1912, (CPAE 5, Doc. 371 and 395) to which we shall refer later. See also the historical discussion in the "Editorial Note" in (CPAE 4, 122–128).

6 "In einer vor kurzem erschienenen Arbeit hat A. Einstein die Hypothese aufgestellt, daß die Geschwindigkeit des Lichtes (c) vom Gravitationspotential (Φ) abhänge. In der folgenden Note entwickle ich eine Theorie der Schwerkraft, welche dem Prinzip der Relativität genügt, und leite aus ihr eine Beziehung zwischen c und Φ ab, die in erster Annäherung mit der Einsteinschen gleichwertig ist." See (Abraham 1912h, 1). A complete English translation of this paper is given in this volume.

7 "In dem wir der Darstellung Minkowskis folgen, betrachten wir (x, y, z) und $(u = il = ict)$ als Koordinaten eines vierdimensionalen Raumes." (Abraham 1912h, 1)

He immediately proceeds to state a differential equation for the gravitational potential, which essentially corresponds to a four-dimensional generalization of the classical Poisson equation:

> Let the "rest density" ν, as well as the gravitational potential Φ, be scalars in this space, and let them be linked through the differential equation:
>
> $$\frac{\partial^2\Phi}{\partial x^2}+\frac{\partial^2\Phi}{\partial y^2}+\frac{\partial^2\Phi}{\partial z^2}+\frac{\partial^2\Phi}{\partial u^2} = 4\pi\gamma\nu. \tag{2}$$
>
> (γ is the gravitational constant.)[8]

Somewhat later Abraham introduced further Minkowskian concepts and terminology:

> We write $\dot{x}\ \dot{y}\ \dot{z}\ \dot{u}$ for the first derivatives of the coordinates of a material "world point" with respect to its "proper time" τ, i.e., for the components of the "velocity" four-vector $\mathfrak{V}$, and $\ddot{x}\ \ddot{y}\ \ddot{z}\ \ddot{u}$ for the second derivatives, i.e., for the components of the "acceleration" four-vector $\dot{\mathfrak{V}}$.[9]

He then observes that, in Minkowski's formalism, the first derivatives must satisfy a certain relation involving the speed of light:

> Between the first derivatives the following identity holds:
>
> $$\dot{x}^2+\dot{y}^2+\dot{z}^2+\dot{u}^2 = -c^2, \tag{3}$$
>
> or
>
> $$\dot{l}^2\left\{\left(\frac{dx}{dl}\right)^2+\left(\frac{dy}{dl}\right)^2+\left(\frac{dz}{dl}\right)^2-1\right\} = -c^2 \ \ldots .^{10} \tag{4}$$

Finally Abraham derives a relation corresponding to the orthogonality relation between four-velocity and acceleration in ordinary Minkowski space but now under the assumption of a variable speed of light:

> Now, by differentiating eq. (4) [eq. (3)] with respect to proper time, Minkowski obtains the condition of "orthogonality" of the velocity and acceleration four-vectors. However, if c is considered to be variable, the place of that condition is taken by the following:
>
> $$\dot{x}\ddot{x}+\dot{y}\ddot{y}+z'\ddot{z}+\dot{u}\ddot{u} = -c\frac{dc}{d\tau}, \tag{5}$$
>
> as the differentiation of eq. (4) [eq. (3)] shows.[11]

---

8 "Die "Ruhdichte" ν sei ein Skalar in diesem Raume, und ebenso das Schwerkraftpotential Φ; sie mögen miteinander verknüpft sein durch die Differentialgleichung: …(1) [eq. (2)]. (γ ist die Gravitationskonstante.)" (Abraham 1912h, 1)

9 "Wir schreiben $\dot{x}\ \dot{y}\ \dot{z}\ \dot{u}$ für die ersten Ableitungen der Koordinaten eines materiellen "Weltpunktes" nach seiner "Eigenzeit" τ, d. h. für die Komponenten des Vierervektors "Geschwindigkeit" $\mathfrak{V}$, und $\ddot{x}\ \ddot{y}\ \ddot{z}\ \ddot{u}$ für die zweiten Ableitungen, d. h. für die Komponenten des Vierervektors "Beschleunigung $\dot{\mathfrak{V}}$." (Abraham 1912h, 1)

10 "Es besteht zwischen den ersten Ableitung die Identität: …(4) [eq. (3)] oder … [eq. (4)]." (Abraham 1912h, 1–2)

In this way, the introduction of the hypothesis of a variable speed of light makes it possible for Abraham to circumvent a problematic restriction imposed by Minkowski's formalism on a relativistic theory of gravitation.[12] In fact, in a theory with variable c, the four-vectors for velocity and for acceleration no longer have to be orthogonal to each other. It follows that the gravitational potential also no longer has to be constant along the world line of a particle, contrary to the conclusion reached in the usual Minkowski's formalism. Although Abraham's line of attack was clearly stimulated by Einstein's earlier use of a variable speed of light, it thus emerges as being so closely associated with a plausible modification of the four-dimensional formalism that this approach may also be conceived as exploring an independent possibility offered by the contemporary state of the gravitation problem.

In Einstein's papers of 1907 and 1911, the variable speed of light was linked to the gravitational potential via the concept of time in an accelerated frame of reference, i.e., via an essentially kinematic relation, independently of the use of the equivalence principle for transferring the results obtained in an accelerated system to a system with a gravitational field. For Abraham, on the other hand, an analogous relation between the speed of light and the gravitational potential follows if the second derivative terms in equation (5) are identified with the acceleration due to the gravitational force, i.e., from an essentially dynamic relation. Indeed, for the relation between four-dimensional gravitational force and four-dimensional gravitational potential Abraham assumed:

$$\begin{array}{ll} F_x = -\dfrac{\partial\Phi}{\partial x} & F_y = -\dfrac{\partial\Phi}{\partial y} \\ F_z = -\dfrac{\partial\Phi}{\partial z} & F_u = -\dfrac{\partial\Phi}{\partial u} \end{array} \tag{6}$$

He further assumed that the relation between four-acceleration and four-force is analogous to Newton's second law:

$$\begin{array}{ll} \ddot{x} = F_x & \ddot{y} = F_y \\ \ddot{z} = F_z & \ddot{u} = F_u \end{array} \tag{7}$$

With these two relations, equation (5) can now be written as a relation between the speed of light and the gravitational potential:

$$\dot{x}\frac{\partial\Phi}{\partial x} + \dot{y}\frac{\partial\Phi}{\partial y} + z'\frac{\partial\Phi}{\partial z} + \dot{u}\frac{\partial\Phi}{\partial u} = c\frac{dc}{d\tau} \tag{8}$$

---

11 "Nun erhält Minkowski, indem er Gl. (4) [eq. (3)] nach der Eigenzeit differenziert, die Bedingung der "Orthogonalität" der Vierervektoren Geschwindigkeit und Beschleunigung. Wenn jedoch c als veränderlich angesehen wird, tritt an Stelle jener Bedingung die folgende: … (5) [eq. 5] wie die Differentiation von Gl. (4) [eq. (3)] zeigt." (Abraham 1912h, 2)

12 See "Einstein, Nordström and the Early Demise of Scalar, Lorentz Covariant Theories of Gravitation," (in vol. 3 of this series).

which can be rewritten as:

$$\frac{d\Phi}{d\tau} = c\frac{dc}{d\tau}. \tag{9}$$

The last equation in turn can be integrated to yield:

$$\frac{c^2}{2} - \frac{c_0^2}{2} = \Phi - \Phi_0, \tag{10}$$

where c_0 and Φ_0 are the speed of light and the gravitational potential at the origin of coordinates. Abraham comments on this formula:

> *The increase of half the square of the speed of light is equal to the increase of the gravitational potential.*
>
> Instead of this relation, which is exactly valid according to our theory, one can, neglecting the square of the quotient of Φ and c^2, take Einstein's formula (loc. cit. p. 906):
>
> $$c = c_0\left(1 + \frac{\Phi - \Phi_0}{c^2}\right). \tag{11}$$
>
> However, eq. (6) [eq. (10)] better serves to manifest the independence from the arbitrarily chosen origin of coordinates.[13]

Abraham had thus achieved all essential elements of Einstein's research in the years 1907 to 1911, albeit in an entirely different way which, in addition, offered a wealth of mathematical resources for the further elaboration of a full-fledged theory of gravitation.

1.2 Abraham's Mathematics versus Einstein's Physics

While the exploitation of Minkowski's formalism had opened up new mathematical possibilities to Abraham, the physical interpretation of his results had yet to be explored. Whereas Einstein had been aware from the beginning that he was transgressing the limits of his original theory of relativity, Abraham seemed to have been initially convinced that he had found the relativistic theory of gravitation that was called for after the establishment of the principle of relativity, as is clear from his introductory remark quoted above. But since the constancy of the speed of light was one of the foundational elements of special relativity, it was questionable with which right Abraham could make use of relations derived from Minkowski's reformulation of special relativity for a theory in which the speed of light depends on the gravita-

---

13 "*Der Zuwachs des halben Quadrats der Lichtgeschwindigkeit ist gleich dem Zuwachs des Schwerkraftpotentials*. Anstelle dieser, nach unserer Theorie exakt gültigen Beziehung kann man, bei Vernachlässigung des Quadrats des Quotienten aus Φ und c^2, die Einsteinsche Formel setzten (loc. cit. S. 906): … [eq. 11]. Indessen läßt die Formel (6) besser die Unabhängigkeit von dem willkürlich wählbaren Koordinatenursprung hervortreten." [Reference is to (Einstein 1911).] (Abraham 1912h, 2)

tional potential. It seems that Einstein, in a personal communication now lost, immediately brought Abraham's attention to this conflict between the formalism of the latter's theory and one of its fundamental assumptions, the variability of the speed of light.[14]

Abraham's use of the straightforward special relativistic generalization of the Poisson equation (2), as well as of other elements of Minkowski's formalism, without really bothering about the gravity of energy, made Einstein skeptical about Abraham's claims. For Einstein, the gravity of energy and the variability of the speed of light were closely connected:

> It is a great pity that the gravitation theory is leading to so little which is observable. But it nevertheless must be taken seriously because the theory of relativity requires such a further development with urgency since the gravitation vector cannot be integrated into the relativity theory with constant c if one requires the *gravitational* mass of energy.[15]

In Einstein's opinion, a special relativistic framework was hence not the appropriate starting point for coping with the concept of a gravity of energy and for predicting effects such as the gravitational deflection of light, while Abraham, in fact, claimed that he could do so:

> But as far as the bending of light rays in the gravitational field is concerned, which can be derived from (6) [eq. (10)] with the help of Huygens' principle, it is identical with the bending of the trajectories of those light particles. This is one of the numerous incomplete analogies between the modern theory of radiation and the emission theory of light.[16]

Einstein, on the other hand, criticized Abraham's theory for not really explaining the bending of light, probably because it so closely resembled a special relativistic theory:

> I am having a controversy with Abraham because of his theory of gravitation. The latter does in reality not give an account of a bending of light rays.[17]

14 For evidence of this personal communication see the "Correction" ("Berichtigung") to (Abraham 1912h), quoted in note 20.

15 "Es ist sehr schade, dass die Gravitationstheorie zu so wenig Beobachtbarem führt. Aber sie muss trotzdem ernst genommen werden, weil die Relativitätstheorie eine derartige Weiterentwicklung gebieterisch verlangt, indem der Gravitationsvektor in die Rel. Theorie mit konstantem c sich nicht einfügen lässt, wenn man die *schwere* Masse der Energie fordert." Einstein to Wilhelm Wien, 17 May 1912, (CPAE 5, Doc. 395, 465).

16 "Was aber die Krümmung der Lichtstrahlen im Schwerkraftfelde anbelangt, die aus (6) [eq. (10)] mit Hilfe des Huygensschen Prinzips sich ableiten läßt, so ist sie identisch mit der Krümmung der Bahnkurve jener Lichtteilchen. Es ist dies eine der Zahlreichen unvollständigen Analogien zwischen der modernen Strahlungstheorie und der Emissionstheorie des Lichtes." (Abraham 1912h, 2) He referred to Einstein's 1911 paper and also undertook a comparison with light deflection in an emission theory of light.

17 "Ich habe eine Kontroverse mit Abraham wegen dessen Theorie der Gravitation. Dieselbe gibt in Wahrheit von einer Krümmung der Lichtstrahlen nicht Rechenschaft." Einstein to Wilhelm Wien, 27 January 1912, (CPAE 5, Doc. 343, 394).

In later comments Einstein acknowledged that light deflection actually follows from Abraham's theory and concentrated his criticism more generally on what he perceived as an incoherent use of the mathematical formalism of special relativity:

> The theory of Abraham, according to which light is also curved, just as it is in my case, is inconsequent from the point of view of the theory of invariants.[18]

What Einstein meant precisely becomes clearer from another contemporary comment:

> The matter is not, however, as simple as Abraham believes it to be. In particular the principle of the constant c and hence the equivalence of the four dimensions is lost.[19]

Einstein must also have addressed such criticism to Abraham directly. The latter, in any case, reacted to Einstein's arguments by pursuing his modification of Minkowski's formalism in greater depth. In a short note published on 15 February 1912 as a reply to Einstein's critique, Abraham revoked the lines with which he had earlier referred to Minkowski's formalism, instead introducing an infinitesimal line element with variable metric, thus effectively extending Minkowski's spacetime to a more general semi-Riemannian manifold:

> In lines 16, 17 of my note "On the Theory of Gravitation" an oversight has to be corrected which was brought to my attention by a friendly note from Mr. A. Einstein. One should read there: 'we consider dx, dy, dz and $du = idl = icdt$ as components of a displacement $\overrightarrow{ds}$ in four-dimensional space'.
>
> Hence
>
> $$ds^2 = dx^2 + dy^2 + dz^2 - c^2dt^2 \tag{12}$$
>
> is the square of the four-dimensional line element where the speed of light c is determined by equation (6) [eq. (10)].[20]

In this way, Abraham had effectively introduced the mathematical representation of the gravitational potential that was to be at the core of later general relativity, the general four-dimensional line element involving a variable metric tensor. However, for the time being, Abraham's expression remained an isolated mathematical formula without context and physical meaning which, at this point, was indeed neither provided by Abraham's nor by Einstein's physical understanding of gravitation. Abraham's expression in particular was neither related to insights about coordinate

---

18 "Die Theorie von Abraham, nach welcher das Licht ebenfalls ebenso wie bei mir gekrümmt ist, ist vom invariantentheoretischen Standpunkt inkonsequent." Einstein to Erwin Freundlich, mid-August 1913, (CPAE 5, Doc. 468, 550).

19 "So einfach, wie Abraham meint, ist die Angelegenheit aber nicht. Insbesondere geht das Prinzip des konstanten c und damit die Gleichwertigkeit der 4 Dimensionen verloren." Einstein to Wilhelm Wien, 11 March 1912, (CPAE 5, Doc. 371, 430).

20 "Auf Z. 16, 17 meiner Note "Zur Theorie der Gravitation" ist ein Versehen zu berichtigen, auf welches ich durch eine freundliche Mitteilung des Herrn A. Einstein aufmerksam geworden bin. Man lese daselbst: 'betrachten wir dx, dy, dz und $du = idl = icdt$ als Komponenten einer Verschiebung $\overrightarrow{ds}$ im vierdimensionalen Raume'. Es ist also $ds^2 = dx^2 + dy^2 + dz^2 - c^2dt^2$ das Quadrat des vierdimensionalen Linienelementes, wobei die Lichtgeschwindigkeit c durch G. (6) bestimmt ist." ("Berichtigung," Abraham 1912h, 176)

systems nor to any ideas about a generalized inertial motion, as is the case for the line element in general relativity. It therefore comes as no surprise that Abraham's formulation did not constitute any noteworthy "discovery" and that it was not even taken very seriously by either Abraham or Einstein.[21]

Einstein did acknowledge the impressive formal advance owed to Abraham's bold "mathematical" approach, in particular when compared with his own sluggish progress; but he quickly realized the theory's impoverished physical meaning. In a letter written shortly after the publication of Abraham's first paper on 27 January 1912, he remarked:

> Abraham has supplemented my gravitation thing, making it into a closed theory, but he made considerable errors in reasoning so that the thing is probably incorrect. This is what happens when one operates formally, without thinking physically![22]

However, he must have been initially impressed by the elegance with which Abraham's formalism yielded essentially the same results as his own, mathematically more pedestrian efforts, and also more than what he himself had achieved. This is evident from a comment Einstein made about two months later, when he was already convinced that Abraham's theory was not tenable:

> Abraham's theory has been created out of thin air, i.e., out of nothing but considerations of mathematical beauty, and is completely untenable. I find it hard to understand how this intelligent man allowed himself to get carried away with such superficiality. *At first (for 14 days!) I too was completely "bluffed" by the beauty and simplicity of his formulas*. [my emphasis][23]

At the beginning of February, Einstein was apparently still willing to concede to Abraham the benefit of the doubt. He wrote:

> Abraham has further developed the new gravitation theory; we are corresponding about this since we are not completely of the same opinion.[24]

As early as mid-February, however, Einstein had formed his firm, negative judgement on Abraham's theory:

> Abraham's theory is completely untenable.[25]

---

21 For a critical view on the notion of discovery, see (Renn et al. 2001).

22 "Abraham hat meine Gravitationssache zu einer geschlossenen Theorie ergänzt, aber bedenkliche Denkfehler dabei gemacht, sodass die Sache wohl unrichtig ist. Das kommt davon, wenn man formal operiert, ohne dabei physikalisch zu denken!" Einstein to Heinrich Zangger, 27 January 1912, (CPAE 5, Doc. 344, 395).

23 "Abrahams Theorie ist aus dem hohlen Bauche, d. h. aus blossen mathematischen Schönheitserwägungen geschöpft und vollständig unhaltbar. Ich kann gar nicht begreifen, wie sich der intelligente Mann zu solcher Oberflächlichkeit hat hinreissen lassen können. Im ersten Augenblick (14 Tage lang!) war ich allerdings auch ganz "geblüfft" durch die Schönheit und Einfachheit seiner Formeln." Einstein to Michele Besso, 26 March 1912, (CPAE 5, Doc. 377, 436–437).

24 "Abraham hat die neue Gravitationstheorie weiter ausgeführt; wir korrespondieren darüber, weil wir nicht vollkommen gleicher Meinung sind." Einstein to Michele Besso, 4 February 1912, (CPAE 5, Doc. 354, 406).

Einstein's harsh judgement may have also been related to the progress he himself was meanwhile making on a theory of gravitation based on the equivalence principle. In fact, the first reference to this theory, later published in Einstein's papers on the static gravitational field, is found in a letter from mid-February:

> The second thing concerns the relationship: gravitational field—acceleration field—velocity of light. Simple and beautiful things emerge here quite automatically. The velocity of light c is variable. It determines the gravitational force. A stationary point with mass 1 is acted upon by the force
>
> $$-\frac{\partial c}{\partial x} - \frac{\partial c}{\partial y} - \frac{\partial c}{\partial z}.$$
>
> In empty space c satisfies Laplace's equation. The inertial mass of a body is m/c, that is, it decreases with the gravitational potential. The equations of motion for the material point agree essentially with those of the customary theory of relativity. Abraham's theory is unfounded in every respect if there really is an equivalence between the gravitational field and the "acceleration field."[26]

It thus seems that around the end of January, Einstein, impressed by "the beauty and simplicity" of Abraham's formulas, had taken up work on a gravitation theory of his own and that he had then, in mid February, after an exchange of letters with Abraham, come to the conclusion that the latter's theory must be untenable, not least because its essential achievements could also be obtained on a physically much more sound basis from the equivalence principle. The relation between Einstein's own work on a gravitation theory and his negative judgement of Abraham's theory is also confirmed by another contemporary comment:

> In the course of my research on gravitation I discovered that Abraham's theory (1st issue of *Phys. Zeitschr.*) is completely untenable.[27]

After having initially limited himself to private communications, Einstein then prepared himself for a public controversy with Abraham:

---

25 "Abrahams Theorie der Gravitation ist ganz unhaltbar." Einstein to Paul Ehrenfest, 12 February 1912, (CPAE 5, Doc. 357, 408).

26 "Die zweite Sache betrifft die Beziehung Gravitationsfeld—Beschleunigungsfeld—Lichtgeschwindigkeit. Es kommen da einfache und schöne Dinge ganz zwangläufig heraus. Die Lichtgeschwindigkeit c ist variabel. Sie bestimmt die Gravitationskraft. Auf einen ruhenden Punkt von der Masse 1 wirkt die Kraft

$$-\frac{\partial c}{\partial x} - \frac{\partial c}{\partial y} - \frac{\partial c}{\partial z}.$$

c erfüllt im leeren Raume die Laplace'sche Gleichung. Die träge Masse eines Körpers ist m/c, sinkt also mit dem Schwerepotential. Die Bewegungsgleichungen des materiellen Punktes stimmen mit denen der gewöhnlichen Relativitätstheorie im Wesentlichen überein. Die Theorie Abrahams ist in allen Teilen unzutreffend, wenn die Aequivalenz zwischen Schwerefeld und "Beschleunigungsfeld" wirklich besteht." See Einstein to Hendrik A. Lorentz, 18 February 1912, (CPAE 5, Doc. 360, 413).

27 "Bei meiner Untersuchung über Gravitation entdeckte ich, dass Abrahams Theorie (1. Heft der phys. Zeitschr.) ganz unhaltbar ist." Einstein to Heinrich Zangger, before 29 February 1912, (CPAE 5, Doc. 366, 421).

> Abraham's theory is completely wrong. I will probably get into a heavy ink fight with him.[28]

As no letters by Abraham on this issue have survived, only the controversy, published later, allows one to infer how he defended himself against Einstein's criticism in his letters. This issue is of importance since Einstein's published criticism of February 1912 addresses a point that is not explicitly made in Abraham's early papers, the admissibility of Lorentz transformations in the infinitesimally small. But Abraham's later papers do suggest that he viewed the correction published in February 1912 as showing that the Lorentz transformation can be at least locally upheld. In the introductory section of a paper published in September 1912 he wrote for example:

> I have just availed myself of the language of the theory of relativity. But it will become clear that this theory cannot be brought into agreement with the views on the force of gravity presented here, in particular because the axiom of the constancy of the speed of light is relinquished. In my earlier papers on gravitation I have attempted to preserve at least in the infinitesimally small the invariance with respect to the Lorentz transformations.[29]

Similarly Abraham wrote in June 1912:

> I had given to the expressions of the gravitation tensor as well as to the equations of motion of the material point in the gravitational field a form which in the infinitesimally small is invariant with respect to Lorentz transformations.[30]

Indeed, in the series of four papers (plus one correction) Abraham published between January and March 1912, starting from his basic paper "On the Theory of Gravitation," via "The Elementary Law of Gravitation," the "Correction," and "The Free Fall," up to "The Conservation of Energy and Matter in the Gravitational Field"[31] Abraham made free use of infinitesimal Lorentz transformations, and even of their integration,[32] without ever explicitly justifying that this procedure is legitimate in the case of a variable speed of light.

In February 1912, Einstein published his criticism of Abraham in the context of his own first paper on a field theory of gravitation. To the editor of the *Annalen der Physik* he wrote:

---

28 "Abrahams Theorie ist ganz falsch. Ich werde wohl ein schweres Tintenduell mit ihm bekommen." Einstein to Ludwig Hopf, after 20 February 1912, (CPAE 5, Doc. 364, 418).

29 " Ich habe mich soeben der Sprache der Relativität bedient. Doch wird sich zeigen, daß diese Theorie mit den hier vorgetragenen Ansichten über die Schwerkraft nicht zu vereinbaren ist, schon darum nicht weil das Axiom von der Konstanz der Lichtgeschwindigkeit aufgegeben wird. Ich habe in meinen früheren Arbeiten über die Gravitation versucht, wenigstens in unendlich kleinen die Invarianz gegenüber den Lorentz-Transformationen zu bewahren." (Abraham1912b, 793–794)

30 "Ich hatte den Ausdrücken des Gravitationstensor, sowie den Bewegungsgleichungen des materiellen Punktes im Schwerefelde eine Form gegeben, die im unendlich kleinen gegenüber Lorentztransformationen invariant ist." (Abraham 1912f, 1057)

31 See (Abraham 1912h, 1912a, 1912c, 1912d) respectively.

32 See (Abraham 1912c, 310).

> I hereby send you a paper for the *Annalen*. Many a drop of sweat is hanging on it, but I now have complete confidence in the matter. Abraham's theory is completely unacceptable. How could anybody have the luck to guess effortlessly equations which are correct! Now I am looking for the dynamics of gravitation. But this will not happen so quickly![33]

In his paper, the aim of criticizing Abraham's theory is immediately announced in the introduction:

> Since then, Abraham has constructed a theory of gravitation which contains the consequences drawn in my first paper as special cases. But we will see in the following that Abraham's system of equations cannot be brought into agreement with the equivalence hypothesis, and that his conception of time and space cannot be maintained even from the purely mathematically formal point of view.[34]

Einstein's argument against Abraham's theory is given in §4 of his paper "General Remarks on Space and Time":

> Which is now the relation of the above theory to the old theory of relativity (i.e. to the theory of the universal c)? According to Abraham's opinion the transformation equations of Lorentz are still valid in the infinitesimally small, that is, there shall be an $x-t$ transformation so that:

$$\begin{aligned} dx' &= \frac{dx - v\,dt}{\sqrt{1-\frac{v^2}{c^2}}}, \\ dt' &= \frac{-\frac{v}{c^2}dx + dt}{\sqrt{1-\frac{v^2}{c^2}}} \end{aligned} \tag{13}$$

> are valid. dx' and dt' must be complete differentials. Therefore the following equations must be valid:

$$\begin{aligned} \frac{\partial}{\partial t}\left\{\frac{1}{\sqrt{1-\frac{v^2}{c^2}}}\right\} &= \frac{\partial}{\partial x}\left\{\frac{-v}{\sqrt{1-\frac{v^2}{c^2}}}\right\}, \\ \frac{\partial}{\partial t}\left\{\frac{-\frac{v}{c^2}}{\sqrt{1-\frac{v^2}{c^2}}}\right\} &= \frac{\partial}{\partial x}\left\{\frac{1}{\sqrt{1-\frac{v^2}{c^2}}}\right\}. \end{aligned} \tag{14}$$

---

33 "Ich sende Ihnen hier eine Arbeit für die Annalen. Hängt mancher Schweisstropfen daran, aber ich habe jetzt alles Vertrauen zu der Sache. Abrahams Theorie der Gravitation ist ganz unannehmbar. Wie könnte einer auch das Glück haben Gleichungen mühelos zu erraten, die richtig sind! Nun suche ich nach der Dynamik der Gravitation. Es wird aber nicht schnell damit gehen!" Einstein to Wilhelm Wien, 24 February 1912, (CPAE 5, Doc. 365, 420).

34 "Seitdem hat Abraham eine Theorie der Gravitation aufgestellt, welche die in meiner ersten Arbeit gezogenen Folgerungen als Spezialfälle enthält. Wir werden aber im folgenden sehen, daß sich das Gleichungssystem Abrahams mit der Äquivalenzhypothese nicht in Einklang bringen läßt, und daß dessen Auffassung von Zeit und Raum sich schon vom rein mathematisch formalen Standpunkte aus nicht aufrecht erhalten läßt." (Einstein 1912, 355)

> In the unprimed system the gravitational field shall now be static. Then c is an arbitrarily given function of x, but independent of t. If the primed system shall be a 'uniformly' moved one, then for fixed x, v must be in any case independent of t. The left-hand sides of the equations, and hence also the right-hand sides must therefore vanish. But the latter is impossible, since for c given in terms of arbitrary functions of x, both right-hand sides cannot be made to vanish by appropriately choosing v in dependence on x. Therefore it is shown that even for infinitesimally small regions of space and time one cannot adhere to the Lorentz transformation as soon as one gives up the universal constancy of c.[35]

In modern terminology, Einstein's argument amounts to showing that it is generally impossible to transform a coordinate system in which the metric does not have the Minkowskian form into another, well-defined one by means of a Lorentz transformation. In a letter somewhat later Einstein summarized his negative attitude toward Abraham's treatment of space and time as follows:

> I have now finished my studies on the statics of gravitation and have great confidence in the results. But the generalization appears to be very difficult. My results are not in agreement with those of Abraham. The latter has worked here, contrary to his usual style, rather superficially. Already his treatment of space and time is untenable.[36]

In spite of Einstein's criticism, Abraham nevertheless continued to use Minkowski's framework for elaborating consequences of his theory of gravitation, some of which will be discussed below. Einstein, in turn, became gradually convinced that it was worthwhile after all to take a closer look at the utility of a modification of this formalism for his version of a gravitational field theory as well. Driven by Abraham's bold and occasionally stubborn persistence, Einstein in May 1912 thus finally recognized that a generalized line element, as suggested by Abraham's note of three months earlier, indeed represents the key to a generally relativistic gravitation theory.[37]

---

35 "In was für einem Verhältnis steht nun die vorstehende Theorie zu der alten Relativitätstheorie (d. H. zu der Theorie des universellen c)? Nach Abrahams Meinung sollen die Transformationsgleichungen von Lorentz nach wie vor im unendlich Kleinen gelten, d. h. es soll eine $x-t$–Transformation geben, so daß [eq. (13)] gelten. dx' und dt' müssen vollständige Differentiale sein. Es sollen also die Gleichungen gelten [eq. (14)]. Es sei nun im ungestrichenen System das Gravitationsfeld ein statisches. Dann ist c eine beliebig gegebene Funktion von x, von t aber unabhängig. Soll das gestrichene System ein "gleichförmig" bewegtes sein, so muß v bei festgehaltenem x jedenfalls von t unabhängig sein. Es müssen daher die linken Seiten der Gleichungen, somit auch die rechten Seiten verschwinden. Letzteres ist aber unmöglich, da bei beliebig in Funktionen von x gegebenem c nicht beide rechten Seiten zum Verschwinden gebracht werden können, indem man v in Funktion von x passend wählt. Damit ist also erwiesen, daß man auch für unendlich kleine Raum-Zeitgebiete nicht an der Lorentztransformation festhalten kann, sobald man die universelle Konstanz von c aufgibt." (Einstein 1912, 368)

36 "Die Untersuchungen über die Statik der Gravitation habe ich nun fertig und setze grosses Vertrauen in die Resultate. Aber die Verallgemeinerung scheint sehr schwierig zu sein. Meine Ergebnisse sind mit denen von Abraham nicht im Einklang. Dieser hat gegen seine sonstige Gewohnheit hier recht oberflächlich gearbeitet. Schon seine Behandlung von Raum und Zeit ist unhaltbar." Einstein to Ludwig Hopf, 12 June 1912, (CPAE 5, Doc. 408, 483).

37 See "Classical Physics in Disarray …" and "The First Two Acts" (both in vol. 1 of this series).

1.3 Abraham's Rejection of Relativity

As time went by, Abraham's reaction to the controversy with Einstein became more and more acrid, until eventually he rejected the theory of relativity altogether. This is in stark contrast to Einstein's ever more resolved search for a generalization of the relativity principle as familiar from classical physics. In fact, Einstein had concluded his criticism of Abraham with the following comment:

> To me the spacetime problem seems to lie as follows. If one limits oneself to a region of constant gravitational potential, the natural laws take on an outstandingly simple and invariant form if one refers them to a spacetime system of that manifold which are connected to each other by the Lorentz transformations with constant c. If one does not restrict oneself to regions of constant c, then the manifold of equivalent systems, just as the manifold of the transformations that leave the natural laws unchanged, will become a larger one, but the laws will, on the other hand, become more complicated.[38]

In his published response to Einstein's critique, Abraham, on the other hand, wrote (the first sentence is quoted above):

> I had given to the expressions of the gravitation tensor as well as to the equations of motion of the material point in the gravitational field a form which in the infinitesimally small is invariant with respect to Lorentz transformations. In the restriction to the infinitesimally small it is already implicit that this invariance shall not be maintained in the finite. In fact, if the gravitational field influences the speed of light, then it is clear from the outset that there is an essential difference between a reference system $\Sigma(xyzt)$ in which the gravitational field is a static one and a reference frame $\Sigma(x'y'z't')$ which is uniformly moved with respect to the former, in which the gravitational field, and hence also the speed of light, is changing with time. There can be no talk about any kind of relativity, i.e., about a correspondence between the two systems, which would express itself in equations between their spacetime parameters $xyzt$ and $x'y'z't'$. Indeed the differential equations between dx', dt' and dx, dt, which contain the Lorentz-transformation in the infinitesimally small, are, as Mr. Einstein observes, not integrable.[39]

Abraham thus drew from the same mathematical fact a consequence that is diametrically opposed to that drawn by Einstein. Instead of calling for an extension of the relativity principle, as Einstein did, Abraham called for a complete rejection of this principle. He therefore continued:

> But in this inherently correct observation I cannot find any justification for the claim that "my conception of time and space cannot be maintained even from the purely mathematically formal point of view." However, any relativistic spacetime conception which would find its expression in relations between spacetime parameters of Σ and Σ' becomes untenable. Such a relativistic spacetime conception is, on the other hand,

---

38 "Mir scheint das Raum-Zeitproblem wie folgt zu liegen. Beschränkt man sich auf ein Gebiet von konstantem Gravitationspotential, so werden die Naturgesetze von ausgezeichnet einfacher und invarianter Form, wenn man sie auf ein Raum-Zeitsystem derjenigen Mannigfaltigkeit bezieht, welche durch die Lorentztransformationen mit konstantem c miteinander verknüpft sind. Beschränkt man sich nicht auf Gebiete von konstantem c, so wird die Mannigfaltigkeit der äquivalenten Systeme, sowie die Mannigfaltigkeit der die Naturgesetze ungeändert lassenden Transformationen eine größere werden, aber es werden dafür die Gesetze komplizierter werden." (Einstein 1912, 368–369)

> entirely far fetched to me. As I have already mentioned elsewhere [Abraham 1912g], to me, the interpretation in the sense of an absolute theory seems rather the appropriate one. If among all reference frames that one is privileged in which the gravitational field is static, or quasi static, then it is permissible to refer to a motion related to this system as 'absolute'.[40]

He proceeds by explaining the relation of his conception to traditional physical ideas about preferred systems of reference, such as that identified by Neumann's "body α," and concludes with the remark:

> Who wishes to do so, may interpret this conception as an argument for the 'existence of the aether'.[41]

Abraham became ever more skeptical about Einstein's attempts to extend the principle of relativity, as his later publications show. Although he actively contributed to developing tools and concepts of a relativistic theory of gravitation, such as expressions for stresses and energy in a gravitational field, he never dealt in detail with the revisions of the concept of time implied by his own theory, let alone those of the concepts of space and time associated with Einstein's rival theories.[42] In a paper submitted in July 1912, Abraham pointed to Einstein's difficulties with implementing the equivalence hypothesis in his own theory of gravitation, but merely restricted himself to the following brief remark on the subject:

> Here I would therefore prefer to develop the new gravitation theory without entering the spacetime problem.[43]

---

39 "Ich hatte den Ausdrücken des Gravitationstensors, sowie den Bewegungsgleichungen des materiellen Punktes im Schwerefelde eine Form gegeben, die im unendlich kleinen gegenüber Lorentztransformationen invariant ist. In der Beschränkung auf das unendlich kleine liegt schon implicite enthalten, daß im endlichen diese Invarianz nicht bestehen soll. In der Tat, wenn das Gravitationsfeld die Lichtgeschwindigkeit beeinflußt, so ist es von vornherein klar, daß ein wesentlicher Unterschied zwischen einem Bezugssystem $\Sigma(xyzt)$ besteht, in welchem das Schwerkraftfeld ein statisches ist, und einem gegen dieses gleichförmig bewegten Bezugssystem $\Sigma(x'y'z't')$, in welchem das Schwerkraftfeld, und mithin auch die Lichtgeschwindigkeit, sich zeitlich verändert. Es kann von irgend einer Art von Relativität, d. h. von einer Korrespondenz der beiden Systeme, die sich in Gleichungen zwischen ihren Raum-Zeit-Parametern $xyzt$ uns $x'y'z't'$ ausdrücken würde, keine Rede sein. In der Tat sind, wie Hr. Einstein bemerkt, die Differenzialgleichungen zwischen dx', dt' und dx, dt, welche die Lorentztransformation im unendlich kleinen enthalten, nicht integrabel." (Abraham 1912f, 1057)

40 "In dieser an sich zutreffenden Bemerkung kann ich freilich keine Rechtfertigung für die Behauptung finden, daß "meine Auffassung von Zeit und Raum sich schon vom rein mathematisch formalen Standpunkt aus nicht aufrecht erhalten läßt". Unhaltbar wird allerdings jede relativistische Raum-Zeit-Auffassung, die in Beziehungen zwischen den Raum-Zeit-Parametern von Σ und Σ' ihren Ausdruck finden würde. Eine solche relativistische Raum-Zeit-Auffassung liegt mir indessen ganz fern. Mir scheint vielmehr, wie ich bereits an anderem Orte erwähnt habe [Abraham 1912g], die Deutung im Sinne einer Absoluttheorie die passende zu sein. Wenn unter allen Bezugssystemen dasjenige ausgezeichnet ist, in welchem das Schwerefeld statisch, oder quasi-statisch ist, so ist es erlaubt, eine auf dieses System bezogene Bewegung "absolut" zu nennen." (Abraham 1912f, 1057)

41 "Wer will, mag diese Vorstellung als Argument für die 'Existenz des Äthers' deuten. (Abraham 1912f, 1058)

Abraham did not just ignore the consequences of the new gravitation theories for the concepts of space and time, he actually rejected them, with no lesser consequence than that with which Einstein pursued them. In short, the same challenging task of creating a gravitation theory compatible with relativity theory found complementary responses in Abraham and Einstein. While both not only acknowledged the difficulties of making them compatible, but also returned to a revision of special relativity, Einstein did so in order to complete the relativity revolution, and Abraham to undo it. Commenting on Einstein's later attempts at a relativistic theory of gravitation, Abraham wrote in a popular review of 1914:

> Hence, at the cliff of gravitation every theory of relativity fails, the special one of 1905, as well as the general one of 1913. The relativistic ideas are obviously not sufficiently advanced to serve as a framework for a complete worldview.
>
> But historical merit does remain for the theory of relativity with regard to its critique of the concepts of space and time. It has taught us that these concepts depend on the ideas we form concerning the behavior of the measurement rods and clocks that we use for the measurement of lengths and intervals of time, and which are subject to change with them [the measurement instruments]. This will secure an honorable funeral for the theory of relativity.[44]

Although Abraham did not accept the new concepts of space and time introduced by Einstein, he remained an acute critic of the latter's ongoing search for a relativistic

---

42 In March 1911 Abraham submitted a paper (Abraham 1912c) in which he used the distinction between proper time and coordinate time in order to draw far-reaching cosmological consequences from his theory. In this paper Abraham explained his definition of time coordinates:
"τ denotes the 'proper time' of the moved point, which is related to the time t measured in the reference system as follows:

$$\frac{du}{d\tau} = ic\frac{dt}{d\tau} = \frac{ic}{\kappa}$$

$$\kappa = \sqrt{1-\frac{v^2}{c^2}}$$

(v magnitude of velocity of the material point).
Clearly, Abraham's explanation of the relation between proper time and coordinate time, although formally analogous to Minkowski's formalism, can neither be based on this formalism, as the speed of light is assumed to be variable, nor be brought into harmony with Einstein's physical motivation for these two notions of time. Contrary to Einstein it appears from the remainder of this paper that Abraham considered only the coordinate time t to have any physical meaning.

43 "Ich möchte es daher hier vorziehen, die neue Gravitationstheorie zu entwickeln, ohne auf das Raum-Zeit-Problem einzugehen." (Abraham 1912b, 794)

44 "An der Klippe der Schwerkraft scheitert also jede Relativitätstheorie, sowohl die spezielle von 1905, wie die allgemeine von 1913. Die relativistischen Ideen sind offenbar nicht weit genug, um einem vollständigen Weltbilde als Rahmen zu dienen. Doch bleibt der Relativitätstheorie ein historisches Verdienst um die Kritik der Begriffe von Raum und Zeit. Sie hat uns gehehrt, dass diese Begriffe von den Vorstellungen abhängen, die wir uns von dem Verhalten der zur Messung von Längen und Zeitintervallen dienenden Massstäbe und Uhren bilden, und die mit ihnen dem Wandel unterworfen sind. Dies sichert der Relativitätstheorie ein ehrenvolles Begräbnis." (Abraham 1914, 26)

theory of gravitation, which, apart from Abraham's reaction, found only little resonance in the contemporary physics community. In a letter to his friend Michele Besso, Einstein wrote:

> Physicists have such a passive attitude toward my work on gravitation. Abraham is still the one who shows the most comprehension. It is true that he complaints violently in 'Scientia' against anything to do with relativity, but with understanding.[45]

In spite, and occasionally perhaps because of his intellectual distance, Abraham noticed problematic features of Einstein's attempts, which eventually became important issues in the development of general relativity, such as the question of the sense in which Einstein's mathematical requirement of a generalized covariance actually also realizes a generalization of the relativity principle as he claimed. In a review paper of 1915, Abraham expressed his doubts concerning Einstein's claim that his *Entwurf* theory of 1913, which is covariant with regard to general linear transformations, actually represents a generalization of the relativity principle of the special theory of 1905:

> The significance of this transformation group lies in the fact that it contains the Lorentz transformations; in the earlier theory of relativity the covariance with respect to this group gave expression to the equivalence of systems of reference in translatory motion with respect to each other. Is this presently also the case in the "generalized theory of relativity"? Does the covariance of the field equations with respect to linear orthogonal transformations imply that in a finite system of mutually gravitating bodies the course of the relative motions is not altered by a uniform translation of the entire system? That this is so has so far not been proven.[46]

Abraham thus pinpoints a distinction that is today considered as being crucial for understanding the covariance properties of a gravitational field equation, the distinction between general covariance as a property of the mathematical formulation of such a field equation, and the symmetry group under which the theory remains invariant, as is the case for the Lorentz group of special relativity. He thus anticipates a conceptual clarification that is usually attributed to Kretschmann.[47] But in spite of this significant insight, Abraham's further explanation of his skeptical attitude shows signs of a still immature understanding of the kinematics within a generic four-

45 "Zur Gravitationsarbeit verhält sich die physikalische Menschheit ziemlich passiv. Das meiste Verständnis hat wohl Abraham dafür. Er schimpft zwar in der "Scienza" kräftig über alle Relativität, aber mit Verstand." Einstein to Michele Besso after 1 January 1914, (CPAE 4, Doc. 499). See (Cattani and De Maria 1989, 171).

46 "Die Bedeutung dieser Transformationsgruppe beruht darauf, daß sie die Lorentzschen Transformationen enthält; die Kovarianz ihnen gegenüber brachte in der früheren Relativitätstheorie die Gleichberechtigung translatorisch gegeneinander bewegter Bezugssysteme zum Ausdruck. Ist dies nun auch in der "verallgemeinerten Relativitätstheorie" der Fall? Bedingt es die Kovarianz der Feldgleichungen gegenüber linearen orthogonalen Transformationen, daß in einem endlichen Systeme gegeneinander gravitierender Körper der Ablauf der relativen Bewegungen durch eine gleichförmige Translation des ganzen Systems nicht geändert wird? Daß dem so sei, ist bisher nicht bewiesen worden." (Abraham 1915, 515)

47 See (Kretschmann 1917) and, for historical discussion, (Norton 1992, 1993; Rynasiewicz 1999).

dimensional spacetime manifold, in particular, concerning his lack of understanding of the relation between different coordinate representations of one and the same physical magnitude:

> Such a proof may already be impossible to conduct for the reason that the concept of "uniform motion" of a finite system is completely up in the air in the new theory of relativity. Since the "natural" space and time measurement is influenced by the values of the space and time potential, observers at different locations in the gravitational field will ascribe different velocities to the same material point. Only for an infinitesimally small region of four-dimensional space—i.e. for one in which the potentials $g_{\mu\nu}$ can be considered a constant—is "velocity" defined at all.[48]

Abraham then returns to the problem of interpreting the covariance property of the field equations in terms of a relativity principle:

> Presumably, only within such an infinitesimal region may the covariance of the gravitational equations with respect to linear orthogonal transformations be interpreted in the sense of an equivalence of systems of reference moving with respect to each other. But if relativity of motion no longer exists for finite systems of gravitating masses in Einstein's theory, with what right does he then assign such great importance to the formal connection to the earlier theory of relativity?[49]

Abraham's criticism was evidently colored by his growing hostility toward any theory of relativity, and motivated, without doubt, by his own failure to successfully establish a relativistic theory of gravitation. But because it raised such crucial issues as the physical significance of the unfamiliar mathematical objects introduced in the course of Einstein's search for such a theory, it nevertheless represents an important intellectual context of the emergence of general relativity. Abraham's sometimes ardent polemics reveal not only the challenging difficulties with which Einstein confronted his contemporaries, but also contributed, at the time, to anchoring his high-flying mathematical artifices in the shared knowledge of contemporary physics, a process made particularly ungainly by the hesitant and cool reaction to Einstein's labors of other contemporary scientists.

---

48 "Ein solcher Beweis dürfte auch schon aus dem Grunde sich nicht führen lassen, weil der Begriff der "gleichförmigen Bewegung" eines endlichen Systems in der neuen Relativitätstheorie völlig in der Luft schwebt. Denn da die "natürliche" Raum- und Zeitmessung durch die lokalen Potentialwerte beeinflußt wird, so werden Beobachter an verschiedenen Orten im Schwerefelde demselben materiellen Punkte verschiedene Geschwindigkeiten zuschreiben. Nur für ein unendlich kleines Gebiet des vierdimensionalen Raumes—d. h. für ein solches, in welchem die Potentiale $g_{\mu\nu}$ als konstant gelten könne—ist die "Geschwindigkeit überhaupt definiert." (Abraham 1915, 515)

49 "Vermutlich dürfte nur innerhalb eines solchen infinitesimalen Gebietes die Kovarianz der Gravitationsgleichungen gegenüber orthogonalen linearen Transformationen im Sinne einer Gleichberechtigung gegeneinander bewegter Bezugssysteme zu deuten sein. Wenn aber in Einsteins Theorie für endliche Systeme gravitierender Massen keine Bewegungsrelativität mehr besteht, mit welchem Rechte legt er dann dem formalen Anschluß an die frühere Relativitätstheorie eine so große Wichtigkeit bei?" (Abraham 1915, 515–516)

2. VISTAS FROM JUST BELOW THE SUMMIT

2.1 A New Theory of Gravitation Meets the Shared Knowledge

Abraham's insights into the nature and effects of the gravitational field conceived in a relativistic context are today even less well known than his theoretical approach. It may seem that they have been rightly forgotten because of their apparent lack of impact on the further development of the theory. It was, however, precisely Abraham's untiring efforts to elaborate his theoretical approach and to draw physical consequences from it that turned his theory into a valuable touchstone and standard of comparison for other contemporary attempts at a relativistic theory of gravitation. Some of the insights that Abraham achieved in the course of his research, such as the possibility and essential properties of gravitational waves remain to this day a standard for a relativistic theory of gravitation.

In the first paper on his theory, Abraham did not go into much detail about the physical consequences of his new theory of gravitation. But he did, of course, have to confront the shared physical knowledge of his time. This ranged from the knowledge embodied in Newton's theory of gravitation—including the deviations from it—via energy and momentum conservation to some of the unusual insights attained by Einstein on the basis of the equivalence principle, such as the deflection of light, which possibly had to be incorporated in this new theory of gravitation as well.

In his paper, Abraham first discusses the equations of motion in a gravitational field which he formulates, as we have seen above, in analogy to Newtonian dynamics. From equations (6) and (7) he obtains:

$$m\frac{d\dot{x}}{d\tau} = -m\frac{\partial\Phi}{\partial x}, \qquad m\frac{d\dot{y}}{d\tau} = -m\frac{\partial\Phi}{\partial y}, \qquad m\frac{d\dot{z}}{d\tau} = -m\frac{\partial\Phi}{\partial z}. \tag{15}$$

These three equations he interpreted as representing momentum conservation. He then added:

> In contrast, the last of the equations (3) [i.e. eq. (7)]:
>
> $$m\frac{dl}{d\tau} = -mi\mathfrak{F}_u = mi\frac{\partial\Phi}{\partial u} = m\frac{\partial\Phi}{\partial l} \tag{16}$$
>
> expresses the law of conservation of *vis viva* in Minkowskian mechanics.[50]

Abraham had thus dealt with one of the inescapable requirements imposed on a new theory by the accumulated knowledge of classical and special relativistic physics. He also went into some detail about the energy balance of physical processes in a gravitational field.

Concerning the observable consequences of his theory, he did not even mention the gravitational redshift predicted by Einstein, probably because at that point he

---

50 See (Abraham 1912h, 2).

could not or did not want to follow Einstein's introduction of two notions of time in a gravitational field.[51] But he did take up Einstein's prediction of a gravitational light deflection, which he compared to that predicted by the emission theory of light. Abraham had, after all, developed his theory of gravitation in reaction to Einstein's 1911 paper and, in particular, to the introduction of a variable speed of light which this paper emphasized. Commenting on equations (10) and (11) he wrote (as partly quoted above):

> It is instructive to compare the relation obtained with the emission theory of light. Let us imagine that the particles emitted from the light source move according to the laws of Galilean mechanics and that they are subject to the gravitational force. They would then experience an increase of *vis viva* (kinetic energy) which is equal to the *decrease* of potential energy. The increase of *vis viva* (kinetic energy) of the light particles calculated according to (6) [i.e. eq. (10)] is, by contrast, equal to the *increase* of their potential energy, i.e., of the same amount but of opposite sign as that following from the emission theory. But as far as the bending of light rays in the gravitational field is concerned, which can be derived from (6) [i.e. eq. (10)] with the aid of Huygens' principle,[1)] it is identical with the bending of the trajectory of those light particles. This is one of the numerous incomplete analogies between the modern theory of radiation and the emission theory of light.[52]

To this passage, after "with the aid of Huygens' principle," Abraham appended the following footnote in which he refers to Einstein:

> A. Einstein has shown that a light ray passing the surface of the Sun is deflected towards the Sun and has drawn the attention of the astronomers to this consequence of the theory which can perhaps serve as its verification.[53]

Apart from this footnote Abraham did not discuss specific observational consequences of his theory. This, however, does not imply that he was not interested in them. On the contrary, at about the same time as Einstein, he independently and actively pursued the astronomical consequences of a relativistic theory of gravitation as is evident from a letter he wrote to Schwarzschild.[54]

---

51 Gravitational redshift is mentioned for the first time with reference to Einstein, in (Abraham 1913, 197).

52 "Es ist lehrreich, die erhaltene Beziehung der Emissionstheorie des Lichts gegenüberzustellen. Wir wollen uns vorstellen, daß die von der Lichtquelle emittierten Teilchen sich gemäß den Gesetzen der Galileischen Mechanik bewegen, und daß sie der Schwerkraft unterworfen seien. Dann würden sie einen Zuwachs an lebendiger Kraft erfahren, welcher gleich der Abnahme der potentiellen Energie ist. Dagegen ist der nach (6) [eq. 10] berechnete Zuwachs der Lichtteilehen an lebendiger Kraft gleich der Zunahme ihrer potentiellen Energie, d. h. von gleichem Betrage, aber von entgegengesetztem Vorzeichen, wie der aus der Emissionstheorie folgende. Was aber die Krümmung der Lichtstrahlen im Schwerkraftfelde anbelangt, die aus (6) [eq. 10] mit Hilfe des Huygensschen Prinzips sich ableiten läßt, so ist sie identisch mit der Krümmung der Bahnkurve jener Lichtteilchen. Es ist dies eine der zahlreichen unvollständigen Analogien zwischen der modernen Strahlungstheorie und der Emissionstheorie des Lichtes." (Abraham 1912h, 2)

53 "A. Einstein hat gezeigt, daß ein die Sonnenoberfläche passierender Lichtstrahl nach der Sonne hin abgelenkt wird, und hat die Aufmerksamkeit der Astronomen auf diese Konsequenz der Theorie gelenkt, die vielleicht zu ihrer Prüfung dienen kann." (Abraham 1912h, 2, fn.1)

In a publication following immediately, “The Elementary Law of Gravitation,”[55] which was also submitted in December 1911, Abraham returned to the question of the relation of his theory to experience. The principal aim of this paper is the derivation from the theory published in the preceding paper of an expression of the gravitational force between two “world points” P and P_0, “an elementary law of gravitation,” as Abraham called it. For this purpose he integrated the field equation of his theory, equation (2), using the Cauchy method of residues and following a procedure by Herglotz.[56] Such an elementary law allowed for a comparison not only with Newton’s law of gravitation but also with earlier adaptations of action-at-a-distance laws to a relativistic framework. Furthermore, it allowed Abraham, at least in principle, to address the astronomical consequences of his theory and to explore deviations from classical predictions.

He concluded his publication with the following summary:

> According to this elementary law the moving force exerted by P on P_0 is represented as a sum of two four-vectors, of which one is parallel to the radius $\mathfrak{R}$, drawn from the world point P_0 to P, the other to the velocity vector $\mathfrak{B}$ of P. This corresponds to the approaches of Poincaré and Minkowski. But our elementary law is simpler, insofar as it does not involve the velocity of the attracted point, and more general, because it also takes into account the acceleration of the attracting point. Its comparison with astronomical observation could serve for the examination of the theory of gravitation developed in the previous note.[57]

Abraham, however, did not specify which astronomical phenomena might be compared to the consequences of his elementary law. But modifications of Newton’s elementary law had earlier been referred to the perihelion anomaly of Mercury[58] so that this may well have been one of the astronomical consequences Abraham had in mind. In other words, the deviation of planetary motion from the implications of Newtonian mechanics was also part of the generally shared knowledge that a new gravitation theory had to confront. On the basis of estimating the order of magnitude of deviations from Newton’s law, Abraham himself later remarked:

---

54 Abraham to Schwarzschild, see “The Continuity between Classical and Relativistic Cosmology in the Work of Karl Schwarzschild,” p. 165, (in this volume).

55 See (Abraham 1912a).

56 See (Herglotz 1904), and also (Sommerfeld 1910, 665; 1911, 51) where according to Abraham the analogous electrical problem is treated.

57 “Diesem Elementargesetz zufolge stellt sich die von P auf P_0 ausgeübte bewegende Kraft als Summe zweier Vierervektoren dar, von denen der eine dem vom Weltpunkte P_0 nach P gezogenen Fahrstrahl $\mathfrak{B}$, der andere dem Geschwindigkeitsvektor $\mathfrak{B}$ von P parallel ist. Dieses entspricht den Ansätzen von Poincaré und von Minkowski. Doch unser Elementargesetz einfacher, insofern als die Geschwindigkeit des angezogenen Punkts nicht eingeht, und allgemeiner, weil es auch die Beschleunigung des anziehenden Punkts berücksichtigt. Seine Vergleichung mit der astronomischen Beobachtung könnte zur Prüfung der in der vorigen Note entwickelten Theorie der Schwerkraft deinen.” (Abraham 1912a, 5)

58 See (Zenneck 1903), compare also (Poincaré 1906 and Minkowski 1909).

> The correction of Newton's law introduced in (16)[59] is hence likely to be too small for noticeably influencing the planetary motion. (Abraham 1912b, 797)

Nevertheless, in the same year 1912, G. Pavanini calculated the perihelion shift of Mercury according to Abraham's theory, finding a value of $14'', 52$, that is, approximately one third of the observed value.[60] Abraham's theory thus made a more accurate prediction than the much more elaborate theory of gravitation Einstein and Grossmann published in 1913.[61]

2.2 Audacious Outlooks

The empirical consequences of Abraham's theory of gravitation were by no means limited to those already envisaged by Einstein, or to those immediately obvious from the body of knowledge covered by the Newtonian theory. Abraham was a master of classical electrodynamics, and, for this reason, was not only used to elaborating in depth and detail the consequences of a complex theory, but also had a specific model that suggested where to look for analogies and differences between the two field theories—electromagnetic and gravitational. In the following, we will briefly look at two of the outstanding achievements that resulted from Abraham's efforts to draw far-reaching physical consequences from his theory. Both concern subjects of great interest to present research in general relativity, spacetime singularities and gravitational waves. Until now, the name Max Abraham has played no role in the history of these subjects, although his contributions are unlikely to have gone entirely unnoticed by those who continued the search for a relativistic theory of gravitation at the time he gave up.

In a paper submitted in March 1912, "The Free Fall" (Abraham 1912c), Abraham considered the motion of free fall in a homogeneous gravitational field, i.e., in exactly the same kind of gravitational field that was used in the formulation of Einstein's equivalence principle, and drew some far-reaching cosmological consequences from his calculations. These consequences essentially depend on equations (10), relating the speed of light and the gravitational potential, and on the relation between coordinate time and proper time. From the law of fall, which Abraham derived from his equations of motion in a gravitational field, he first concluded that there is a point in (coordinate) time, corresponding to a certain distance of fall, at which the speed of light becomes zero. This point corresponds to a singularity in the relation between proper time and coordinate time. From the condition that the speed of light must always remain larger than zero, he then concluded, with the help of (10), that:[62]

---

59 A law of the form $Force = \frac{A}{r^2} - \frac{B}{r^3}$.

60 See (Pavanini 1912). See also the brief discussion of these results in (Abraham 1915, 488). For the corresponding calculation in Nordström's theory, see (Behacker 1913, 989).

61 See (Einstein and Grossmann 1913). See also "What Did Einstein Know …" (in vol. 2 of this series).

62 See (Abraham 1912c, 311).

$$\Phi_0 - \Phi < \frac{c_0^2}{2}. \tag{17}$$

As a first step, he interpreted this relation as implying that the existence of a homogeneous gravitational field of infinite extension is excluded and then turned to further, as he called them, "cosmogonic" consequences of this relation.
He considered a star of mass m with the potential:

$$\Phi = -\gamma \cdot \frac{m}{r} \tag{18}$$

and compared the potential difference between an infinite distance and the surface of the star at distance $r = a$:

$$\Phi_0 - \Phi = \gamma \cdot \frac{m}{a}. \tag{19}$$

From equation (17) he then inferred that there is a maximum for the quotient of mass and radius:

> Therefore
>
> $$\gamma \cdot \frac{m}{a} < \frac{c_0^2}{2}\left(= \frac{9}{2} \cdot 10^{20}\right) \tag{20}$$
>
> must hold. For the Sun one has in *cgs* units
>
> $$\left(\gamma \cdot \frac{m}{a}\right)_s = 2 \cdot 10^{15}. \tag{21}$$
>
> Thus, the *quotient of mass and radius for an arbitrary star must satisfy the inequality*
>
> $$\frac{m}{a} < 2,25 \cdot 10^5 \cdot \left(\frac{m}{a}\right)_s. \tag{22}$$
>
> For stars whose mean density is equal to that of the Sun, the quotients m/a are proportional to the squares of the radii. Hence, the following must hold
>
> $$a < 500(a)_s, \tag{23}$$
>
> and thus
>
> $$m < 10^8(m)_s. \tag{24}$$
>
> That is, *the mass of a star whose mean density is equal to that of the Sun cannot become larger than one hundred million times the mass of the Sun.*
>
> Since this limit is rather high, no difficulties arise from this for our theory.[63]

Abraham, in March 1912, was hence the first to hit upon a singularity in a field theory of gravitation and to calculate what was later called the "Schwarzschild radius." Although his understanding of the relation between proper time and coordinate time does not correspond to the modern one, his consideration of their relation for a freely falling object in the field of a point mass corresponds to a procedure still applied in modern general relativity. From his introductory reference to the "cosmogonic" implications of his theory one may further infer that he did not consider this singularity to be fictitious, contrary to the first explorers of the Schwarzschild singularity within general relativity, Eddington and Lemaitre.[64] Abraham did not, however, directly relate his result to light deflection, although the vanishing of the speed of light at the singularity clearly implies that light cannot escape from it. But Abraham did not make this consequence explicit and rather limited himself to interpreting the singularity in terms of limits on the size of stars.

In a lecture presented in October 1912 and published the following year (Abraham 1913) Abraham was also the first to discuss the possibility of gravitational waves in a relativistic field theory of gravitation. There he wrote:

> According to our theory, light and gravitation have the same speed of propagation; but whereas light waves are transverse, gravitational waves are longitudinal. Incidentally, the problem of the oscillating particle can be treated in a similar manner as that of the oscillating electron; the strength of the emitted gravitational waves depends on the product of gravitational mass and the acceleration of the particle. Is it possible to detect these gravitational waves?
>
> This hope is futile. Indeed, to impart an acceleration to one particle, another particle is necessary which, according to the law of action and reaction, is driven in the opposite side. But now, the strength of the emitted gravitational waves depends on the sum of the products of the gravitational mass and the acceleration of the two particles, while, according to the reaction principle, the sum of the products of inertial mass and acceleration is equal to zero.[65] Therefore, although the existence of gravitational waves is compatible with the assumed field mechanism, through the equality of gravitational and inertial mass the possibility of its production is practically excluded. It follows from this that the planetary system does not lose its mechanical energy through radiation, whereas an analogous system consisting of negative electrons circling around a positive nucleus gradually radiates its energy away. The life of the planetary system is thus not threatened by such a danger.[66]

Abraham's argument amounts to showing that, because of momentum conservation, there can be no dipole moment in gravitational waves; it is an argument still used in

---

63 Es muß also sein: [(eq. 20)]. Für die Sonne hat man, in C.G.S.Einheiten [(eq. 21)]. Daher muß *der Quotient aus Masse und Radius für einen beliebigen Stern der Ungleichung genügen* [(eq. 22)]. Für Sterne, deren mittlere Dichte derjenigen der Sonne gleich ist, stehen die Quotienten m/a im Verhältnis der Quadrate der Radien; es muß dann sein [(eq. 23)] und somit [(eq. 24)]. D h. *die Masse eines Sternes, dessen mittlere Dichte derjenigen der Sonne gleich ist, kann nicht größer werden als das Hundertmillionenfache der Masse der Sonne*. Da diese Grenze recht hoch ist, so entsteht hieraus keine Schwierigkeit für unsere Theorie." (Abraham 1912c, 311)

64 See (Eisenstaedt 1989).

65 *Abraham's footnote*: Here the momentum of the gravitational field has, however, been neglected; but this [momentum] practically comes as little into consideration as the energy of the field. ("Hierbei ist allerdings der Impuls des Schwerefeldes unberücksichtigt geblieben; aber dieser kommt ebensowenig wie die Energie des Feldes praktisch in Betracht.")

modern textbooks on general relativity.[67] His allusion to an electrical analogue is clearly a reference to a Bohr-like atomic model, which classically would, of course, be unstable.

Abraham nevertheless took the idea of gravitational waves so seriously that he published another detailed study comparing electromagnetic and gravitational waves, including quantitative comparisons.[68] In a later review he took up this earlier study and discussed the possibility that gravitational waves are emitted during the emission of α particles by radioactive atoms:

> One could now surmise that during the emission of α particles by radioactive atoms, in which very large accelerations occur, the hypothetical gravitational waves are excited in noticeable strength. But at the same time there are electrical waves which are excited and, as the quantitative discussion shows, the force which is exerted by the gravitational waves emitted by an α particle upon another α particle amounts to maximally 10^{-36} of the electrical force.[69]

If, or rather when, gravitational waves are one day directly verified, Abraham's papers will certainly not constitute a relevant theoretical reference point. From the point of view of a history of knowledge, however, they do represent a reference point for gauging the possibilities open to the development of a theory of gravitation around 1912.[70] While Abraham's efforts are hardly suited to detracting from Einstein's triumph of late 1915, they do make evident the extent to which this solitary tri-

---

66 "Nach unserer Theorie haben Licht und Schwere die gleiche Fortpflanzungsgeschwindigkeit; aber während die Lichtwellen transversal sind, sind die Schwerewellen longitudinal. Übrigens kann das Problem des schwingenden Massenteilchens in ähnlicher Weise behandelt werden wie dasjenige des schwingenden Elektrons; die Stärke der ausgesandten Gravitationswellen hängt von dem Produkt aus schwerer Masse und Beschleunigung des Teilchens ab. Ist es möglich, diese Schwerewellen zu entdecken?
Diese Hoffnung ist vergeblich. In der Tat, um einem Teilchen eine Beschleunigung zu erteilen, bedarf es eines anderen Teilchens, welches, vermöge des Gesetzes von Wirkung und Gegenwirkung, nach der entgegengesetzten Seite getrieben wird. Nun hängt aber die Stärke der ausgesandten Schwerewelle von der Summe der Produkte aus schwerer Masse und Beschleunigung der beiden Teilchen ab, während nach dem Gegenwirkungsprinzip die Summe der Produkte aus träger Masse und Beschleunigung gleich null ist.[1)] [see previous note] Es ist also zwar die Existenz der Gravitationswellen mit dem angenommenen Feldmechanismus verträglich, aber durch die Identität von schwerer und träger Masse wird die Möglichkeit ihrer Erzeugung praktisch ausgeschlossen. Hieraus geht hervor, daß das Planetensystem nicht seine mechanische Energie durch Strahlung verliert, während ein analoges System, bestehend aus negativen Elektronen, die um einen positiven Kern kreisen, allmählich seine Energie ausstrahlt. Das Leben des Plantensystems ist also nicht durch eine solche Gefahr bedroht." (Abraham 1913, 208–209)

67 See, e.g., (Wald 1984, 83).

68 See (Abraham 1912g).

69 "Man könnte nun vermuten, daß während der Emission von α -Strahlen durch radioaktive Atome, bei der sehr große Beschleunigungen auftreten, die hypothetischen Schwerkraftwellen in merklicher Stärke erregt werden. Indessen werden dabei gleichzeitig elektrische Wellen erregt und, wie die quantitative Diskussion zeigt, beträgt die Kraft, welche die von einem α -Teilchen entsandten Schwerewellen auf ein zweites α -Teilchen ausüben, höchstens 10^{-36} der elektrischen Kraft." (Abraham 1915, 487)

umph merely realized a potential inherent in the shared knowledge of the time. After all, even if considered in hindsight, Abraham's thoughts on maximal sizes of stars and gravitational waves represent the grandiose vistas offered by an outlook from considerable heights, even if it is also clear in hindsight that Abraham was not the one who reached the highest summit.

REFERENCES

Abraham, Max. 1912a. "Das Elementargesetz der Gravitation." *Physikalische Zeitschrift* (13) 4–5.

———. 1912b. "Das Gravitationsfeld." *Physikalische Zeitschrift* 13: 793–797.

———. 1912c. "Der freie Fall." *Physikalische Zeitschrift* 13: 310–311. (English translation in this volume.)

———. 1912d. "Die Erhaltung der Energie und der Materie im Schwerkraftfelde." *Physikalische Zeitschrift* (13) 311–314.

———. 1912e. "Nochmals Relativität und Gravitation. Bemerkungen zu A. Einsteins Erwiderung." *Annalen der Physik* (38) 444–448.

———. 1912f. "Relativität und Gravitation. Erwiderung auf eine Bemerkung des Herrn A. Einstein." *Annalen der Physik* (38) 1056–1058.

———. 1912g. "Sulle onde luminose e gravitazionali." *Nuovo Cimento* (6), 3: 211.

———. 1912h. "Zur Theorie der Gravitation." *Physikalische Zeitschrift* 13: 1–4, "Berichtigung," p. 176. (English translation in this volume.)

———. 1913. "Eine neue Gravitationstheorie." *Archiv der Mathematik und Physik* (20): 193–209. (English translation in this volume.)

———. 1914. "Die neue Mechanik." *Scientia* 15: 8–27.

———. 1915. "Neuere Gravitationstheorien." *Jahrbuch der Radioaktivität und Elektronik* 11: 470–520. (English translation in this volume.)

Abraham, Max and August Föppl. 1904–1908. *Theorie der Elektrizität* (2 vols). Leipzig: Teubner.

Behacker, M. 1913. "Der freie Fall und die Planetenbewegung in. Nordströms Gravitationstheorie." *Physikalische Zeitschrift*: 14: 989–992.

Cattani, Carlo, and Michelangelo De Maria. 1989. "Max Abraham and the Reception of Relativity in Italy: His 1912 and 1914 Controversies with Einstein." In D. Howard and J. Stachel (eds.), *Einstein and the History of General Relativity.* (*Einstein Studies* vol. 1.) Boston: Birkhäuser, 160–174.

CPAE 3. 1993. Martin J. Klein, A. J. Kox, Jürgen Renn, and Robert Schulmann (eds.), *The Collected Papers of Albert Einstein*. Vol. 3. *The Swiss Years: Writings, 1909–1911*. Princeton: Princeton University Press.

CPAE 4. 1995. Martin J. Klein, A. J. Kox, Jürgen Renn, and Robert Schulmann (eds.), *The Collected Papers of Albert Einstein*. Vol. 4. *The Swiss Years: Writings, 1912–1914*. Princeton: Princeton University Press.

CPAE 5. 1993. Martin J. Klein, A. J. Kox, and Robert Schulmann (eds.), *The Collected Papers of Albert Einstein*. Vol. 5. *The Swiss Years: Correspondence, 1902–1914*. Princeton: Princeton University Press.

Einstein, Albert. 1911. "Über den Einfluß der Schwerkraft auf die Ausbreitung des Lichtes." *Annalen der Physik* 35: 898–908, (CPAE 3, Doc. 23).

———. 1912. "Lichtgeschwindigkeit und Statik des Gravitationsfeldes." *Annalen der Physik* 38: 355–369, (CPAE 4, Doc. 3).

Einstein, Albert, and Marcel Grossmann. 1913. "Entwurf einer verallgemeinerten Relativitätstheorie und einer Theorie der Gravitation." *Zeitschrift für Mathematik und Physik* (62) 3: 225–261.

Eisenstaedt, Jean. 1989. "The Early Interpretation of the Schwarzschild Solution." In D. Howard and J. Stachel (eds.), *Einstein and the History of General Relativity*, (*Einstein Studies* vol. 1). Boston: Birkhäuser, 213–233.

Herglotz, G. 1904. "Zur Elektronentheorie." *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen. Mathematisch-physikalische Klasse* 357–382.

Kennefick, Daniel. 1999. "Controversies in the History of the Radiation Reaction Problem in General Relativity." In H. Goenner, J. Renn, J. Ritter and T. Sauer (eds.), *The Expanding Worlds of General Relativity.* (*Einstein Studies* vol. 7.) Boston: Birkhäuser.

70 For further discussion, see (Kennefick 1999, 2006).

———. 2006. *Traveling at the Speed of Thought: Einstein and the Quest for Gravitational Waves*. Princeton: Princeton University Press.

Kretschmann, Erich. 1917. "Über den physikalischen Sinn der Relativitätspostulate, A. Einsteins neue und seine ursprüngliche Relativitätstheorie." *Annalen der Physik* (53) 16: 575–614.

Minkowski, Hermann. 1908. "Die Grundgleichungen für elektromagnetischen Vorgänge in bewegten Körpern." *Nachrichten der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, 53–111. (English translation of the appendix "Mechanics and the Relativity Postulate" in this volume.)

Norton, John. 1992. "The Physical Content of General Covariance." In J. Eisenstaedt and A. J. Kox (eds.), *Studies in the History of General Relativity*, (*Einstein Studies* vol. 3). Boston: Birkhäuser, 281–315.

———. 1993. "General Covariance and the Foundations of General Relativity: Eight Decades of Dispute." *Reports on Progress in Physics* (56) 791–858.

Pavanini, G. 1912. "Prime consequenze d'una recente teoria della gravitazione." *Rendic. d. R. Acc. dei Lincei* XXI ^{2}, 648; XXII^{1}, 369, 1913.

Poincaré, Henri. 1906. "Sur la dynamique de l'électron." *Rendiconto del Circolo Matematico di Palermo* 21: 129–175. (English translation in this volume.)

Renn, Jürgen, Peter Damerow and Simone Rieger. 2001. "Hunting the White Elephant: When and How did Galileo Discover the Law of Fall? (with an Appendix by Domenico Giulini)." In J. Renn (ed.), *Galileo in Context*. Cambridge: Cambridge University Press, 29–149.

Rynasiewicz, Robert. 1999. "Kretschmann's Analysis of Covariance and Relativity Principles." In H. Goenner, J. Renn, J. Ritter and T. Sauer (eds.), *The Expanding Worlds of General Relativity*, (*Einstein Studies* vol. 7). Boston: Birkhäuser, 431–462.

Sommerfeld, Arnold. (ed). 1903–1926. *Physik*, 3 vols. (*Encyklopädie der mathematischen Wissenschaften mit Einschluss ihrer Anwendungen,* vol. 5.) Leipzig: Teubner.

Sommerfeld, Arnold. 1910. "Zur Relativitätstheorie II. Vierdimensionale Vektoranalysis." *Annalen der Physik* (33) 649–689.

———. 1911 *Sitzungsberichte d. Bayer. Akad. d. Wissensch*. p. 51.

Wald, Robert M. 1984. *General Relativity.* Chicago: Chicago University Press.

Zenneck, Jonathan. 1903. "Gravitation." In (Sommerfeld 1903–1926), 1: 25–67. (Printed in this volume.)

MAX ABRAHAM

ON THE THEORY OF GRAVITATION

Originally published in Rendiconti della R. Accademia dei Lincei. German translation by the author published as "Zur Theorie der Gravitation" in Physikalische Zeitschrift 13, 1912, pp. 1–4; correction, p. 176. Received December 14, 1911. Author's date: Milan, 1911. English translation taken from the German.

In a recently published paper A. Einstein[1] proposed the hypothesis that the speed of light (c) depends on the gravitational potential (Φ). In the following note, I develop a theory of the gravitational force which satisfies the principle of relativity and derive from it a relation between c and Φ, which in first approximation is equivalent to Einstein's. This theory attributes values to the densities of the energy and the energy flux of the gravitational field which differ from those hitherto assumed.

Following Minkowski's[2] presentation, we consider

$$x, y, z \qquad \text{and} \qquad u = il = ict$$

as the coordinates of a four-dimensional space. Let the "rest density" ν, as well as the gravitational potential Φ, be scalars in this space, and let them be linked through the differential equation:

$$\frac{\partial^2\Phi}{\partial x^2} + \frac{\partial^2\Phi}{\partial y^2} + \frac{\partial^2\Phi}{\partial z^2} + \frac{\partial^2\Phi}{\partial u^2} = 4\pi\gamma\nu. \tag{1}$$

(γ is the gravitational constant.)
The "accelerating force" [*bewegende Kraft*] acting on the unit mass in the gravitational field is a four-vector

$$\mathfrak{F} = -\mathrm{Grad}\Phi \tag{2}$$

with components

$$\mathfrak{F}_x = -\frac{\partial\Phi}{\partial x}, \quad \mathfrak{F}_y = -\frac{\partial\Phi}{\partial y}, \quad \mathfrak{F}_z = -\frac{\partial\Phi}{\partial z}, \quad \mathfrak{F}_u = -\frac{\partial\Phi}{\partial u}. \tag{2a}$$

1 A. Einstein, *Ann. d. Phys*. 35, p. 898, 1911.
2 H. Minkowski, *Göttinger Nachr*. 1908, p. 53.

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 3
Gravitation in the Twilight of Classical Physics: Between Mechanics, Field Theory, and Astronomy.

According to eqs. (1) and (2), the gravitational force propagates with the speed of light as required by the principle of relativity; whereas, however, light waves are transverse, the *gravitational waves are longitudinal*.

We write $\dot{x}, \dot{y}, \dot{z}, \dot{u}$ for the first derivatives of the coordinates of a material "world point" with respect to its "proper time" τ, i.e., for the components of the "velocity" four-vector $\mathfrak{V}$ and $\ddot{x}, \ddot{y}, \ddot{z}, \ddot{u}$ for the second derivatives, i.e., for the components of the "acceleration" four-vector $\dot{\mathfrak{V}}$. Then the *equations of motion*[3] are:

$$\ddot{x} = \mathfrak{F}_x, \quad \ddot{y} = \mathfrak{F}_y, \quad \ddot{z} = \mathfrak{F}_z, \quad \ddot{u} = \mathfrak{F}_u. \tag{3}$$

Between the first derivatives exist the identity: |

[2]
$$\dot{x}^2 + \dot{y}^2 + \dot{z}^2 + \dot{u}^2 = -c^2, \tag{4}$$

or

$$\dot{l}^2 \left\{ \left(\frac{dx}{dl}\right)^2 + \left(\frac{dy}{dl}\right)^2 + \left(\frac{dz}{dl}\right)^2 - 1 \right\} = -c^2.$$

Therefore, if one sets

$$\beta^2 = \left(\frac{dx}{dl}\right)^2 + \left(\frac{dy}{dl}\right)^2 + \left(\frac{dz}{dl}\right)^2, \qquad \kappa = \sqrt{1-\beta^2}, \tag{4a}$$

it follows that

$$\dot{l} = \frac{c}{\sqrt{1-\beta^2}} = c\kappa^{-1}. \tag{4b}$$

Now, by differentiating eq. (4) with respect to proper time, Minkowski obtains the condition of the "orthogonality" of the velocity and acceleration four-vectors. However, if c is considered to be variable, the place of that condition is taken by the following:

$$\dot{x}\ddot{x} + \dot{y}\ddot{y} + \dot{z}\ddot{z} + \dot{u}\ddot{u} = -c\frac{dc}{d\tau}, \tag{5}$$

as the differentiation of eq. (4) shows. By introducing here, instead of the acceleration, the accelerating force according to (3) and (2), we obtain

$$\dot{x}\frac{\partial\Phi}{\partial x} + \dot{y}\frac{\partial\Phi}{\partial y} + \dot{z}\frac{\partial\Phi}{\partial z} + \dot{u}\frac{\partial\Phi}{\partial u} = c\frac{dc}{d\tau}$$

---

3 As I have shown, when energy of a non-mechanical nature is supplied to matter, the Minkowskian equations of motion require modification. See this journal 10, 737, 1909; 11, 527, 1910. But, now, we are dealing with purely mechanical effects so that this modification need not be considered.

or

$$\frac{d\Phi}{d\tau} = c\frac{dc}{d\tau}.$$

The integration yields

$$\frac{c^2}{2} - \frac{c_0^2}{2} = \Phi - \Phi_0, \tag{6}$$

where c_0 and Φ_0 represent the speed of light and the gravitational potential at the origin of coordinates. Equation (6) implies: *The increase of half the square of the speed of light is equal to the increase of the gravitational potential.*

Instead of this relation, which is exactly valid according to our theory, one can, neglecting the square of the quotient of Φ and c^2, take Einstein's formula (loc. cit. p. 906):

$$c = c_0\left(1 + \frac{\Phi - \Phi_0}{c^2}\right).$$

However, eq. (6) better serves to manifest the independence from the arbitrarily chosen origin of coordinates.

It is instructive to compare the relation obtained with the emission theory of light. Let us imagine that the particles emitted from the light source move according to the laws of Galilean mechanics and that they are subject to the gravitational force. They would then experience an increase of *vis viva* [*lebendiger Kraft*] equal to the *decrease* of potential energy. The increase of *vis viva* of the light particles calculated according to (6) is, by contrast, equal to the *increase* of their potential energy, i.e., of the same amount but of opposite sign as that following from the emission theory. But as far as the bending of light rays in the gravitational field is concerned, which can be derived from (6) with the aid of Huygens' principle,[4] it is identical with the bending of the trajectories of those light particles. This is one of the numerous incomplete analogies between the modern theory of radiation and the emission theory of light.

We consider the motion of a material point of mass m in a gravitational field. The first three equations of motion yield:

$$m\frac{d\dot{x}}{d\tau} = -m\frac{\partial\Phi}{\partial x}, \qquad m\frac{d\dot{y}}{d\tau} = -m\frac{\partial\Phi}{\partial y}, \qquad m\frac{d\dot{z}}{d\tau} = -m\frac{\partial\Phi}{\partial z}; \tag{7}$$

they contain the law of conservation of momentum [*Impulssatz*]. In contrast, the last of the eqs. (3):

4 A. Einstein has shown that a light ray passing the surface of the Sun is deflected towards the Sun and has drawn the attention of the astronomers to this consequence of the theory which can perhaps serve as its verification.

$$m\frac{dl}{d\tau} = -mi\mathfrak{F}_u = mi\frac{\partial\Phi}{\partial u} = m\frac{\partial\Phi}{\partial l} \tag{8}$$

expresses the law of conservation of *vis viva* in Minkowskian mechanics. In particular, if the gravitational field depends on the position but not on the time, then one obtains, multiplying (8) by c and taking (4b) into consideration:

$$mc\frac{d}{d\tau}(c\kappa^{-1}) = 0. \tag{9}$$

Now, by considering c as constant, Minkowski interprets $m(c^2\kappa^{-1} - 1)$ as the kinetic energy of the material point. In contrast, in the theory developed here, which takes c to be variable, this would not be permissible. It also seems impossible in this case to give a general expression for the energy of the material point whose decrease would be precisely equal to the energy extracted from the gravitational field.

However, we can convince ourselves that, at least for small velocities, the theorem of conservation of energy, which is confirmed by experience in this realm, follows from (9). From (9) one obtains

$$mc\kappa^{-1} = \text{const.} \tag{9a}$$

[3] | By neglecting squares and products of β^2 and $\dfrac{\Phi - \Phi_0}{c^2}$, according to (4a) we write:

$$\kappa^{-1} = (1-\beta^2)^{-1/2} = 1 + \frac{1}{2}\beta^2, \qquad \beta^2 = \frac{v^2}{c_0^2},$$

and, according to (6):

$$c = \{c_0^2 + 2(\Phi - \Phi_0)\}^{1/2} = c_0 + \frac{\Phi - \Phi_0}{c_0}.$$

Then, we obtain

$$mc\kappa^{-1} = mc_0\left(1 + \frac{1}{2}\beta^2\right) + m\frac{\Phi - \Phi_0}{c_0}.$$

Therefore, if we multiply (9a) by c_0, it follows that

$$\frac{1}{2}mv^2 + m\Phi = \text{const.}, \tag{9b}$$

i.e., the law of the conservation of energy in the usual form. *Thus, in the limiting case of small velocities, the new mechanics coincides with the old.* As a consequence of the relation (6) between the velocity of light c and the gravitational potential Φ, it also now emerges clearly that not only the "*kinetic*" *energy* $1/2\ mv^2$, but also the *potential energy* $m\Phi$ *are associated with the material point.*

We now consider two material points of masses m and m_0 moving with small velocities in a stationary gravitational field. Each of the two points possesses a potential energy, of which the part depending upon the mutual distance r is:

$$-\gamma \frac{m_0 m}{r} = -E.$$

Therefore, the total of this variable part of the potential energy for the two points amounts to $-2E$. Hence, if, under their mutual attraction, the two points approach one another, then the *increase in their total vis viva is equal to half of the decrease of their total potential energy.*

Where now does the other half reside? Obviously in the gravitational field. Indeed, as we shall see, in this case, our theory ascribes to the field energy the value E, which is equal and opposite to the one previously assumed. In this way, the difficulty emphasized by Maxwell[5]—that the energy density in a gravitational field, when set to zero for vanishing forces, would become negative elsewhere—disappears. *The expression (13) for the energy density in a gravitational field at which we will arrive is strictly positive*. But, besides the field energy E, the total energy of the system also contains the energy of the matter whose potential part in the stationary field is $-2E$.

The credit for having extended the concept of energy flux to the gravitational field goes to V. Volterra.[6] However, his expression for the energy flux is based on Maxwell's assumption concerning the energy distribution in the field. Accordingly, the theory developed here will thus lead to different results also with respect to energy flux.

The fictitious stresses, the energy flux as well as the densities of energy and momentum for a field depend on a "*four-dimensional tensor*".[7] We write for the *ten components of the gravitational tensor*:

5 [J.] Clerk Maxwell, *Scientific papers* 1, 570.

6 V. Volterra, *Nuovo Cimento* 337, 1899, 1.

7 Regarding these "world tensors" or "ten-tensors" see M. Abraham, *Rendiconti del circolo matematico di Palermo* 1910, 1; A. Sommerfeld, *Ann. d. Phys.* 32, 749, 1910; M. Laue, "Das Relativitätsprinzip", Braunschweig 1911, p. 73.

$$\left.\begin{aligned} X_x &= \frac{1}{4\pi\gamma}\left\{-\left(\frac{\partial\Phi}{\partial x}\right)^2+\Psi\right\} \\ Y_y &= \frac{1}{4\pi\gamma}\left\{-\left(\frac{\partial\Phi}{\partial y}\right)^2+\Psi\right\} \\ Z_z &= \frac{1}{4\pi\gamma}\left\{-\left(\frac{\partial\Phi}{\partial z}\right)^2+\Psi\right\} \\ X_y &= Y_x = -\frac{1}{4\pi\gamma}\frac{\partial\Phi}{\partial x}\frac{\partial\Phi}{\partial y} \\ Y_z &= Z_y = -\frac{1}{4\pi\gamma}\frac{\partial\Phi}{\partial y}\frac{\partial\Phi}{\partial z} \\ Z_x &= X_z = -\frac{1}{4\pi\gamma}\frac{\partial\Phi}{\partial z}\frac{\partial\Phi}{\partial x} \end{aligned}\right\} \tag{10}$$

$$\left.\begin{aligned} X_u &= U_x = -\frac{1}{4\pi\gamma}\frac{\partial\Phi}{\partial u}\frac{\partial\Phi}{\partial x} \\ Y_u &= U_y = -\frac{1}{4\pi\gamma}\frac{\partial\Phi}{\partial u}\frac{\partial\Phi}{\partial y} \\ Z_u &= U_z = -\frac{1}{4\pi\gamma}\frac{\partial\Phi}{\partial u}\frac{\partial\Phi}{\partial z} \end{aligned}\right\} \tag{10a}$$

$$U_u = \frac{1}{4\pi\gamma}\left\{-\left(\frac{\partial\Phi}{\partial u}\right)^2+\Psi\right\}. \tag{10b}$$

Here, Φ and

$$\Psi = \frac{1}{2}\left\{\left(\frac{\partial\Phi}{\partial x}\right)^2+\left(\frac{\partial\Phi}{\partial y}\right)^2+\left(\frac{\partial\Phi}{\partial z}\right)^2+\left(\frac{\partial\Phi}{\partial u}\right)^2\right\} \tag{10c}$$

are four-dimensional scalars. Thus, it becomes immediately clear that the components (10), (10a), (10b) of the ten-tensor indeed transform like the squares and products of the coordinates x, y, z, u. From them, | the components of the *accelerating force* per unit volume are derived as follows: [4]

$$\left.\begin{aligned}
\nu\mathfrak{F}_x &= \frac{\partial X_x}{\partial x} + \frac{\partial X_y}{\partial y} + \frac{\partial X_z}{\partial z} + \frac{\partial X_u}{\partial u} \\
\nu\mathfrak{F}_y &= \frac{\partial Y_x}{\partial x} + \frac{\partial Y_y}{\partial y} + \frac{\partial Y_z}{\partial z} + \frac{\partial Y_u}{\partial u} \\
\nu\mathfrak{F}_z &= \frac{\partial Z_x}{\partial x} + \frac{\partial Z_y}{\partial y} + \frac{\partial Z_z}{\partial z} + \frac{\partial Z_u}{\partial u} \\
\nu\mathfrak{F}_u &= \frac{\partial U_x}{\partial x} + \frac{\partial U_y}{\partial y} + \frac{\partial U_z}{\partial z} + \frac{\partial U_u}{\partial u}
\end{aligned}\right\} \tag{11}$$

By substituting the above expressions for the components of the ten-tensor and by taking into consideration the differential equation (1), one arrives at the value

$$\nu\mathfrak{F} = -\nu\mathrm{Grad}\Phi. \tag{12}$$

for the accelerating force per unit volume, in agreement with (2).

We discuss the formulae for the components of the gravitational tensor. The first six components (10) determine the "*fictitious stresses*" in the gravitational field. In the *stationary field* there exist the *normal stresses*:

$$\left.\begin{aligned}
X_x &= \frac{1}{8\pi\gamma}\left\{-\left(\frac{\partial\Phi}{\partial x}\right)^2 + \left(\frac{\partial\Phi}{\partial y}\right)^2 + \left(\frac{\partial\Phi}{\partial z}\right)^2\right\}, \\
Y_y &= \frac{1}{8\pi\gamma}\left\{\left(\frac{\partial\Phi}{\partial x}\right)^2 - \left(\frac{\partial\Phi}{\partial y}\right)^2 + \left(\frac{\partial\Phi}{\partial z}\right)^2\right\}, \\
Z_z &= \frac{1}{8\pi\gamma}\left\{\left(\frac{\partial\Phi}{\partial x}\right)^2 + \left(\frac{\partial\Phi}{\partial y}\right)^2 - \left(\frac{\partial\Phi}{\partial z}\right)^2\right\},
\end{aligned}\right\} \tag{12a}$$

and the *shear stresses*

$$\left.\begin{aligned}
X_y &= Y_x = -\frac{1}{4\pi\gamma}\frac{\partial\Phi}{\partial x}\frac{\partial\Phi}{\partial y}, \\
Y_z &= Z_y = -\frac{1}{4\pi\gamma}\frac{\partial\Phi}{\partial y}\frac{\partial\Phi}{\partial z}, \\
Z_x &= X_z = -\frac{1}{4\pi\gamma}\frac{\partial\Phi}{\partial z}\frac{\partial\Phi}{\partial x}.
\end{aligned}\right\} \tag{12b}$$

They correspond to a *pressure along the lines of force and to a pull perpendicular to the lines of force, and their total value is proportional to the field strength and, for a* *stationary field, is equal to the energy density* (ε) (see eq. (13)). For a temporally varying field there is still a *hydrostatic pressure* to be added:

$$p = \frac{1}{8\pi\gamma}\left(\frac{\partial\Phi}{\partial l}\right)^2, \tag{12c}$$

which has to be subtracted from the normal stress (12a).

The *energy density* in the *gravitational field* is:

$$\varepsilon = U_u = \frac{1}{8\pi\gamma}\left\{\left(\frac{\partial\Phi}{\partial x}\right)^2 + \left(\frac{\partial\Phi}{\partial y}\right)^2 + \left(\frac{\partial\Phi}{\partial z}\right)^2 + \left(\frac{\partial\Phi}{\partial l}\right)^2\right\}. \tag{13}$$

This turns out to be *strictly positive* and, in particular, to be *proportional to the square of the gravitational force* $\mathfrak{F}$ in the *stationary field*.

The energy flux has the components:[8]

$$\mathfrak{S}_x = icU_x, \qquad \mathfrak{S}_y = icU_y, \qquad \mathfrak{S}_z = icU_z,$$

with which, from (10a), follows:

$$\mathfrak{S} = -\frac{1}{4\pi\gamma}\frac{\partial\Phi}{\partial t}\mathrm{grad}\Phi = \frac{1}{4\pi\gamma}\frac{\partial\Phi}{\partial t}\cdot\mathfrak{F}, \tag{14}$$

i.e. the energy flux has the direction of the gravitational force; its magnitude is proportional to the product of the magnitude of gravitational force and the temporal increase of the potential.

If one multiplies the last of the eqs. (11) by $-ic$, then it becomes:[8]

$$-ic\nu\mathfrak{F}_u = -\mathrm{div}\mathfrak{S} - \frac{\partial\varepsilon}{\partial t}. \tag{14a}$$

Therefore

$$-ic\nu\mathfrak{F}_u = ic\nu\frac{\partial\Phi}{\partial u} = \nu\frac{\partial\Phi}{\partial l} \tag{14b}$$

yields the energy, which, per unit volume and time, is extracted from the field and is supplied to the matter.

Since, furthermore, as can be seen from (11),[8]

$$\frac{i}{c}X_u, \qquad \frac{i}{c}Y_u, \qquad \frac{i}{c}Z_u,$$

8 Here, the speed of light c is considered as constant, and therefore the influence of the potential on the speed of light discussed above is neglected.

are the components of the momentum density (g) of the gravitational field, it follows from (10a) that:

$$\mathfrak{g} = \frac{1}{c^2}\mathfrak{S} = \frac{1}{4\pi\gamma c^2}\frac{\partial\Phi}{\partial t}\cdot\mathfrak{F}. \tag{15}$$

The symmetry of the ten-tensor implies this relation between the energy flux and the momentum density.[9]

CORRECTION

In lines 8 and 9 [lines 16 and 17 in the original] of my note "On the Theory of Gravitation" an oversight has to be corrected which was brought to my attention by a friendly note from Mr. A. Einstein. Hence one should read there "we consider dx, dy, dz and $du = idl = icdt$ as components of a displacement ds in four-dimensional space." [176]

Hence:

$$ds^2 = dx^2 + dy^2 + dz^2 - c\ dt^2$$

is the square of the four-dimensional line element, where the speed of light c is determined by eq. (6).

9 M. Planck, this journal, 9, 828, 1908, has put forth the claim that also a mechanical energy flux always implies a corresponding momentum.

MAX ABRAHAM

THE FREE FALL

Originally published in Rendiconti del R. Instituto Lombardo di scienze e lettere. German translation by the author published as "Der freie Fall" in Physikalische Zeitschrift 13, 1912, pp. 310–311. Received March 11, 1912. Author's date: Milan, March 1912. English translation taken from the German.

In a recently published communication[1] I have developed a new theory of gravitation. In this theory, the speed of light c is linked to the gravitational potential Φ through the relation:

$$\frac{1}{2}c^2 - \frac{1}{2}c_0{}^2 = \Phi - \Phi_0. \tag{1}$$

The equations of motion of a material point in a gravitational field are:

$$\frac{d^2x}{d\tau^2} = -\frac{\partial\Phi}{\partial x}, \qquad \frac{d^2y}{d\tau^2} = -\frac{\partial\Phi}{\partial y}, \qquad \frac{d^2z}{d\tau^2} = -\frac{\partial\Phi}{\partial z}, \tag{2}$$

$$\frac{d^2u}{d\tau^2} = -\frac{\partial\Phi}{\partial u}; \tag{2a}$$

τ denotes the "*proper time*" of the moving point and is related to the time t measured in the reference frame as follows:

$$\frac{du}{d\tau} = ic\frac{dt}{d\tau} = \frac{ic}{\kappa}, \tag{3}$$

$$\kappa = \sqrt{1 - \frac{v^2}{c^2}} \tag{4}$$

(v is the magnitude of the velocity of the material point).

In the static field, the last (2a) of the equations of motion

1 This journal, 13, 1, 1912.

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 3
Gravitation in the Twilight of Classical Physics: Between Mechanics, Field Theory, and Astronomy.

$$\frac{d}{d\tau}\left(\frac{ic}{\kappa}\right) = \frac{i}{c}\frac{\partial\Phi}{\partial t}$$

assumes the form

$$\frac{c}{\kappa} = \text{const}; \tag{5}$$

it corresponds to the law of conservation of energy.

In the following, the *free fall in empty space* shall be treated on the basis of this theory. The gravitational field is assumed to be homogeneous and parallel to the x- axis. Therefore

$$-\frac{\partial\Phi}{\partial x} = g, \qquad -\frac{\partial\Phi}{\partial y} = -\frac{\partial\Phi}{\partial z} = 0.$$

Then, eqs. (2) yield:

$$\frac{d^2x}{d\tau^2} = g, \qquad \frac{d^2y}{d\tau^2} = 0, \qquad \frac{d^2z}{d\tau^2} = 0. \tag{6}$$

These equations of motion differ from those of Galilean mechanics merely by the proper time τ of the falling point taking the place of t. From these it follows immediately that *the trajectory of the point is a parabola also in the new theory.*

We want to restrict ourselves to the consideration of motion parallel to the x- axis, i.e. along the field, and, in particular, with the following initial conditions:

$$\tau = 0: \qquad x = 0, \qquad \frac{dx}{d\tau} = 0. \tag{7}$$

The gravitational potential may be referred to the plane $(x = 0)$:

$$\Phi_0 = 0, \qquad \Phi = -gx;$$

then, according to (1), one has to set:

$$c^2 = c_0^2 - 2gx; \tag{8}$$

where c_0 denotes the speed of light for $x = 0$. Since we assume the initial speed of the falling point to be equal to zero, initially, we have, according to (4), $\kappa = 1$. Therefore eq. (5) reads:

$$\frac{c}{\kappa} = c_0. \tag{9}$$

From this and from (3) follows the relation

$$\frac{dt}{d\tau} = \frac{1}{\kappa} = \frac{c_0}{c}, \tag{10}$$

which connects the time t with the proper time τ; as a consequence of (8) it can be written as:

$$\frac{dt}{d\tau} = \frac{1}{\sqrt{1-\frac{2gx}{c_0^2}}}. \tag{11}$$

By integrating the first of the equations of motion (6), subject to the initial conditions (7), we obtain

$$\frac{dx}{d\tau} = g\tau, \tag{12}$$

$$x = \frac{1}{2}g\tau^2, \tag{13}$$

i.e. the free fall velocity is proportional to the proper time, the free fall distance to half its square.

The task is now to replace the proper time τ by introducing the time t measured in the coordinate system. In order to determine t as a function of τ, we use eq. (11); taking into account (13) yields

$$\frac{dt}{d\tau} = \frac{1}{\sqrt{1-\left(\frac{g\tau}{c_0}\right)^2}}, \tag{14}$$

and thus

$$t = \int_0^\tau \frac{d\tau}{\sqrt{1-\left(\frac{g\tau}{c_0}\right)^2}} = \frac{c_0}{g}\text{arc }\sin\left(\frac{g\tau}{c_0}\right). \tag{15}$$

By taking the inverse of this functional relation we express τ through t: |

$$\tau = \frac{c_0}{g}\sin\left(\frac{gt}{c_0}\right). \tag{16}$$ [311]

Therefore it follows that

$$\frac{d\tau}{dt} = \cos\left(\frac{gt}{c_0}\right). \tag{16a}$$

Thus, the *free fall velocity* becomes:

$$v = \frac{dx}{dt} = \frac{dx}{d\tau}\frac{d\tau}{dt} = g\tau\frac{d\tau}{dt} = c_0\sin\left(\frac{gt}{c_0}\right)\cos\left(\frac{gt}{c_0}\right). \tag{17}$$

On the other hand, since according to (10):

$$c = c_0 \frac{d\tau}{dt} = c_0 \cos\left(\frac{gt}{c_0}\right), \tag{17a}$$

we then have:

$$\frac{v}{c} = \sin\left(\frac{gt}{c_0}\right) \tag{18}$$

for the quotient of the velocity of fall and of the speed of light at the point in question.

Finally, the *distance of fall* follows from (13) and (16):

$$x = \frac{c_0^2}{2g} \sin^2\left(\frac{gt}{c_0}\right). \tag{19}$$

It is understood that all of these relations are valid only for

$$gt < c_0 \cdot \frac{\pi}{2} \tag{20}$$

Indeed, for $gt = c_0 \cdot \frac{\pi}{2}$, one would have

$$gx = \frac{c_0^2}{2},$$

and therefore

$$\Phi - \Phi_0 = -gx = -\frac{c_0^2}{2}.$$

However, for this value of x, the speed of light c would, according to (1), become zero which is inadmissible.

It is obvious that, by imposing the condition

$$\Phi_0 - \Phi < \frac{c_0^2}{2} \tag{21}$$

on the gravitational potential, the fundamental relation (1) excludes the existence of a homogeneous gravitational field of infinite extent

From the condition (21) one can derive similar consequences of the new theory of gravitation which are of interest for cosmogony.

The potential of a star of mass m is

$$\Phi = -\gamma \frac{m}{r} \tag{22}$$

(γ gravitational constant). Thus, the potential difference between infinite distance $(r = \infty)$, and the surface of the star $(r = a)$ is:

$$\Phi_0 - \Phi = \gamma \frac{m}{a}.$$

Therefore,

$$\gamma \frac{m}{a} < \frac{c_0^2}{2} \left(= \frac{9}{2} \cdot 10^{20} \right) \tag{23}$$

must hold. For the Sun one has in *cgs* units

$$\left(\gamma \cdot \frac{m}{a} \right)_s = 2 \cdot 10^{15}.$$

Thus, the *quotient of mass and radius for an arbitrary star must satisfy the inequality*

$$\frac{m}{a} < 2,25 \cdot 10^5 \cdot \left(\frac{m}{a} \right)_s. \tag{24}$$

For stars whose mean density is equal to that of the Sun, the quotients m/a are proportional to the squares of the radii. Hence, the following must hold

$$a < 500(a)_s, \tag{24a}$$

and thus

$$m < 10^8 (m)_s. \tag{24b}$$

That is, *the mass of a star whose mean density is equal to that of the Sun cannot become larger than one hundred million times the mass of the Sun*.

Since this limit is rather high, no difficulties arise from this for our theory.

MAX ABRAHAM

A NEW THEORY OF GRAVITATION

Lecture presented on October 19, 1912 to the Societa italiana per il progresso delle scienze. German translation by the author published as "Eine neue Gravitationstheorie" in Archiv der Mathematik und Physik. Third series 20, 1913, pp. 193–209. English translation taken from the German.

Modern physics does not allow forces that propagate with an infinite speed. It does not hold that Newton's law is the true fundamental law of gravitation; rather, it endeavors to obtain this action-at-a-distance law from differential equations attributing a finite speed of propagation to the gravitational force.

A model of such a theory of local action is provided to us by Maxwell's theory of the electromagnetic field. Its fundamental laws are differential equations connecting the electric vector to the magnetic vector. The electromagnetic energy is, according to this theory, distributed throughout the field. When the field changes in time, an energy flow determined by the Poynting vector obtains. If, for example, an electron starts to oscillate, it sends out electromagnetic waves. With the waves the electromagnetic energy flows from the neighborhood of the electron to the initially undisturbed regions of space; this radiation of energy results in the damping of the oscillations of the electron.

The analogy between Coulomb's and Newton's law suggests a similar interpretation of the gravitational force. It was Maxwell himself who developed a theory of gravitation, modelled on electrostatics, and emphasized the difficulties of such a theory.[1] The different signs of the forces—attraction of masses compared to repulsion of charges of equal sign—entails that the energy density of the gravitational field becomes negative as soon as one assumes that when the field vanishes so does the energy. One would need to drop this last assumption and imagine | that if there were no [194] gravitational field, a certain amount of energy resides in the aether, which decreases upon excitation of the field. But even so, one still does not avoid all objections.

Let us consider for example a material particle which is initially at rest and then is set into oscillation. In theories of the electromagnetic type it emits waves similar to light waves, i.e. transverse waves propagating with the speed of light. With the waves, the gravitational field enters into previously undisturbed regions of space. Hence, the

1 J. Cl. Maxwell, *Scientific papers* I, p. 570.

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 3
Gravitation in the Twilight of Classical Physics: Between Mechanics, Field Theory, and Astronomy.

energy density of the aether decreases in these regions; i.e. the energy flows towards the oscillating particle whose energy increases at the expense of the aether energy. This influx of energy results in an increase of the oscillation of the particle; its equilibrium is therefore unstable.

Similar difficulties arise in this way for all gravitational theories of Maxwellian type; among these, especially the theory of H. A. Lorentz[2] and R. Gans[3] has found followers among physicists. Starting from the hypothesis that matter consists of positive and negative electrons and that the attraction between electrons of opposite charge is slightly greater than the repulsion between those of like charge, it arrives at field equations of Maxwellian character connecting the gravitational vector, which corresponds to the electric vector, to a second one analogous to the magnetic vector. In this theory, the energy flux of the gravitational field is expressed by a vector corresponding to the Poynting vector but with opposite sign.[4] Here, the above mentioned difficulty appears; indeed, R. Gans found that the radiation reaction force has the opposite sign as in the dynamics of the electron. Thus, as a result of the emitted radiation, the acceleration of a neutral particle would increase rather than decrease as for electrons. Hence, the equilibrium of the particle would not be stable.

Therefore, we have to dispense with the close analogy between gravitation and electromagnetism without thereby relinquishing the essential notions of Maxwell's theory, namely: *The fundamental laws must be differential equations describing the excitation and propagation of the gravitational field; associated with this field is a positive energy density as well as an energy flux.* |

[195] The problem of gravitation is even more urgent, as modern physics has discovered interesting relations between mass (m) and energy (E). According to the theory of relativity one should have

$$E = mc^2 \quad (c \text{ is the speed of light in the vacuum}). \tag{1}$$

Since we still lack a satisfactory theory of gravitation, one has been able to derive this relation only for the inertial mass. Is it valid also for the gravitational mass?

We displace a body in a gravitational field. Its potential energy, and consequently the first term in equation (1), changes with the gravitational potential. It follows that one of the factors on the right-hand side, or even both, must depend on the gravitational potential. We want to consider the hypothesis that c, *the speed of light, depends on the gravitational potential*. This hypothesis was first enunciated by A. Einstein.[5] Taking it as a starting point, I undertook to develop a theory of the gravitational field[6] which I then, in constructive competition with Mr. A. Einstein,[7] gave a

2 H. A. Lorentz, *Verlag. Akad. v. Wetensch. te Amsterdam* 8, 1900, p. 603.
3 R. Gans. *Physik. Zeitschrift* 1905, p. 803.
4 R. Gans, *H. Weber-Festschrift* 1912, p. 75.
5 A. Einstein, *Ann. d. Physik* 35, (1911), p. 898.
6 M. Abraham, *Physik. Zeitschrift* 1912, p. 1 and p. 311.
7 A. Einstein, *Ann. d. Physik* 38 (1912), p. 355 and p. 433.

more satisfactory form.[8] In the following, I want to present the essential features of the new theory of gravitation.

Postulate I. The surfaces c = constant *coincide with the equipotential surfaces of the gravitational field*. Or, in other words: *The negative gradient of* c *gives the direction of the gravitational force*.

If the speed of light varies in the gravitational field, then, according to Huygens' principle, a light ray in this field will be refracted as in an inhomogeneous medium. This consequence was derived by A. Einstein; he showed that a light ray passing the surface of the Sun must be deflected, in fact, just as if it were attracted by the Sun. However, this deflection, only observable at a total eclipse of the Sun, lies at the limit of observability.

For bodies at rest in the gravitational field we apply the usual geometry; hence, we assume that the unit of length, the meter, is independent of c and can thus serve to measure length in arbitrary regions of the gravitational field. We now want to consider two regions in which c may have different values, c_1 and c_2. We bring an antenna having a length of one meter from the | first region into the second. Since the [196]
length of the antenna remains constant, obviously the periods of its electromagnetic normal modes change in inverse proportion to the speed of light c:

$$\tau_1 : \tau_2 = c_2 : c_1. \tag{2}$$

If one could construct a clock whose rate were independent of c, then one could determine the change in period by transporting the clock, with the antenna, from one region into the other. However, we rule out the possibility of such a clock by putting forward the following postulate:

Postulate II. An observer belonging to a material system is unable to perceive that he, together with the system, is brought into a region in which c *has a different value*.

From this postulate,[9] it follows that the duration of an arbitrary process changes in the same proportion to c as does the period of the normal modes of the antenna; because otherwise the observer would be able to determine the latter change. Hence, from the second postulate follows the general theorem: The duration of an arbitrary process in a system changes in inverse proportion to c if the system's location in a gravitational field changes. Or more briefly:

"The times are of degree c^{-1}*."*

Within this theory the unit of time has only a local significance, whereas a universal validity will be attributed to the unit of length, at least for the state of rest.

---

8 M. Abraham, *Physik. Zeitschrift* 1912, p. 793.

9 It is understood that this postulate refers only to the value of c itself, and not to its derivative; the value of c has no influence on the events in the system which present themselves to an observer belonging to that system. The gradient of c, however, i.e. the gravitational force, of course influences these events.

According to postulate II, there exists a certain relativity. Let us imagine that the Earth reaches locations in space in which the gravitational potential, and therefore also c, has a different value. According to the latter postulate it would not be possible for us to determine this fact through any terrestrial measurement. All measurements, and hence also all the constants of physics, would remain unchanged; also the measurement of the speed of light, e.g. according to the method of Fizeau, would produce the same result as before since with respect to a terrestrial clock its changes are compensated for by the changes of the angular velocity of the gear wheel. Obviously, all
[197] velocities change generally in proportion to the light | velocities, i.e. they are of degree c, because the lengths are of degree c^0 and the times of degree c^{-1}.

It is, however, in no way ruled out that an observer not belonging to the system discovers this particular influence of the gravitational potential on the periods. For example, a terrestrial observer measuring the Fraunhofer lines of the Sun and comparing them to the corresponding lines of terrestrial sources, should find that their frequencies behave as

$$\nu_2 : \nu_1 = \tau_1 : \tau_2 = c_2 : c_1 .$$

From this would follow a relative shift of the Sun's lines:

$$\frac{\nu_2 - \nu_1}{\nu_1} = \frac{c_2 - c_1}{c_1},$$

namely towards the red end of the spectrum, because the gravitational potential on the Sun, and hence also c, has a smaller value than on Earth. For this relative displacement of the Sun's lines Einstein found the value $2 \cdot 10^{-6}$, which according to Doppler's principle would correspond to a speed of 0.6 kilometers per second. Astrophysicists are now well able to measure shifts of this order, and indeed they have found such shifts, and precisely in the sense required by our theory. However, they interpret the shift partially as the Doppler effect of descending flows of absorbing gases, and partially as pressure effects. But perhaps the totality of the phenomena on the surface of the Sun can be better interpreted if one takes the predicted gravitational shifts of the lines into account.

We now turn from kinematics to dynamics. The second postulate implies that *all mechanical quantities of the same class are of the same degree in* c. For example, a system of forces which maintain themselves in equilibrium in a region, where c has the value c_1, must still do so when the system is brought into a region of the field in which c is equal to c_2. Therefore, all forces must change in the same proportion with c. All forms of energy must be of the same degree in c as well, because if two forms of energy, e.g. the kinetic and the potential, were of different degree, the conversion of energy from one form into the other would give rise to a periodic process whose frequency is not of the correct degree, i.e. that of c. Similar considerations apply for other dynamical quantities. Since all these quantities are constituted from mass,
[198] length and time, it is sufficient to know the *degree of the mass* to determine all | their degrees since we already know the degree of the lengths (c^0) and of the times (c^{-1}).

I raised the question earlier whether, as for the inertial mass, the gravitational mass is proportional to the energy. What would take place during the transformation of radioactive elements if their continual loss of energy as heat would cause a decrease in their inertial mass, but not in their weight? Obviously, uranium and its daughter element radium suspended from strings of equal lengths would result in pendulums with unequal periods of oscillations. Even the equilibrium position of such pendulums would be different since the attraction of the Earth acts on the gravitational, whereas the centrifugal force acts on the inertial mass. This contradicts experience. Either one must give up any relation between mass and energy, or assume that the weight too, like the inertia, is proportional to energy. The first alternative would mean the collapse of the new mechanics. We prefer the second and hence propose the following postulate.

Postulate III. The forces which are acting on two bodies at the same location of the gravitational field are in proportion to their energy.

Here, one has to take into consideration not only the potential and kinetic energy of the molar and molecular motion, but also the chemical and electromagnetic energy. For example, the electrons in a metal carry electromagnetic energy; hence they are subject to gravitation.[10] Similarly, the thermal radiation in the interior of a cavity will also acquire weight. If one interprets the third postulate in this manner, then *the laws of the conservation of energy and of the conservation of weight merge into one*.

Proceeding with the mathematical presentation, we start with the expression for the *Lagrangian function*, which for the dynamics of the electron is:

$$L = -mc^2 f\left(\frac{v}{c}\right); \tag{3}$$

v denotes the velocity, and m the rest mass of the electron.[11] From | the *Lagrangian* [199]
function we derive in the well known manner the values for the momentum and energy as follows:

$$G = \frac{\partial L}{\partial v}, \tag{3a}$$

$$E = v\frac{\partial L}{\partial v} - L, \tag{3b}$$

while the *equations of motion* are:

---

10 This is also supported by comments of J. Koenigsberger (*Verh. d. D. physik. Ges. XIV* (1912), p. 185).

11 In order for m to have this meaning, we must have $f''(0) = -f(0) = -1$ (this holds for example in the theory of relativity, where $f(v/c) = \sqrt{1-(v/c)^2}$); or else, an insignificant numerical factor becomes associated with m in equation (3).

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}}\right) - \frac{\partial L}{\partial x} = 0 \text{ etc.} \tag{4}$$

The new mechanics has, in the discussion of these equations, so far restricted itself to the case of constant c; in this case the Lagrangian function depends only on the velocity but not on the location. Hence, only the first terms containing the time derivatives of the momentum components enter into the equations of motion:

$$\frac{d}{dt}\left(\frac{\partial L}{\partial v} \cdot \frac{\dot{x}}{v}\right) = \frac{d}{dt}\left(G \cdot \frac{\dot{x}}{v}\right) = \frac{d\mathfrak{G}_x}{dt} \text{ etc.} \tag{4a}$$

In our theory of gravitation however, through the speed of light c, the Lagrangian function also depends on the coordinates; hence one must retain the second terms in Lagrange's equations:

$$\frac{\partial L}{\partial x} = \frac{\partial L}{\partial c}\frac{\partial c}{\partial x} \text{ etc.} \tag{4b}$$

They represent the components of a force proportional to the gradient of c. According to postulate I, *this force* is now precisely *the gravitational force*. In vector form, the equations of motion (4) are:

$$\frac{d\mathfrak{G}}{dt} = \frac{\partial L}{\partial c} \operatorname{grad} c. \tag{5}$$

One recognizes that the Lagrangian equations are nothing else but the analytic expression of our first postulate. (They hold exactly for the free motion of material points in the gravitational field, but can also be applied to systems whose extensions are so small that they can be considered equivalent to a material point.) Especially now, while the world of mathematics prepares for the centennial celebration of the great founder of analytical mechanics, we do not want to omit drawing attention to the significance of his work for the new mechanics.

Now we want to introduce the third postulate, which for a given location in the [200] field, sets the gravitational force proportional to the energy of the | moving point; in order to fulfill it we must write:

$$\frac{\partial L}{\partial c} = -\chi \cdot E, \tag{6}$$

where χ depends only on c, but not on v. Then the second term of (5), i.e. the gravitational force, becomes:

$$\mathfrak{K} = -\chi(c) \cdot E \cdot \operatorname{grad} c. \tag{6a}$$

Setting

$$\log \varphi(c) = \int^{c} dc \chi(c), \tag{6b}$$

it follows from (6) that

$$E = -\frac{1}{\chi(c)}\frac{\partial L}{\partial c} = -\varphi\frac{\partial L}{\partial \varphi}. \tag{6c}$$

This and (3b) yields

$$L = v\frac{\partial L}{\partial v} + \varphi\frac{\partial L}{\partial \varphi}. \tag{7}$$

On the basis of a well known theorem of Euler we infer: the Lagrangian function is a homogeneous linear function of v and φ. We wish to write it in the form:

$$L = -M\varphi \cdot f\left(\frac{v}{\varphi}\right). \tag{7a}$$

M denotes a constant (mass constant) belonging to the material mass point in question that is independent of φ and hence of c. The comparison with the expression (3) yields:

$$\varphi = c, \quad M = mc. \tag{7b}$$

From this follows the value for the rest mass m:

$$m = \frac{M}{c}. \tag{8}$$

Therefore, the rest mass is of degree c^{-1}.

Since we now know the degree of the lengths (c^0), times (c^{-1}) and masses (c^{-1}), we are in a position to derive the degree of each class of dynamical quantities. For example, *energies are of degree* c, *likewise forces are of degree* c, *actions* (having dimension of energy times time) *of degree* c^0.

From (6b) and (7b) it follows that

$$\chi(c) = \frac{1}{c},$$

so that the expression (6a) for the *gravitational force* becomes:

$$\mathfrak{K} = -\frac{E}{c}\operatorname{grad}c. \tag{9}$$

| As one sees, the first postulate is satisfied *because the force is proportional to the negative gradient of* c, as well as the third, because it is *proportional to the energy.* [201]

A potential energy is associated with a material point resting in a gravitational field; one obtains it from (3b) and (3) by setting v equal to zero:

$$E = -L = +mc^2;$$

therefore, according to (8):

$$E = M \cdot c. \tag{9a}$$

From this expression for the *energy* of a *resting point* follows, according to (9), the value for the *gravitational force* for the case of rest:

$$\mathfrak{K} = -M \operatorname{grad} c. \tag{9b}$$

The work done by this force is, as it must be, equal to the decrease of the potential energy (9a).

We are now in a position to determine for a given gravitational field, i.e. for a given field of the scalar c, the force acting on a material point. But, we still have not solved the problem initially posed, to find the gravitational field which corresponds to a given distribution of matter. Now, the interrelation between action and reaction suggests we assume that like the attracted, so also the *attracting mass is proportional to the energy,* and hence that we consider the energy as the source of the gravitational field. However, the question arises: Besides the energy of the matter, is the energy of the gravitational field itself to be taken into account? If we knew the energy of the field, then, through the application of the principle of virtual work, without further ado, at least the statics of the gravitational field can be derived. Consequently, we prefer to start from reasonable assumptions [*Ansätze*] for the field energy and the energy flux, and to obtain from these the relations connecting the gravitational vector with the density of matter and energy, respectively.

The simplest assumption would be that the energy density of the static field is proportional to the square of the gradient of c, on which, according to (9b), the field strength depends. However, as we have seen, the energy and hence also the density is of degree c, while the square of the gradient of c is obviously of degree c^2. This consideration leads to the introduction of the auxiliary variable

$$u = \sqrt{c} \tag{10}$$

[202] | and to assign to the *energy density of the gravitational field* the value:

$$\varepsilon = \frac{1}{2\alpha}\left\{\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2 + \left(\frac{\partial u}{\partial z}\right)^2 + \left(\frac{1}{c}\frac{\partial u}{\partial t}\right)^2\right\} \tag{11}$$

(α denotes a universal constant of degree c^0).

In our theory, this expression is valid also for the dynamical field; it is supplemented by the ansatz for the *energy flux in the gravitational field*

$$\mathfrak{S} = -\frac{1}{\alpha}\frac{\partial u}{\partial t} \cdot \operatorname{grad} u. \tag{12}$$

Since according to (11) the field energy is always positive, and vanishes only when the field vanishes, one sees that the energy always flows with the wave by which the disturbance is propagated. Let us imagine for example that such a disturbance, for which $u < u_0$, enters a region in which initially $u = u_0$. Then, the gradient is pointing from u towards the undisturbed region, whereas, during the passing of the wave, the time derivative of u has a negative sign. Hence, the expression (12) for the energy flux has the correct sign. It retains it when $u > u_0$ for the disturbance, because then the gradient as well as the time derivative change sign. *In the theory presented here, the energy flux always corresponds to an emission of energy from the disturbed region*; thus the above mentioned objection is not raised against it. The radiation reaction always implies a decrease of the acceleration of the material particle; its *equilibrium* is hence *not unstable*.

We now want to consider a field containing matter at rest. As we are dealing with a continuous distribution of matter over a volume V, we set

$$M = \int \mu dV, \tag{13}$$

and call

$$\mu = \lim_{V=0}\left(\frac{M}{V}\right)$$

the "*specific density*" of the matter. Since M does not depend on c, then for an incompressible fluid, whose particles do not change in volume, μ too is independent of the location in the gravitational field.

The rest energy of matter with respect to the unit volume is according to (9a),

$$\eta = \lim_{V=0}\left(\frac{Mc}{V}\right) = \mu c$$

or, according to (10)

$$\eta = \mu u^2. \tag{14}$$

| Since μ is constant for the case of rest, from this follows: [203]

$$\frac{\partial \eta}{\partial t} = 2\mu u \frac{\partial u}{\partial t} \tag{14a}$$

as the temporal increase of the rest energy per unit volume, caused by a temporal variation of the gravitational field.

Now we apply the energy equation which demands that the convergence of the energy flux is equal to the sum of the increases of the energy densities of the field (ε) and of the matter (η):

$$-\mathrm{div}\mathfrak{S} = \frac{\partial \varepsilon}{\partial t} + \frac{\partial \eta}{\partial t}. \tag{15}$$

From the expressions (11) and (12) we derive

$$\frac{\partial \varepsilon}{\partial t} = \frac{1}{\alpha}\left\{\frac{\partial u}{\partial x}\frac{\partial^2 u}{\partial x \partial t} + \frac{\partial u}{\partial y}\frac{\partial^2 u}{\partial y \partial t} + \frac{\partial u}{\partial z}\frac{\partial^2 u}{\partial z \partial t} + \frac{1}{c}\frac{\partial u}{\partial t}\frac{\partial}{\partial t}\left(\frac{1}{c}\frac{\partial u}{\partial t}\right)\right\}, \tag{15a}$$

$$-\mathrm{div}\mathfrak{S} = \frac{1}{\alpha}\left\{\frac{\partial^2 u}{\partial x \partial t}\frac{\partial u}{\partial x} + \frac{\partial^2 u}{\partial y \partial t}\frac{\partial u}{\partial y} + \frac{\partial^2 u}{\partial z \partial t}\frac{\partial u}{\partial z} + \frac{\partial u}{\partial t}\Delta u\right\}, \tag{15b}$$

with

$$\Delta u = \mathrm{div}\,\mathrm{grad} u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}. \tag{15c}$$

If, in addition, we set

$$\square u = \Delta u - \frac{1}{c}\frac{\partial}{\partial t}\left(\frac{1}{c}\frac{\partial u}{\partial t}\right), \tag{16}$$

then from (15), taking into account (15a, b) and (14a), it follows that

$$\square u = 2\alpha\mu u. \tag{17}$$

Here, the second term relates to matter at rest. According to the previous discussion, for the case of motion one has to introduce its energy density, and replace (17) by the more general formulation:

$$\square u = 2\alpha\frac{\eta}{u}. \tag{17a}$$

This is the fundamental equation relating the gravitational field to the energy of the matter. It can function as the analytic expression of the fourth postulate of our theory, which relates the attracting mass of a body to its energy. For a static field, according to (16), the fundamental equation takes the form:

$$\Delta u \equiv \mathrm{div}\,\mathrm{grad} u = 2\alpha\frac{\eta}{u} = 2\alpha\mu u. \tag{17b}$$

Therefore, in this case the *divergence of the gradient of* $u = \sqrt{c}$ *is proportional to the energy density of matter.*

Now, what about the energy of the gravitational field itself? At least for the static
[204] field (17b), it is easy to put the basic equation | into a form in which the energy density of the field appears on the right-hand side. One has

$$\frac{\partial^2 c}{\partial x^2} = \frac{\partial^2 u^2}{\partial x^2} = 2u\frac{\partial^2 u}{\partial x^2} + 2\left(\frac{\partial u}{\partial x}\right)^2 \text{ etc.,}$$

thus,

$$\Delta c = 2u\Delta u + 2(\mathrm{grad} u)^2;$$

according to (11), the energy density of the static field is now:

$$2\alpha\varepsilon = (\mathrm{grad}\,u)^2.$$

Therefore, we can write (17b) as

$$\Delta c \equiv \mathrm{div}\,\mathrm{grad}\,c = 4\alpha(\eta + \varepsilon). \tag{18}$$

Hence, in the static field, the divergence of the gravitational vector, i.e. of the gradient of c, is proportional to the density of the total energy. Integrating (18) over a volume V bounded by the surface f, we obtain

$$\int df \frac{\partial c}{\partial n} = 4\alpha \int dV(\eta + \varepsilon) = 4\alpha E, \tag{19}$$

i.e. in the static gravitational field, *the flux of the gravitational vector through a closed surface is proportional to the enclosed energy.*

To highlight the significance of these statements, we consider a special case: *a sphere at rest* (e.g. the Sun), which is composed of homogeneous concentric layers. The basic equation (17b) yields:

$$\Delta u = 2\alpha\mu u \text{ for } r < a, \tag{20}$$

i.e. for the interior of the sphere,

$$\Delta u = 0 \text{ for } r > a, \tag{20a}$$

i.e. for the exterior of the sphere.

This last equation is Laplace's differential equation; its symmetric integral

$$u = u_0\left(1 - \frac{\vartheta}{r}\right) \qquad (r > a) \tag{21}$$

determines the gravitational field outside the attracting sphere (where u_0 denotes the value of $u = \sqrt{c}$ for $r^{-1} = 0$, ϑ a different constant).
From (21) follows

$$\frac{du}{dr} = \frac{u_0 \cdot \vartheta}{r^2}. \tag{21a}$$

Thus, the radial gradient of c becomes:

$$\frac{dc}{dr} = 2u\frac{du}{dr} = \frac{2c_0\vartheta}{r^2}\left(1 - \frac{\vartheta}{r}\right). \tag{22}$$

| Now, according to (9b) the gravitational force is proportional to this gradient. *It follows from this that Newton's law is not strictly valid in our theory. The center of the Sun attracts the planet* (considered as a material point) *with a force which contains besides the term with r^{-2} also a term with r^{-3}.* [205]

According to the theorem (eq. (19)) just developed, the flux of the gradient of c passing through a sphere of radius r is proportional to the total energy enclosed by the sphere:

$$4\pi r^2 \frac{dc}{dr} = 4\alpha E. \tag{23}$$

A sphere of infinite radius encloses the total energy (E_t) of the sphere and of its gravitational field; hence, from (22) and (23) it follows

$$E_t = \frac{2\pi c_0}{\alpha} \cdot \vartheta. \tag{23a}$$

In contrast, the sphere of radius $r = a$ contains only the internal energy of the sphere; therefore,

$$E_i = \frac{2\pi c_0}{\alpha} \cdot \vartheta\left(1 - \frac{\vartheta}{a}\right). \tag{23b}$$

Hence, the energy of the external field has the value

$$E_a = E_t - E_i = \frac{\vartheta}{a} E_t. \tag{23c}$$

We write

$$\psi = \frac{\vartheta}{a} = \frac{E_a}{E_t}. \tag{24}$$

This quantity ψ, i.e. the ratio of the external to the total energy, enters into expression (22) for the value of the gravitational force outside of the sphere:

$$\frac{dc}{dr} = \frac{2c_0 a\psi}{r^2}\left(1 - \psi\frac{a}{r}\right) = \frac{\alpha E_t}{\pi r^2}\left(1 - \frac{E_a}{E_t} \cdot \frac{a}{r}\right). \tag{24a}$$

We can therefore say: *The deviation from Newton's law is caused by the energy of the external gravitational field.*

To determine the quotient ψ, one must integrate equation (20), which is valid for the interior of the sphere. We want to compare this differential equation with Poisson's; in Poisson's equation, the right-hand side depends only on the density of the attracting masses, whereas on the right of (20), the potential (u) itself enters as a factor. This implies that in our theory the contributions of the individual mass elements do not superimpose; rather, along with the value of u at a specific location, the con-
[206] tribution of the | masses at the same location decreases as well. Since now the neighboring masses cause a decrease of the potential, it is apparent that large accumulations of matter produce here a smaller attraction than they would according to the usual theory. However, as we will see, the difference is, even for the Sun, still not noticeable.

Taking into account the symmetry of the field, the differential equation (20) yields:

$$\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{du}{dr}\right) = 2\alpha\mu u,$$

or

$$\frac{d^2(ru)}{dr^2} = 2\alpha\mu \cdot ur \qquad (\text{for } r \le a). \tag{25}$$

To integrate the equation, one must know the distribution of the specific density μ along a radius. We want to restrict ourselves to the specific case of a sphere consisting of an incompressible fluid (μ = constant). By setting

$$\kappa^2 = 2\alpha\mu, \tag{25a}$$

we find u expressed through the hyperbolic sine

$$u = \frac{A}{r} \cdot \sinh\Phi(\kappa r). \tag{26}$$

The integral of (25) remains finite for $r = 0$. The constant A is determined by the condition that for $r = a$, the values of u and of

$$\frac{du}{dr} = A\left\{\frac{\kappa}{r}\cosh(\kappa r) - \frac{1}{r^2}\sinh(\kappa r)\right\} \tag{26a}$$

agree with those valid outside the sphere, which are given by (21) and (21a) respectively:

$$\left.\begin{aligned} \frac{A}{a}\sinh(\kappa a) &= u_0\left(1 - \frac{\vartheta}{a}\right), \\ \frac{A}{a}\{\kappa a\cosh(\kappa a) - \sinh(\kappa a)\} &= \frac{u_0\vartheta}{a}. \end{aligned}\right\} \tag{26b}$$

In addition, these two equations still determine the value of the quotient ψ (cf. 24); only this is of interest to us here. We find

$$\frac{\psi}{1-\psi} = \frac{\kappa a\cosh(\kappa a) - \sinh(\kappa a)}{\sinh(\kappa a)} = \frac{\kappa a - \tanh(\kappa a)}{\tanh(\kappa a)} = \xi$$

and therefore,

$$\psi = \frac{\xi}{1+\xi} = 1 - \frac{\tanh(\kappa a)}{\kappa a}. \tag{27}$$

The gravitational force (24a) which will be exerted by the incompressible sphere on an external point depends on this quantity, which determines the ratio of the external to the total energy of the sphere. |

[207] We first want to evaluate ψ for the realistic case:

I. κa *small*: One has

$$\frac{\tanh(\kappa a)}{\kappa a} = \frac{1 + \frac{(\kappa a)^2}{3!} + \cdots}{1 + \frac{(\kappa a)^2}{2!} + \cdots} = 1 - \frac{(\kappa a)^2}{3},$$

and thus

$$\psi = \frac{\kappa^2 a^2}{3} = \frac{2\alpha\mu a^2}{3} = \frac{\alpha M}{2\pi a} \qquad \left(M = \mu \cdot \frac{4\pi a^3}{3}\right). \tag{27a}$$

Since in this case ψ is very small, then so also is the proportion of the external field energy to the total energy of the sphere; ignoring it, we find from (24a) and (27a) the force acting on a resting material point P':

$$M'\frac{dc}{dr} = \frac{\alpha c_0}{\pi}\frac{MM'}{r^2}.$$

Going over to the usual units, we set

$$M = cm, \quad M' = c'm'$$

and neglect the difference between c, c' and c_0. The force then becomes

$$\frac{\alpha}{\pi}c_0^3\frac{mm'}{r^2} = \gamma\frac{mm'}{r^2},$$

where γ is the usual gravitational constant. Therefore, one has

$$\alpha = \frac{\pi\gamma}{c_0^3}$$

and hence, according to (27a):

$$\psi = \frac{\gamma m}{2ac_0^2}, \tag{27b}$$

which for the Sun gives the value:

$$\psi = 10^{-6}.$$

This is thus, indeed, still the limiting case of a small ψ, or rather, κa. For the purposes of astronomy, the deviation from Newton's law due to the external field energy is to be neglected, and the Sun is to be replaced by a mass point.[12] |

II. Although it is apparent from what has been said that the *converse limiting case* [208] does not correspond to reality, it shall be briefly considered:

$$\kappa a \text{ large } \tanh(\kappa a) = 1 .$$

Here, according to (27)

$$\psi = \frac{E_a}{E_t} = 1 - \frac{1}{\kappa a} \tag{27c}$$

becomes only slightly less than one, i.e. nearly the entire energy resides in the external gravitational field. In the interior resides only the small fraction

$$1 - \psi = \frac{E_i}{E_t} = \frac{1}{\kappa a}.$$

For the gravitational force, from (24a) it follows that

$$\frac{dc}{dr} = \frac{2c_0 a}{r^2}\left(1 - \psi\frac{a}{r}\right).$$

Thus, at large distances the gravitational force is not proportional to the volume of the sphere, but to its radius. Here, the above mentioned screening effect of large accumulations of masses becomes apparent (i.e., of a very large value of the radius a or of the density μ).

We want to return to the basic equation (17a):

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} - \frac{1}{c}\frac{\partial}{\partial t}\left(\frac{1}{c}\frac{\partial u}{\partial t}\right) = 2\alpha\frac{\eta}{u}.$$

From it one derives the disturbance in the gravitational field caused by the motion of matter. Outside the matter, it indicates a propagation of the disturbance with the speed of light (c). However, a rigorous treatment of the problem of propagation is made more difficult because the disturbance of the field itself influences the value of u, and thereby the value of the speed of propagation, c. The same difficulty appears as well with the propagation of sound, whose speed depends on pressure, and thereby varies upon the passing of the sound wave. But in all practical cases, the variation of u is so minute that one can consider the speed of propagation of gravitation to be constant.

12 With reference to the principle of action and reaction, one also obtains from this a clue as to the range of the domain of validity of the above considerations based on Lagrange's equations, wherein the object was replaced by a material point, and where, upon the calculation of the mass, only the self energy, and not the energy of the gravitational field produced by the body, was taken into consideration. As one can now see, this procedure is still permissible for objects of the order of the fixed stars.

According to our theory, light and gravitation have the same speed of propagation; but whereas light waves are transverse, gravitational waves are longitudinal. Incidentally, the problem of the oscillating particle can be treated in a similar manner as that of the oscillating electron; the strength of the emitted gravitational waves depends on the product of the gravitational mass and the acceleration of the particle. Is it possible to detect these gravitational waves? |

[209] This hope is futile. Indeed, to impart an acceleration to one particle, another particle is necessary which, according to the law of action and reaction, is driven in the opposite direction. But now, the strength of the emitted gravitational wave depends on the sum of the products of the gravitational mass and the acceleration of the two particles, while, according to the reaction principle, the sum of the products of inertial mass and acceleration is equal to zero.[13] Therefore, although the existence of gravitational waves is compatible with the assumed field mechanism, through the equality of gravitational and inertial mass the possibility of its production is practically excluded. It follows from this that the planetary system does not lose its mechanical energy through radiation, whereas an analogous system consisting of negative electrons circling around a positive nucleus gradually radiates its energy away. The life of the planetary system is thus not threatened by such a danger.

Our theory of gravitation based on the assumption of a variable c contradicts from the outset the second axiom of the theory of relativity. However, in a vacuum, the invariance with respect to the Lorentz transformations is preserved as is shown by the form of the fundamental equation (17a) which applies there:

$$u\Box u = 0.$$

Hence, outside of matter, the Lorentz group still applies in the infinitesimally small. *It is the matter which breaks the invariance under the Lorentz group*, because in the equation

$$u\Box u = 2\alpha\eta$$

the first term is an invariant of the group, whereas the second term, proportional to the energy density, is not. It is precisely the very plausible hypothesis that the attracting mass is proportional to the energy that forces us to abandon the Lorentz group in the infinitesimally small as well.

Thus, Einstein's relativity theory of 1905 turns to dust. Will there, like the Phoenix rising out of the ashes, emerge a new, more general principle of relativity? Or, will one return to absolute space, and beckon back the much scorned aether, so that it can support, in addition to the electromagnetic, also the gravitational field?

---

13 Here however, the momentum of the gravitational field has not been taken into consideration; but this is of just as little practical importance as the energy of the field.

MAX ABRAHAM

RECENT THEORIES OF GRAVITATION

Originally published as "Neuere Gravitationstheorien" in Jahrbuch der Radioaktivität und Elektronik, 11, 1915, pp. 470–520. Received 15 December 1914. Author's Date: Milan, December 1914.

CONTENTS

I. INTRODUCTION [473]

A. The Vector Theory of Gravitation

Since I. Newton postulated his action-at-a-distance law of attraction of masses, theoretical physics has endeavored to reduce this law to one of local interactions. In this attempt the older theories of gravitation[1] were based on concepts derived from the theory of elasticity, hydrodynamics and the kinetic theory of gases. Electromagnetic theories of the gravitational field appeared only at the end of the nineteenth century, encouraged by Maxwell's local field theory of electrodynamics. The best received theory was probably one developed by H. A. Lorentz[2] which adapted a hypothesis

1 P. Drude, *Wied. Ann*. 62, 1, 1897 and J. Zenneck, *Enzyklopädie d. math. Wissensch*. V, 1, article 2 [in this volume], give overviews of the state of the theory of gravitation at the end of the last century.

2 H. A. Lorentz, *Verslag. Akad. v. Wetenschapen te Amsterdam*, 8, 603, 1900.

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 3
Gravitation in the Twilight of Classical Physics: Between Mechanics, Field Theory, and Astronomy.

already pursued by Aepinus, Mosotti and Zöllner to the framework of electron theory. According to this hypothesis, the attraction of unlike electric charges should be somewhat larger than the repulsion of like ones. For electrically neutral matter, whose atoms are supposed to consist of positive and negative electrons, Newton's law for the attraction of masses at rest follows from Coulomb's law of electrostatic forces. The force ($\mathfrak{E}^g$) acting on a unit of mass of the stationary matter arises thus as the resultant of two electric forces of opposite direction and nearly equal magnitude, act-
[474] ing on the | positive and negative electrons. However, if the matter moves, then a magnetic force is associated with each of these two electric forces, related to them by the electromagnetic field equations. The two magnetic forces act on the convection current of the positive and the negative electrons, respectively. The resultant ($\mathfrak{H}^g$) of the two magnetic vectors, likewise opposite in direction and of nearly equal magnitude, determines a mechanical force acting on the moving matter. All in all, it follows that if v denotes the velocity vector and c the speed of light, then the expression for the gravitational force [*Schwerkraft*] per unit mass is

$$\mathfrak{F}^g = \mathfrak{E}^g + \frac{1}{c}[\mathfrak{v}\mathfrak{H}^g]. \tag{1}$$

This expression corresponds precisely to the one which represents the force ($\mathfrak{F}^e$) per unit charge in the electron theory. The results of the Lorentzian theory of gravitation thus can be summarized as follows: with the electromagnetic vector pair ($\mathfrak{E}, \mathfrak{H}$) is associated a second pair, ($\mathfrak{E}^g, \mathfrak{H}^g$), characterizing the gravitational field, which determines the gravitational force in accordance with (1). This pair is linked to the density and velocity of matter by differential equations that agree with the field equations of the electron theory, except for the signs of the charge and mass density respectively.

We will call a theory that represents the gravitational field by means of two electromagnetically interrelated vectors simply a "*vector theory of gravitation*." As is apparent from eq. (1), in such a theory the force acting on a body at a given point in the gravitational field depends not only on its mass, but also on its velocity; a moving material point acts on another point just like as one moving charge on another. However, since such an influence of the state of motion on the gravitational force has been discovered neither in physics nor astronomy, the vector $\mathfrak{H}^g$ plays a merely hypothetical role. In order to explain the absence of the force arising from $\mathfrak{H}^g$, and represented by the second term in (1), the vector theory appeals to the smallness of the velocity of the bodies in comparison to the speed of light c. Indeed, for celestial bodies the quotient v : c is of the order of 10^{-4}. In addition, the quotient of the magnitudes of the vectors $\mathfrak{H}^g$ and $\mathfrak{E}^g$ is itself of the same order, so that even according to the vector
[475] theory the planetary system | satisfies the Newtonian law up to terms of order 10^{-8}; deviations of this order are of course permissible astronomically as well.

According to the vector theory, the precise law of interaction of two moving mass points should agree up to sign with the fundamental law of electrodynamics. Therefore, a counterpart to the "forces of induction" of electrodynamics ought to exist in mechanics. In the same way as the acceleration of an electric charge produces forces which act on a neighboring charge in a sense to oppose the acceleration, so should the acceleration of a body generate induced gravitational forces which, because of the different signs, act to accelerate neighboring bodies. Accordingly, it seems possible that a system of masses set in motion by a small force further accelerates by itself through internal forces. Thus, the equilibrium of a gravitating system of masses would not be stable.

This instability is related to a difficulty which arises in the vector theory of gravitation, already noted by Maxwell.[3] The differential equations of the static gravitational field are here:

$$\operatorname{div}\mathfrak{E}^g = -\mu, \tag{2}$$

$$\mathfrak{E}^g = -\operatorname{grad}\varphi, \tag{3}$$

where μ denotes the mass density and φ the gravitational potential in suitably chosen units.

Now, the expression for the potential energy of a system of gravitating masses in the action-at-a-distance theory is:

$$E = \int dV \frac{1}{2}\mu\varphi. \tag{4}$$

If, as in electrostatics, one transforms this equation, through integrating by parts, into the following form:

$$E = -\int dV \frac{1}{2}(\mathfrak{E}^g)^2, \tag{4a}$$

then one sees that the interpretation in the sense of the local field theory, leads to a distribution of energy in the field with a density

$$\eta^g = -\frac{1}{2}(\mathfrak{E}^g)^2. \tag{4b}$$

Hence, in contrast to electrostatics, and due to | the opposite sign in eq. (2), *in the vector theory the energy density of the gravitational field turns out to be negative*. [476]
The same applies for a changing gravitational field whose energy density, according to the vector theory, should be

$$\eta^g = -\frac{1}{2}\{(\mathfrak{E}^g)^2 + (\mathfrak{H}^g)^2\}. \tag{4c}$$

Accordingly, a region of space, if it does contain a gravitational field, ought to contain less energy than if it were without a field. And when the gravitational field—

---

3 J. Cl. Maxwell, *Scientific Papers,* Vol. I, p. 570.

spreading, say, in the nature of a wave—enters into a previously field-free region, energy ought to flow in a direction opposite to that of the propagation of the waves. This paradoxical conclusion is peculiar to the vector theory of the gravitational field.

The so-called "*theory of relativity*" arose from the Lorentzian electrodynamics of moving bodies. This theory demands, that in empty space all forces propagate with the same speed as light. Therefore, according to the theory of relativity, the gravitational force as well must propagate with the speed of light. H. Poincaré(2) raised the question as to how this requirement can be brought into agreement with the view of Laplace, that the speed of propagation of gravitation, if at all finite, must be far greater than that of light (c). He remarked, that this view applies only, if quantities of first order (with respect to the quotient $v : c$) enter into the fundamental law of attraction of two masses. But if one formulates the fundamental law so that it fits into a relativistic scheme, yet deviates only in terms of second order and higher from Newton's, then Laplace's reservations lose their significance. Poincaré gives such formulations of the fundamental law of gravitation. These approaches contain also the force term of H. Minkowski,(3) which, by the way, can be obtained by transferring the fundamental law of the theory of electrons to gravitation.(4) The interaction laws for the attraction of masses developed by these pioneers of the theory of relativity are accordingly quite readily compatible with the vector theory of gravitation sketched above. On the other hand, that vector theory gave also an account of those processes in the field through whose mediation one mass transfers energy and momentum to the other; these fundamental relativistic laws, however, lack the derivation from the field equations. Thus the difficulties associated with the vector theory were not resolved but only concealed. What remained unresolved was the problem of developing field |
[477] equations of the gravitational field that yield a propagation of the gravitational force with the speed of light and also ascribe to the field a positive energy, transferred via an energy flux, and a momentum transferred by means of fictitious stresses.

B. Conservation Laws of Momentum and Energy: World Tensors

The above mentioned work by H. Poincaré(2) already contains the beginnings of the four-dimensional vector calculus, which was then further developed by H. Minkowski(3) In our notation, we will mainly follow the presentation of A. Sommerfeld.(4) The four-dimensional formulation of the theory of relativity, as is well known, interprets the group of Lorentz transformations as a rotation group of a four-dimensional space, whose coordinates are the Cartesian coordinates x, y, z of ordinary space and $u = ict$; these transformations leave the expression

$$x^2 + y^2 + z^2 + u^2$$

invariant. Poincaré had already dealt with four-vectors whose components transform like the coordinates x, y, z, u. H. Minkowski introduced the concept of the six-vector, and characterized the electromagnetic field $\{\mathfrak{E}, \mathfrak{H}\}$ in a vacuum through such a vector. The six-vector $\{\mathfrak{E}, \mathfrak{H}\}$ is derived from the four-potential $\{\mathfrak{A}, \Phi\}$, whose 4 components are given by the electromagnetic vector- and scalar potential. The vector theory of the gravitational force sketched above, which, of course, is formally identical with the electromagnetic field theory, accordingly also derives the six-vector of the gravitational field $\{\mathfrak{E}^g, \mathfrak{H}^g\}$ from a four-potential. Also in this four-dimensional sense—namely insofar as the gravitational potentials form a four-vector—one will accordingly have to denote that theory of gravitation as a "vector theory."

Besides the two kinds of "world vectors," the "*world tensors*" are important for the theory of relativity.[4] The ten components of such a symmetric tensor T transform like the squares and products of the four coordinates x, y, z, u. From it one derives a four-vector which we want to call "the divergence of the ten-tensor T," and whose components are to be formed according to the following scheme: |

$$\operatorname{div} T = \left\{ \begin{array}{l} \dfrac{\partial T_{xx}}{\partial x} + \dfrac{\partial T_{xy}}{\partial y} + \dfrac{\partial T_{xz}}{\partial z} + \dfrac{\partial T_{xu}}{\partial u}, \\ \dfrac{\partial T_{yx}}{\partial x} + \dfrac{\partial T_{yy}}{\partial y} + \dfrac{\partial T_{yz}}{\partial z} + \dfrac{\partial T_{yu}}{\partial u}, \\ \dfrac{\partial T_{zx}}{\partial x} + \dfrac{\partial T_{zy}}{\partial y} + \dfrac{\partial T_{zz}}{\partial z} + \dfrac{\partial T_{zu}}{\partial u}, \\ \dfrac{\partial T_{ux}}{\partial x} + \dfrac{\partial T_{uy}}{\partial y} + \dfrac{\partial T_{uz}}{\partial z} + \dfrac{\partial T_{uu}}{\partial u}. \end{array} \right\} \tag{5}$$ [478]

In the Maxwell-Lorentz electrodynamics, the electromagnetic force acting on a unit volume is determined by the four-vector

$$\rho\mathfrak{F}^e = \operatorname{div} T^e, \tag{6}$$

where T_e designates the "*electromagnetic ten-tensor*." Its components

$$T^e_{xx}, T^e_{yy}, T^e_{zz}$$

represent the fictitious normal stresses, and

$$T^e_{xy} = T^e_{yx}, \quad T^e_{yz} = T^e_{zy}, \quad T^e_{zx} = T^e_{xz} \tag{6a}$$

represent the shear stresses; the symmetry relations (6a) imply the vanishing of the torques of these fictitious surface forces [*Flächenkräfte*]. The three remaining symmetry conditions of the ten-tensor T^e

$$T^e_{xu} = T^e_{ux}, \quad T^e_{yu} = T^e_{uy}, \quad T^e_{zu} = T^e_{uz} \tag{6b}$$

---

4 M. Abraham, *Rendiconti del circolo matematico di Palermo,* XXVIII^2, 17, 1909. M. Laue, *Das Relativitätsprinzip*, p. 4 of the 2nd ed. (1913).

have the following significance:

$$T^e_{xu} = -ic\mathfrak{g}^e_x, \quad T^e_{yu} = -ic\mathfrak{g}^e_y, \quad T^e_{zu} = -ic\mathfrak{g}^e_z \tag{6c}$$

determine the components of the momentum density $(\mathfrak{g}^e)$ of the electromagnetic field,

$$T^e_{ux} = -\frac{i}{c}\mathfrak{S}^e_x, \quad T^e_{uy} = -\frac{i}{c}\mathfrak{S}^e_y, \quad T^e_{uz} = -\frac{i}{c}\mathfrak{S}^e_z \tag{6d}$$

determine those of the electromagnetic energy flux $(\mathfrak{S}^e)$; the eqs. (6b) contain therefore the "*theorem of the momentum of electromagnetic energy flux*":

$$\mathfrak{g}^e = \frac{\mathfrak{S}^e}{c^2}. \tag{6e}$$

Since according to (6c)

$$\frac{\partial T^e_{xu}}{\partial u} = \frac{\partial T^e_{xu}}{ic\partial t} = -\frac{\partial \mathfrak{g}^e_x}{\partial t} \qquad \text{etc.,}$$

the first three of the equations (6), to be formed according to the scheme (5), express the *law of conservation of momentum* for the electromagnetic field, as they derive from the stresses and the momentum density of the field the momentum given by the
[479] field to the unit of volume of matter. | The last of the eqs. (6), however, contains the law of *conservation of energy.* Indeed, if one equates the last tensor component to the electromagnetic energy density:

$$T^e_{uu} = \eta^e = \frac{1}{2}(\mathfrak{E}^2 + \mathfrak{H}^2), \tag{6f}$$

and furthermore

$$\rho\mathfrak{F}^e_u = \frac{i}{c}\frac{dA^e}{dt}, \tag{6g}$$

where dA^e is the work done on a unit volume, then the last of eqs. (6) reads, taking into consideration (6d):

$$\frac{dA^e}{dt} = -\operatorname{div}\mathfrak{S}^e - \frac{\partial \eta^e}{\partial t}. \tag{6h}$$

It therefore expresses the law of conservation of energy for the electromagnetic field, because it relates the energy transferred from the field to a unit volume of matter to the energy flux and the energy density of the field.

One would now wish to retain the validity of the conservation laws of momentum and energy, and of the theorem of the momentum of energy flux for the gravitational field by deriving the gravitational force per unit volume of matter as a four-vector

$$\mu\mathfrak{F}^g = \operatorname{div} T^g \tag{7}$$

from a symmetric "*gravitational tensor* T^g." The vector theory of gravitation satisfies this demand by expressing the tensor T^g through the vectors $\mathfrak{E}^g$, $\mathfrak{H}^g$, in the same way in which $-T^e$ is determined through the electromagnetic vectors $\mathfrak{E}$, $\mathfrak{H}$. We have shown above that precisely the difference in sign causes difficulties for the vector theory. In the following, we will become acquainted with other possibilities for representing the gravitational tensor T^g.

C. Inertia and Gravity[1]

If the theorem of the momentum of energy flux is valid, then a momentum, and thus an inertial mass, is associated with a convectively moving quantity of energy. A well-known example is the "electromagnetic mass" of the electron which, in the dynamics of the electron, is derived from the momentum of the energy flux flowing in the electromagnetic field of the moving electron. If one considers the theorem of the momentum of energy flux as valid for arbitrary kinds of energy flows, then it follows that the momentum of a uniformly moving body isolated from external effects is |

$$\mathfrak{G} = \frac{\mathfrak{v}}{c^2}E. \tag{8}$$ [480]

To this corresponds an inertial rest-mass of value

$$m_0 = \frac{E_0}{c^2}. \tag{8a}$$

Here, E_0 denotes the "rest-energy" of the body, i.e. its energy with respect to a coordinate system Σ_0 in which the body is at rest; m_0 measures the inertia of the body accelerated from rest. Equation (8a) expresses "*the theorem of the inertia of energy*."

If all forces in nature can be fitted into the scheme of a symmetric world tensor, then the theorem of the momentum of energy flux and that which is derived from it, the theorem of the inertia of energy, gain general validity. The inertial mass of a body is then proportional to its energy content; it can be increased by the gain of energy, or decreased through the loss of energy. However, the denominator c^2 in (8a) entails that the changes in energy that occur in the usual chemical reactions are too minute to cause measurable changes in mass. However, the radioactive transformations, in view of their enormous heat production, should be accompanied by a noticeable decrease in mass. Such changes in mass—still quite small at any rate—could, however, be determined only with a scale. The scale, however, does not measure inertia but weight [*Schwere*]. Hence, the question arises: Is there also gravitational mass associated with energy? Is there, as a counterpart to the theorem of the inertia of energy, a theorem of the weight of energy? This question leads us back to the problem of gravitation.

Experience teaches us that all bodies fall with equal acceleration in a vacuum and that the period of a pendulum is independent of its chemical composition. Therefore, *the gravitational mass is proportional to the inertial mass*. The most precise test of the law of proportionality is due to B. Eötvös;[5] with the aid of a torsion balance, this researcher investigated whether the gravitational force acts on all bodies on the surface of the Earth in the same direction. Since the gravitational force is the resultant of the Earth's attraction of masses [*Massenanziehung*] and of the centrifugal force, this resultant, i.e. the vertical, would have a different direction for different bodies if strict
[481] proportionality between the two masses did not hold. This is because | the attraction of masses is determined by the gravitational mass and the centrifugal force by the inertial mass. Although Eötvös recently refined his measurements so that deviations of the order 10^{-8} would not have escaped him, he nevertheless could not find such. With the corresponding accuracy, the law of proportionality of gravitational and inertial mass has been shown to be valid.

For the time being, the investigations of Eötvös do not cover radioactive bodies. If at all, one should most likely expect a departure in the behavior of gravitational and inertial mass in radioactive transmutations, namely in the case that the former remains constant while the latter decreases because of the emission of energy. Uranium oxide for example transmutes by radioactive decay into lead oxide, while emitting 8 α-particles per atom, during which the fraction $2.3 \cdot 10^{-4}$ of the energy is emitted.[6] According to the theorem of the equivalence of inertia and energy, the inertial mass should decrease by the same fraction; if the gravitational mass did not remain constant, or if it changed in a different ratio, then the proportionality between gravitational and inertial mass could not be maintained under radioactive transformations. However, L. Southerns,[7] using pendulum observations to which he ascribed an accuracy of $5 \cdot 10^{-6}$, discovered no difference with respect to the mass ratio between uranium oxide and lead oxide. Thus, here too, the law of proportionality is valid.

Classical mechanics introduces the proportionality of gravitational mass and inertia as an empirical law without deeper justification. The theorem of the inertia of energy suggests that, in the new mechanics, the equivalence of the two masses be explained by the gravitational mass also being proportional to the energy. Then furthermore, the law of conservation of (gravitational) mass, which takes an isolated position in the traditional physics and chemistry, would merge with the law of conservation of energy into a single one. However, if not the energy itself, but another quantity were the determining factor for the gravitational mass of a body, then still, in all practical cases, this quantity should be proportional to the inertial mass with the above stated accuracy. It is the merit of the newer theories of gravitation, on which we want to report in the following, to have put the discussion of the relation between

5 B. Eötvös, *Mathem. u. naturwiss. Ber. aus Ungarn*, 8, 65, 1890.

6 See also R. Swinne, *Physik. Zeitschr.,* 14, 145, 1913.

7 L. Southerns, *Proc. Roy. Soc. London*, 84, 325, 1910.

inertia, gravitation and energy on a rational foundation and, in so doing, to have prepared the way for the exploration of these relations. |

II. SCALAR THEORIES [482]

A. Energy Density and Energy Flux in the Gravitational Field

In the vector theory of gravitation, the expression (4b) gave the energy density of the static gravitational field. If one assumes that a similar expression is also valid for time varying fields, when only the static force $\mathfrak{E}^g$ is replaced by the dynamic gravitational force $\mathfrak{F}^g$, then

$$\eta^g = -\frac{1}{2}(\mathfrak{F}^g)^2. \tag{9}$$

In a remarkable paper on the energy flux in a gravitational field, V. Volterra(1)[2] takes this expression as the basis for the energy density. He divides the energy flux into two parts:

$$\mathfrak{S} = \mathfrak{S}^g + \mathfrak{S}^m, \tag{10}$$

of which the first

$$\mathfrak{S}^g = -\frac{\varphi\partial\mathfrak{F}^g}{\partial t} \tag{10a}$$

is caused by the variation of the gravitational field with respect to time, whereas the second

$$\mathfrak{S}^m = \varphi\mu\mathfrak{v} \tag{10b}$$

represents the energy transport by moving matter. The mass density μ is related via

$$\operatorname{div}\mathfrak{F}^g = -\mu \tag{11}$$

with the gravitational force per unit mass, which, by virtue of

$$\mathfrak{F}^g = -\operatorname{grad}\varphi \tag{11a}$$

is derived from the scalar gravitational potential φ. Insofar as the existence of a scalar potential is also assumed in a dynamic gravitational field, Volterra's theory is to be counted among the "scalar theories of gravitation" (in a three-dimensional sense).

We want to convince ourselves that Volterra's approaches satisfy the energy equation. According to (9) one has

$$-\frac{\partial\eta^g}{\partial t} = \left(\mathfrak{F}^g \cdot \frac{\partial\mathfrak{F}^g}{\partial t}\right),$$

whereas from (10), (10a), and (10b) it follows that

$$-\mathrm{div}\mathfrak{S} = \mathrm{div}\varphi\frac{\partial\mathfrak{F}^g}{\partial t} - \mathrm{div}\varphi\mu\mathfrak{v}$$

$$= \varphi\mathrm{div}\frac{\partial\mathfrak{F}^g}{\partial t} + \left(\frac{\partial\mathfrak{F}^g}{\partial t}\cdot\mathrm{grad}\varphi\right) - \varphi\ \mathrm{div}\mu\mathfrak{v} - (\mu\mathfrak{v}\cdot\mathrm{grad}\varphi),$$

and taking into consideration (11), (11a) and the equation of continuity of matter |

$$\frac{\partial\mu}{\partial t} = -\mathrm{div}\mu\mathfrak{v},$$

[483]

$$-\mathrm{div}\mathfrak{S} = -\left(\mathfrak{F}^g\cdot\frac{\partial\mathfrak{F}^g}{\partial t}\right) + (\mu\mathfrak{v}\cdot\mathfrak{F}^g).$$

Therefore, one obtains

$$(\mu\mathfrak{F}^g\cdot\mathfrak{v}) = -\mathrm{div}\mathfrak{S} - \frac{\partial\eta^g}{\partial t}. \tag{11b}$$

Here, on the left hand side, is the work done by the gravitational force per unit volume and time; on the right-hand side, is the energy gain caused by the influx of energy and by the temporal decrease of the energy density. The law of conservation of energy is indeed satisfied.

However, it is not possible, by means of constructing a corresponding world tensor T^g, to reconcile the approaches of Volterra with the first three of the eqs. (7), which formulate the laws of conservation of momentum. It is also remarkable that in (10b) an energy transport by moving matter is introduced, without a corresponding energy density of the matter. If one takes account of such an energy density

$$\eta^m = \varphi\mu \tag{12}$$

then one obtains the correct value for the total energy of a system of masses at rest

$$E = \int dV\frac{1}{2}\mu\varphi = E^m + E^g,$$

if one sets

$$\eta^g = \frac{1}{2}(\mathfrak{F}^g)^2 \tag{12a}$$

for the *energy density of the gravitational field*.

Indeed, according to (11) and (11a), the identity

$$\int dV\frac{1}{2}\mu\varphi = -\int dV\frac{1}{2}(\mathfrak{F}^g)^2$$

holds, and it follows that

$$\int dV\left\{\mu\varphi + \frac{1}{2}(\mathfrak{F}^g)^2\right\} = \int dV\frac{1}{2}\mu\varphi = E.$$

The value (12a) for the *energy density is positive*. Hence, the difficulty appearing in the vector theory of gravitation is eliminated if one succeeds in pairing that energy density with an energy flow.

We equate[6)] the *energy flux in a gravitational field to the gravitational force per unit mass multiplied by the time derivative of the gravitational potential:*

$$\mathfrak{S}^g = \frac{\partial\varphi}{\partial t}\mathfrak{F}^g, \tag{13}$$

| whereas, for the energy flux of matter, we retain the expression (10b). Then the *total* [484]
energy flux becomes

$$\mathfrak{S} = \mathfrak{S}^m + \mathfrak{S}^g = \varphi\mu\mathrm{v} + \frac{\partial\varphi}{\partial t}\mathfrak{F}^g, \tag{14}$$

whereas the *total energy density,* is according to (12, 12a)

$$\eta = \eta^m + \eta^g = \varphi\mu + \frac{1}{2}(\mathfrak{F}^g)^2. \tag{14a}$$

One can easily convince oneself that the energy equation is satisfied. One has

$$\frac{\partial\eta}{\partial t} = \varphi\frac{\partial\mu}{\partial t} + \mu\frac{\partial\varphi}{\partial t} + \left(\mathfrak{F}^g \cdot \frac{\partial\mathfrak{F}^g}{\partial t}\right),$$

$$\mathrm{div}\mathfrak{S} = \varphi\mathrm{div}\mu\mathfrak{v} + (\mu\mathfrak{v}\cdot\mathrm{grad}\varphi) + \frac{\partial\varphi}{\partial t}\mathrm{div}\mathfrak{F}^g + \left(\mathfrak{F}^g\cdot\mathrm{grad}\frac{\partial\varphi}{\partial t}\right),$$

and from this, taking into consideration (11, 11a) and the equation of continuity, the law of conservation of energy follows

$$(\mu\mathfrak{v}\cdot\mathfrak{F}^g) = -\mathrm{div}\mathfrak{S} - \frac{\partial\eta}{\partial t}. \tag{14b}$$

As we will see immediately, the expressions (12a) and (13) for the energy density and the energy flux in a gravitational field, fit readily into the scheme of a symmetric gravitational tensor.[8]

B. Abraham's First Theory

So far we have been talking about "scalar theories" of gravitation merely in the context of a three-dimensional vector calculus, and taking them to be theories which derive the gravitational force from a scalar potential φ also for a dynamic field. In the four-dimensional context, *a theory of gravitation is to be designated as a scalar theory, in which the gravitational potential is a scalar, i.e. an invariant with respect to rotation of the four-dimensional space of the* (x, y, z, u). The gravitational force with respect to a unit volume, considered as a four-vector, shall have the components |

[485]

$$\left.\begin{aligned} \mu\mathfrak{F}_x^g &= -\nu\frac{\partial\varphi}{\partial x}, & \mu\mathfrak{F}_y^g &= -\nu\frac{\partial\varphi}{\partial y}, \\ \mu\mathfrak{F}_z^g &= -\nu\frac{\partial\varphi}{\partial z}, & \mu\mathfrak{F}_u^g &= -\nu\frac{\partial\varphi}{\partial u}; \end{aligned}\right\} \tag{15}$$

ν is an, for now, undetermined scalar factor. As the comparison with (6g) teaches,

$$\frac{c}{i}\mu\mathfrak{F}_u^g = ic\nu\frac{\partial\varphi}{\partial u} = \nu\frac{\partial\varphi}{\partial t} \tag{15a}$$

represents the energy that is transferred per unit space and time from the gravitational field to the matter. In the symbolism of the four-dimensional vector analysis(4), the eqs. (15) are written as

$$\mu\mathfrak{F}^g = -\nu\operatorname{grad}\varphi. \tag{15b}$$

Now, corresponding to the scheme (7), the gravitational force should be derived from a symmetric gravitational tensor T^g, such that the expressions (12a, 13) for the energy density and the energy flux correspond to the appropriate tensor components. M. Abraham(6) defines the ten-tensor T^g in the following manner:

8 Here, I have presented the train of thought which I pursued in the development of my first theory of gravitation(6) in such detail, because G. Jaumann (*Physik. Zeitschr.,* 15, 159, 1914) formulated an inappropriate hypothesis concerning the psychological origin of this theory. Jaumann's theory (*Wien. Ber.,* 121, 95, 1912), of which I became aware only after that first publication, is so far removed from the conceptual realm of the investigations summarized in this report, that it appears to me that this is not the place for its discussion.

$$\left.\begin{aligned}
T^g_{xx} &= -\frac{1}{2}\left(\frac{\partial\varphi}{\partial x}\right)^2 + \frac{1}{2}\left(\frac{\partial\varphi}{\partial y}\right)^2 + \frac{1}{2}\left(\frac{\partial\varphi}{\partial z}\right)^2 + \frac{1}{2}\left(\frac{\partial\varphi}{\partial u}\right)^2,\\
T^g_{yy} &= \frac{1}{2}\left(\frac{\partial\varphi}{\partial x}\right)^2 - \frac{1}{2}\left(\frac{\partial\varphi}{\partial y}\right)^2 + \frac{1}{2}\left(\frac{\partial\varphi}{\partial z}\right)^2 + \frac{1}{2}\left(\frac{\partial\varphi}{\partial u}\right)^2,\\
T^g_{zz} &= \frac{1}{2}\left(\frac{\partial\varphi}{\partial x}\right)^2 + \frac{1}{2}\left(\frac{\partial\varphi}{\partial y}\right)^2 - \frac{1}{2}\left(\frac{\partial\varphi}{\partial z}\right)^2 + \frac{1}{2}\left(\frac{\partial\varphi}{\partial u}\right)^2,\\
T^g_{uu} &= \frac{1}{2}\left(\frac{\partial\varphi}{\partial x}\right)^2 + \frac{1}{2}\left(\frac{\partial\varphi}{\partial y}\right)^2 + \frac{1}{2}\left(\frac{\partial\varphi}{\partial z}\right)^2 - \frac{1}{2}\left(\frac{\partial\varphi}{\partial u}\right)^2,\\
T^g_{xy} &= T^g_{yx} = -\frac{\partial\varphi}{\partial x}\frac{\partial\varphi}{\partial y},\\
T^g_{yz} &= T^g_{zy} = -\frac{\partial\varphi}{\partial y}\frac{\partial\varphi}{\partial z},\\
T^g_{zx} &= T^g_{xz} = -\frac{\partial\varphi}{\partial z}\frac{\partial\varphi}{\partial x},\\
T^g_{xu} &= T^g_{ux} = -\frac{\partial\varphi}{\partial x}\frac{\partial\varphi}{\partial u},\\
T^g_{yu} &= T^g_{uy} = -\frac{\partial\varphi}{\partial y}\frac{\partial\varphi}{\partial u},\\
T^g_{zu} &= T^g_{uz} = -\frac{\partial\varphi}{\partial z}\frac{\partial\varphi}{\partial u}.
\end{aligned}\right\} \tag{16}$$

As the comparison with (6c, d, e, f) shows, the appropriate expressions for the *momentum density, energy flux* and *energy density* are accordingly |

$$c^2\mathfrak{g}^g = \mathfrak{S}^g = -\frac{\partial\varphi}{\partial t}\mathrm{grad}\varphi, \tag{16a}$$ [486]

$$\eta^g = T^g_{uu} = \frac{1}{2}\left\{\left(\frac{\partial\varphi}{\partial x}\right)^2 + \left(\frac{\partial\varphi}{\partial y}\right)^2 + \left(\frac{\partial\varphi}{\partial z}\right)^2 + \frac{1}{c^2}\left(\frac{\partial\varphi}{\partial t}\right)^2\right\}. \tag{16b}$$

Equation (16a) agrees with (13) and, in addition, contains the theorem of the momentum of energy flux. Equation (16b) becomes (12a) for static fields; but also for the dynamic field the energy density calculated according to (16b), is always positive. The fictitious stresses contained in T^g agree for the static field, apart from signs, with the Maxwellian electrostatic stresses. Let us write the system (16), which derives a ten-tensor T^g from the scalar φ, symbolically as

$$T^g = \mathrm{Ten}\varphi. \tag{16c}$$

We now apply the scheme (5), and obtain for the divergence of the gravitational tensor:

$$\operatorname{div} T^g = \operatorname{div}\,\operatorname{ten}\varphi = -\Box\varphi \cdot \operatorname{grad}\varphi, \tag{17}$$

using the abbreviation

$$\Box\varphi = \operatorname{div}\,\operatorname{grad}\varphi = \frac{\partial^2\varphi}{\partial x^2} + \frac{\partial^2\varphi}{\partial y^2} + \frac{\partial^2\varphi}{\partial z^2} + \frac{\partial^2\varphi}{\partial u^2}. \tag{17a}$$

The laws of conservation of momentum and energy summarized in (7) thus yield

$$\mu\mathfrak{F}^g = -\Box\varphi \cdot \operatorname{grad}\varphi. \tag{17b}$$

In order to obtain agreement with (15b), the gravitational potential is required to satisfy the *field equation*

$$\Box\varphi = \nu, \tag{18}$$

which is to be taken as a generalization of Poisson's equation.

The theory of relativity demands that ν be a scalar in the four-dimensional sense, like the gravitational potential φ and the operator $\Box$; the "rest-mass density" is such a scalar. For this reason, in his first communication,(6) Abraham used this for ν. By integrating (18), he calculates the gravitational potential of a moving mass point and thus obtains the "*fundamental law of gravitation*."(7) This law is simpler than the fundamental law of electrodynamics, insofar as the *direction of the attracting force is independent of the state of motion of the attracted point*. This is a peculiarity of the scalar theories, which determine the gravitational force by means of a single vector $\mathfrak{F}^g$.

In the vacuum, ν is equal to zero and the gravitational potential therefore satisfies the differential equation |

[487]
$$\Box\varphi = \frac{\partial^2\varphi}{\partial x^2} + \frac{\partial^2\varphi}{\partial y^2} + \frac{\partial^2\varphi}{\partial z^2} - \frac{1}{c}\frac{\partial}{\partial t}\left(\frac{1}{c}\frac{\partial\varphi}{\partial t}\right) = 0, \tag{18a}$$

the so-called "wave equation." This results in the possibility of gravitational waves, which propagate with the speed of light. However, in the scalar theories the *gravitational waves are longitudinal,* whereas in the vector theory they are considered as transverse, in analogy with the electromagnetic waves. According to both theories, the factor determining the amplitude of the gravitational wave emitted by an oscillating mass particle is the product of gravitational mass and acceleration. One could now surmise that during the emission of α-rays from radioactive atoms, where very large accelerations occur, the hypothetical gravitational waves will be excited in noticeable strength. However, electric waves are simultaneously excited and, as the quantitative discussion(10) shows, the force excited by gravitational waves from one emitted α-particle on a second α-particle is at most 10^{-36} of the electric force. That the gravitational waves play no role in the balance of nature has however still a deeper reason. If one mass particle imparts an acceleration to another through collision or long range forces, then the second particle also acts to accelerate the first in such a way that the vector sum of the products of inertial mass and acceleration is equal to zero. What determines the amplitude of the gravitational wave emitted from the system of the two particles is the vector sum of the products of gravitational mass and acceleration. Due to the proportionality of inertial and gravitational mass, this sum is also equal to zero. Thus, even theoretically, one cannot provide a means to excite gravitational waves of noticeable strength.

One aspect of Abraham's first theory has so far not been mentioned. Following an hypothesis already proposed by A. Einstein,(5) one considers the gravitational potential φ to be a function of the speed of light c, and, in particular, the following relation between the two quantities results from the assumed form of the equation of motion of material points:

$$\varphi - \varphi_0 = \frac{c^2}{2} - \frac{c_0{}^2}{2}. \tag{19}$$

Since, according to this, c is variable in a gravitational field, Einstein and Abraham thus give up the postulate of the constancy of the speed of light, which was a fundamental requirement of Einstein's theory of relativity | of 1905. The covariance [488]
of the physical laws with respect to the Lorentz transformations demanded by that theory exists then only in infinitesimally small spacetime regions in which c can be considered to be constant. Only with regard to the processes in an infinitesimal region are two uniformly moving systems of reference to be considered as equivalent. But if one regards finite systems of gravitating masses, then there exists no equivalence whatsoever between such systems of reference. Indeed, G. Pavanini(18) was able to show that, according to Abraham's theory, secular perturbations of the Keplerian motion occur in a two-body system, which depend on the "absolute" state of motion of the system. If the center of mass of the system is at rest, then only a secular motion of the perihelion occurs; for Mercury it would amount to $14'', 52$ in a hundred years, that is only about one third of the actual motion of the perihelion.

Our planetary system moves within the gravitational field of the other celestial bodies. The particular system of reference in which the external gravitational field can be taken as static, plays a distinguished role according to Abraham(10) (13) and is to coincide with the "absolute" system of reference, which, at least in the case of rotational motion, makes itself felt through centrifugal and Coriolis forces, and which the action-at-a-distance theory of C. Neumann anchors in a hypothetical "body α." This view, however, did not meet with the approval of Einstein. He attempted to salvage the principle of relativity of 1905(14), wherein, however, he had to restrict the equivalence between systems of reference in uniform translatory motion with respect to each other to processes in "isolated systems," and to the "limiting case of constant gravitational potential." As Abraham remarked,(15) to the contrary, it is not possible to shield a system against the gravitational force, and the limiting case of constant gravitational potential corresponds to the absence of gravitational force. If, on the other hand, those theoreticians who define the gravitational potential through c consider gravitation as an essential property of matter, then they must abandon "yesterday's theory of relativity."

On the occasion of that dispute, Einstein,(14) incidentally, revealed the prospect of a more general "tomorrow's principle of relativity," encompassing accelerated and rotational motion. To what extent the "generalized theory of relativity," published by him in association with M. Grossmann the following year (1913), fulfills that promise is to be discussed in detail later in (III B). |

[489] *C. Abraham's Second Theory*

In that discussion there emerged at least this much agreement, that in the development of the theory of gravitation attention has to be paid to the relation between weight and energy. In a lecture(16) at the international congress of mathematics at Cambridge (August 1912), M. Abraham gave the highest priority to the *postulate of the weight of energy*. He proved that the gravitational force on a moving point mass can be strictly proportional to its energy only if its Lagrangian function is a linear homogeneous function of the velocity and of the gravitational potential. The Lagrangian of the earlier theory of relativity

$$\mathfrak{L} = -m_0 c^2 \sqrt{1 - \left(\frac{v}{c}\right)^2} \tag{20}$$

is to be adapted to this demand by interpreting the speed of light c itself as a potential and considering

$$m_0 c = M \text{ (mass constant)} \tag{20a}$$

as independent of v and c. Then the *Lagrangian of the mass-point* becomes

$$\mathfrak{L} = -M \cdot \sqrt{c^2 - v^2}, \tag{21}$$

and the *rest-energy*

$$E_0 = -\mathfrak{L}_0 = Mc. \tag{21a}$$

Hence, the rest-energy decreases with decreasing c, whereas the *rest-mass*

$$m_0 = \frac{E_0}{c^2} = \frac{M}{c}$$

increases with decreasing c.

Now one must demand that all mechanical quantities of the same class possess the same degree in c, thus all energies possess the degree c, all masses the degree c^{-1}. Later the author derives this requirement(17) from the following postulate:[3] "An observer belonging to a mechanical system must not perceive that he, together with the system, is brought into a region in which c has a different value" Since lengths should not depend on c, time intervals are of degree c^{-1}. The energy density as well as the remaining components of the ten-tensor T^g are of degree c; but since they are of the second degree in the derivative of φ with respect to the coordinates (cf. 16), so φ is to be replaced not by c, but by $\sqrt{c}$. Correspondingly, Abraham chooses the gravitational tensor T^g according to the scheme (16, 16c):

$$T^g = \text{ten}\sqrt{c}. \tag{22}$$

| Then from (5) one obtains, similar to (17), [490]

$$\text{div}T^g = \text{div ten}\sqrt{c} = -\Box(\sqrt{c})\cdot\text{grad}\sqrt{c}, \tag{22a}$$

where we have set

$$\Box(\sqrt{c}) = \frac{\partial^2\sqrt{c}}{\partial x^2}+\frac{\partial^2\sqrt{c}}{\partial y^2}+\frac{\partial^2\sqrt{c}}{\partial z^2}-\frac{1}{c}\frac{\partial}{\partial t}\left(\frac{1}{c}\frac{\partial\sqrt{c}}{\partial t}\right). \tag{22b}$$

In order for the postulate of the weight of energy to be valid, the *field equation* is to be formulated as follows

$$\sqrt{c}\cdot\Box(\sqrt{c}) = 2\eta^m. \tag{23}$$

Then, according to (22a), the laws of conservation of momentum and energy, summarized in (7), yield

$$\mu\mathfrak{F}^g = -\frac{2\eta^m}{\sqrt{c}}\text{grad}\sqrt{c} = -\frac{\eta^m}{c}\text{grad}c. \tag{23a}$$

for the gravitational force per unit volume. Integration over a body of sufficiently small volume yields the *gravitational force*

$$\mathfrak{K}^g = -\frac{E^m}{c}\text{grad}c. \tag{24}$$

Therefore, the weight of a body is proportional to its energy; by the way, in (24), the gravitational field determined by the gradient of c has been assumed to be homogeneous over the extension of the body, and correspondingly the gravitational field produced by the body itself and its energy has been neglected.

In a lecture(17) presented at the congress of the “Società italiana per il progresso delle scienze,” the author treats the role of the self-excited field and its energy for the case of rest. Since here η^m denotes the density of rest-energy of matter, it is useful to define μ as the mass constant with respect to the unit volume, and correspondingly to write (21a) as

$$\eta^m = \mu c, \tag{24a}$$

so that according to (23a)

$$\mathfrak{F}^g = -\mathrm{grad}\, c \tag{24b}$$

denotes the gravitational force per unit of mass as measured by M. For incompressible fluids, for example, μ is a constant also independent of c.

For the static gravitational field, the differential eq. (23) is written as

$$\Delta(\sqrt{c}) = \frac{2\eta^m}{\sqrt{c}} = 2\mu\sqrt{c}. \tag{25}$$

Since, according to (16b), the energy density of the static field is

$$\eta^g = \frac{1}{2}(\mathrm{grad}\sqrt{c})^2,$$

[491] | and since the identity

$$\Delta c = 2\sqrt{c}\Delta\sqrt{c} + 2(\mathrm{grad}\sqrt{c})^2$$

applies, one can write (25) as

$$\Delta c = 4(\eta^m + \eta^g) = 4\eta. \tag{25a}$$

Thus, the divergence of the gradient of c is proportional to the density of the total energy, i.e., according to Gauss' theorem: *The flux of the gravitational force vector* $\mathfrak{F}^g$ (cf. 24b) *through a closed surface is equal to four times the enclosed energy.*

For a stationary homogeneous sphere of radius α we have, according to (25),

$$\Delta(\sqrt{c}) = 0 \qquad \text{for} \quad r > a, \tag{26}$$

$$\Delta(\sqrt{c}) = 2\mu\sqrt{c} \qquad \text{for} \quad r < a. \tag{26a}$$

The integration is not difficult to perform.[17] According to (26), the gradient of $\sqrt{c}$ decreases outside the sphere as r^{-2}; the gradient of c, the determining factor for the gravitational force per unit mass according to (24b), is obtained as

$$-\mathfrak{F}_r^g = \frac{dc}{dr} = \frac{E}{\pi r^2}\left(1 - \psi\frac{a}{r}\right), \tag{26b}$$

where

$$\psi = E_a/E \tag{26c}$$

denotes the quotient of the energy E_a of the external gravitational field and the total energy E of the sphere and of its gravitational field. This result is in agreement with the above theorem concerning the flux of the vector $\mathfrak{F}^g$. The energy of the gravitational field outside the sphere is responsible for the *deviation from Newton's law* because it causes the difference in the force flux through two concentric spheres. Incidentally, eq. (26b) applies also to a mass distribution homogeneous within concentric layers.

The integration of (26a) for the interior of a homogeneous sphere leads to hyperbolic functions. For a body with the mass and the radius of the Sun the quantity ψ, given in (26c), turns out to be of the order of 10^{-6}. This provides an idea of the order of magnitude of the quotient of the external and of the total energy, as well as of the order of magnitude of the deviation from the Newtonian law determined by it. A celestial body may be considered as a material point if that quotient can be neglected under the given circumstances.

We return to the Lagrangian function (21) of the material point; it implies the values for *momentum* and *energy* |

$$\mathfrak{G} = \frac{\partial \mathfrak{L}}{\partial v} \cdot \frac{\mathfrak{v}}{v} = \frac{M\mathfrak{v}}{\sqrt{c^2 - v^2}}, \tag{27}$$ [492]

$$E = v\frac{\partial \mathfrak{L}}{\partial v} - \mathfrak{L} = \frac{Mc^2}{\sqrt{c^2 - v^2}}. \tag{27a}$$

Lagrange's equations for the free motion of a point mass

$$\frac{d}{dt}\left(\frac{\partial \mathfrak{L}}{\partial x}\right) - \frac{\partial \mathfrak{L}}{\partial x} = 0 \qquad \text{etc.}$$

can be written

$$\frac{d\mathfrak{G}}{dt} = \operatorname{grad}\mathfrak{L} = \frac{\partial \mathfrak{L}}{\partial c}\operatorname{grad}c = \mathfrak{K}^g; \tag{28}$$

They contain the inertial force as well as the gravitational force, which is

$$\mathfrak{K}^g = \frac{\partial \mathfrak{L}}{\partial c}\operatorname{grad}c = -\frac{Mc}{\sqrt{c^2 - v^2}}\operatorname{grad}c, \tag{28a}$$

or, according to (27a),

$$\mathfrak{K}^g = -\frac{E}{c}\operatorname{grad}c. \tag{28b}$$

For a given material point in a gravitational field, the gravitational force is, also in the case of motion, thus proportional to its energy; this corresponds to the primary postulate of the weight of energy.

The substitution of (27) and (28a) in (28) yields, upon the cancellation of the constant M, the *equation of motion of free material points in a gravitational field*

$$\frac{d}{dt}\left(\frac{\mathfrak{v}}{\sqrt{c^2 - v^2}}\right) = -\frac{c}{\sqrt{c^2 - v^2}}\operatorname{grad}c. \tag{29}$$

The energy equation, which connects the temporal increase of the energy E of the material point with the local temporal differential quotient of c with respect to time, is, by the way, joined to the momentum eq. (28) as a fourth equation. In a static field, where c depends only on position, E remains constant; in this case, according to (27a), the *energy equation* is

$$\frac{c^2}{\sqrt{c^2 - v^2}} = \text{const.} \tag{29a}$$

If one divides the equation of motion (29) by this constant expression, which then can be moved under the derivative, one obtains

$$\frac{d}{dt}\left(\frac{\mathfrak{v}}{c^2}\right) = -\frac{1}{c}\operatorname{grad} c \tag{29b}$$

[493] | *as the differential equation of motion of free material points in a static gravitational field*. The equations of motion (29) and (29b), incidentally, have been given first by A. Einstein,(11) and the energy eq. (29a) previously by M. Abraham.(8)

From it [eq, (29b)] one can derive the *theory of the free fall*. If v_0 is the initial velocity, and c_0 the speed of light appropriate to the initial location, then

$$v^2 = c^2\left(1 - \frac{c^2}{c_0^2} + c^2\frac{v_0^2}{c_0^4}\right), \tag{30}$$

and, in particular, if the initial velocity v_0 is equal to zero

$$v = c\sqrt{1 - \frac{c^2}{c_0^2}}. \tag{30a}$$

While c decreases from the initial value c_0, the velocity of the fall v increases at first, reaches a maximum

$$v_m = \frac{1}{2}c_0 \quad \text{for} \quad c = \frac{c_0}{\sqrt{2}}, \tag{30b}$$

and finally tends towards the limit c as c continues to decrease.

In order to determine the distance of fall as a function of time, c must be given as a function of z. In his note about the free fall,(9) M. Abraham puts

$$c^2 = c_0^2 + 2\mathfrak{g}z, \tag{31}$$

corresponding to a homogeneous mass-free field in his first theory. Then (30a) is satisfied by

$$\left.\begin{aligned}\frac{v}{c} &= \sin\left(\frac{gt}{c_0}\right)\\ \frac{c}{c_0} &= \cos\left(\frac{gt}{c_0}\right)\end{aligned}\right\} \qquad 0 \le t < \frac{\pi}{2}\frac{c_0}{g}. \tag{31a), (31b}$$

If one integrates the velocity of fall

$$-\frac{dz}{dt} = v = c_0 \cos\left(\frac{gt}{c_0}\right)\sin\left(\frac{gt}{c_0}\right) \tag{31c}$$

with respect to time, then the distance of fall becomes

$$-z = \frac{c_0^2}{2g}\sin^2\left(\frac{gt}{c_0}\right), \tag{31d}$$

which also satisfies (31). At the time $t = \pi c_0/2g$, i.e. for the distance of fall $c_0^2/2g$, c and v would both become zero. Of course, one cannot produce homogeneous fields of such extension. Practically, only times of fall have to be considered that are small | compared to the limiting time; then the above laws of fall become the Galilean ones. [494]

B. Caldonazzo[(23)] has concerned himself with the trajectories of freely point masses in homogeneous gravitational fields. He compares them to light rays whose paths follow from Fermat's principle of minimum light travel time,

$$\delta t = \delta\int\frac{ds}{c} = 0, \tag{32}$$

if c is given as a function of position. The differential equation for a *light ray trajectory* is, with $\mathfrak{t}$ denoting a tangential unit vector

$$\frac{d}{ds}\left(\frac{\mathfrak{t}}{c}\right) = \mathrm{grad}\left(\frac{1}{c}\right). \tag{32a}$$

On the other hand, the equation of motion (29b) implies the equation for the *trajectories of a material points* in a static gravitational field

$$\frac{d}{ds}\left(\frac{v\mathfrak{t}}{c^2}\right) = \frac{c}{v}\mathrm{grad}\left(\frac{1}{c}\right). \tag{32b}$$

An examination of the two differential equations (32a, b) shows that the second goes over into the first if one replaces v by c. From this remark follows an interesting relation between the trajectories of material points and those of light-points. Let us follow a material point P and a light-point L, which both emerge from O in the direction defined by the unit vector $\mathfrak{t}_0$. The initial speed of L is the speed of light c_0

associated with the point O, whereas we let that of P be equal to v_0. The trajectory of L is then uniquely determined by (32a), the one of P by (32b), where for v one has to use the function of position given by the energy eq. (30). Different initial speeds v_0 of P correspond to different trajectories; this set of trajectories has as the limiting curve the one corresponding to the initial speed c_0. But for $v_0 = c_0$ eq. (30) implies $v = c$, whence (32b) goes over into (32a). *Thus, the set of trajectories of material points shot from a point O in a given direction with different initial speed v_0, and moving in a static gravitational field, contain as a limiting curve $(v_0 = c_0)$ the light path that emerges from O in the same direction.*

Caldonazzo treats in more detail the case of the homogeneous gravitational field,
[495] where the equipotential surfaces, $c = \text{const.}$, are horizontal | planes. If the initial velocity was horizontal, it follows from (29b) that

$$v_x = \pm v_0 \frac{c^2}{c_0^2} \tag{33}$$

for the horizontal velocity component, and, hence, on account of (30), the vertical velocity component of the material points becomes

$$v_z = -\sqrt{v^2 - v_x^2} = -c\sqrt{1 - \frac{c^2}{c_0^2}}\,. \tag{33a}$$

The velocity components of the light-point are, according to the theorem just proved, obtained from this by setting $v_0 = c_0$:

$$c_x = \pm\frac{c^2}{c_0}, \tag{33b}$$

$$c_z = -c\sqrt{1 - \frac{c^2}{c_0^2}}\,. \tag{33c}$$

It follows that

$$v_z = c_z, \qquad v_x = \frac{v_0}{c_0}\,c_x. \tag{34}$$

Therefore, all the material points shot horizontally with different initial speeds v_0 have the same velocity of fall as the light-point, whereas the horizontal velocities are in the same ratio as the initial velocities. Thus all these points fall by equal distances during equal times; the horizontal projections of paths at equal time are, however, related to that of the light path as the initial velocities v_0 and c_0. Thus the trajectories of material points are obtained from the light curve to which they are tangent at the vertex by a contraction in the horizontal direction. Based on this result, Caldonazzo constructs, in a suitably chosen coordinate scale, the trajectories for the following three cases:

$$c = \sqrt{z}. \tag{I}$$

This corresponds to a homogeneous mass-free field in Abraham's first theory, where $c^2 = z$ should satisfy the Laplace equation (cf. (18a)). The light curves are cycloids. The trajectories resulting from them by contraction in the horizontal direction are so called "Fermat cycloids."

$$c = z. \tag{II}$$

Here, c itself satisfies Laplace's equation. According to the Einstein's equivalence hypothesis (III A), this should be the case in mass-free fields. | The light curves [496]
are circles, hence the trajectories of material points are ellipses.

$$c = z^2. \tag{III}$$

$\sqrt{c}$ satisfies Laplace's equation as demanded by the postulate of the weight of energy. The light-curves are elastic curves [*elastische Kurven*], the trajectories of material points are affine to them.

If one considers c as variable in the gravitational field, how are the electromagnetic field equations to be formulated? According to I. Ishiwara,(19) as follows:

$$\left.\begin{aligned}
\operatorname{curl}(\mathfrak{H}\sqrt{c}) &= \frac{\partial}{\partial t}\left(\frac{\mathfrak{E}}{\sqrt{c}}\right) + \rho\frac{\mathfrak{v}}{\sqrt{c}},\\
\operatorname{div}\left(\frac{\mathfrak{H}}{\sqrt{c}}\right) &= 0,\\
-\operatorname{curl}(\mathfrak{E}\sqrt{c}) &= \frac{\partial}{\partial t}\left(\frac{\mathfrak{H}}{\sqrt{c}}\right),\\
\operatorname{div}\left(\frac{\mathfrak{E}}{\sqrt{c}}\right) &= \frac{\rho}{\sqrt{c}}.
\end{aligned}\right\} \tag{35}$$

By retaining the usual expressions for the components of the electromagnetic tensor T^e, he obtains for its divergence, i.e., for the energy-momentum transfer of the electromagnetic field, the expression:

$$\rho\mathfrak{F}^e + \frac{\overset{e}{\eta}}{c}\operatorname{grad} c = \operatorname{div} T^e. \tag{35a}$$

The first term represents the electromagnetic four-force per unit volume; it corresponds to a transfer of momentum and energy from the electromagnetic field to the electrically charged matter. The second term, however, *shows a transfer of momentum and energy from the electromagnetic field to the gravitational field, which should take place everywhere where there exists electromagnetic energy*. This conception corresponds to the postulate of the weight of energy; and, according to it, the *electro-*

magnetic energy, just like the energy of matter, must *generate a gravitational field*, so that the field equation (23) is written as

$$\sqrt{c}\Box(\sqrt{c}) = 2(\eta^m + \eta^e). \tag{36}$$

Then, (cf. (23a)) the momentum and energy transferred from the gravitational field per unit volume and time becomes

$$-\left(\frac{\eta^m + \eta^e}{c}\right)\mathrm{grad}c = \mathrm{div}T^g. \tag{36a}$$

[497] | Here, the term containing η^m shows the transfer of momentum and energy to matter, the term containing η^e the transfer to the electromagnetic field; the latter is equal and opposite to the corresponding term appearing in (35a), justifying the above interpretation. Upon adding of the divergences of the tensors T^e and T^g it [η^e] cancels. The *total four-force acting on the unit volume of matter* is[4]

$$\rho\mathfrak{F}^e - \frac{\eta^m}{c}\mathrm{grad}c = \mathrm{div}(T^e + T^g). \tag{36b}$$

It is composed of the electromagnetic force acting on the charge, and of the gravitational force acting on the energy of matter.

The field equations (35) thus agree with the postulate of the weight of energy, and so fit into the theory presented here. The coupling of the two fields is, by the way, not a simple one. For, according to (35) the gravitational potential $\sqrt{c}$ influences the electromagnetic field; on the other hand, according to (36), the gravitational potential depends on the distribution of energy within the electromagnetic field.

D. Theories of Nordström and Mie

In Abraham's first theory (II B), the postulate of the constancy of the speed of light is renounced, and thus the validity of the theory of relativity is restricted to infinitely small spacetime regions. Abraham's second theory gives up the validity of the principle of relativity even in this restricted sense, because the two sides of its field equation (23) exhibit different behavior under Lorentz transformation.

The theories of gravitation of G. Nordström and G. Mie, however, view the speed of light as constant; their equations are invariant with respect to Lorentz transformations.

Following Abraham's first theory, G. Nordström(20) derives the gravitational tensor T^g from a scalar gravitational potential φ according to the scheme (16). This [potential] satisfies the field equation (18), in which ν denotes the "rest-mass density" of matter. Then, as in (II B, eq. (17), (17b)), on the basis of the conservation laws of momentum and energy there results,

$$\mu\mathfrak{F}^g = -\nu\,\mathrm{grad}\varphi \tag{37}$$

as the *gravitational force per unit of volume*.

We integrate over the volume of a moving body. Since | its volume elements experience a Lorentz contraction as a result of the motion, we have [498]

$$dV = dV_0\sqrt{1-\frac{v^2}{c^2}} \qquad (V_0 \text{ rest-volume}). \tag{38}$$

If, furthermore, one takes the rest-mass of the body to be

$$m_0 = \int \nu dV_0, \tag{38a}$$

then, in case the body is so small that the gravitational field is considered to be homogeneous over its extent, one obtains the expression

$$\mathfrak{K}^g = \int \mu \mathfrak{F}^g dV = -\frac{m_0}{c}\sqrt{c^2-v^2} \cdot \operatorname{grad}\varphi \tag{39}$$

for the *resultant gravitational force*. It becomes readily apparent that the appearance of the square root as a factor in the expression for the gravitational force is a necessary consequence following from the theory of relativity, because that theory cannot avoid assigning to the scalar φ, by means of (18), a ν invariant under Lorentz transformations. *In the theory of relativity the gravitational force of a moving body is therefore not proportional to its energy, but to its Lagrangian function* (cf. (20)):

$$\mathfrak{L} = -m_0 c\sqrt{c^2-v^2}. \tag{40}$$

At low speeds this means that not the sum but the *difference* of the *potential* and *kinetic energy should be the determining factor for the weight of a body.*

As c is constant, the gravitational potential φ can enter the Lagrangian function (40) only through the multiplicative constant m_0. This must take place in such a way that in the law of conservation of momentum formulated in the Lagrangian manner (see (28)),

$$\frac{d\mathfrak{G}}{dt} = \operatorname{grad}\mathfrak{L} = \frac{\partial\mathfrak{L}}{\partial\varphi}\operatorname{grad}\varphi, \tag{40a}$$

the right-hand side, i.e., the gravitational force

$$\mathfrak{K}^g = \frac{\partial\mathfrak{L}}{\partial\varphi}\operatorname{grad}\varphi = -\frac{dm_0}{d\varphi}c\sqrt{c^2-v^2} \cdot \operatorname{grad}\varphi, \tag{40b}$$

agrees with (39). Therefore, for the rest-mass, the differential equation

$$\frac{dm_0}{d\varphi} = \frac{m_0}{c^2} \tag{40c}$$

must apply, whose integral is

$$m_0 = \frac{M}{c} e^{\frac{\varphi}{c^2}}; \tag{41}$$

[499] | M denotes a “mass-constant” which is independent of φ. Thus, the gravitational potential enters exponentially into the *Lagrangian function of the first theory of Nordström,*

$$\mathfrak{L} = -M e^{\frac{\varphi}{c^2}} \cdot \sqrt{c^2 - v^2}, \tag{41a}$$

and enters similarly into the *expressions for momentum* and *energy*

$$\mathfrak{G} = \frac{\partial \mathfrak{L}}{\partial v} \frac{\mathfrak{v}}{v} = \frac{M}{\sqrt{c^2 - v^2}} e^{\frac{\varphi}{c^2}} \cdot \mathfrak{v}, \tag{41b}$$

$$E = v \frac{\partial \mathfrak{L}}{\partial v} - \mathfrak{L} = \frac{Mc^2}{\sqrt{c^2 - v^2}} \cdot e^{\frac{\varphi}{c^2}}, \tag{41c}$$

and into the *expression for the gravitational force* (40b)

$$\mathfrak{K}^g = -\frac{M}{c^2} e^{\frac{\varphi}{c^2}} \sqrt{c^2 - v^2} \cdot \operatorname{grad}\varphi. \tag{41d}$$

The equation of motion of material points, or equivalent bodies, in a gravitational field

$$\frac{d\mathfrak{G}}{dt} = \mathfrak{K}^g \tag{42}$$

takes, according to (41b, d), the form

$$\frac{d}{dt}\left(e^{\frac{\varphi}{c^2}} \cdot \frac{\mathfrak{v}}{\sqrt{c^2 - v^2}}\right) = -\frac{1}{c^2} \cdot e^{\frac{\varphi}{c^2}} \cdot \sqrt{c^2 - v^2} \cdot \operatorname{grad}\varphi. \tag{42a}$$

In a static field we have the energy integral (cf. (41c))

$$\frac{e^{\frac{\varphi}{c^2}}}{\sqrt{c^2 - v^2}} = \text{const.}, \tag{42b}$$

so that in Nordström’s theory, the *equation of motion of material points in a static field* is:(22)

$$\frac{d\mathfrak{v}}{dt} = -\left(1 - \frac{v^2}{c^2}\right) \operatorname{grad}\varphi. \tag{42c}$$

It is apparent that in a homogeneous gravitational field—in contrast with (II C)—the greater the velocity of horizontally projected bodies, the slower they fall. The light-curve is, however, here too to be considered as a limiting curve of the family of trajectories; because for $v = c$, the acceleration is zero, thus the motion is uniform and rectilinear, like that of a light-point under the presupposition of the constancy of the speed of light.

For *free fall* in a homogeneous mass-free field, the equation of motion

$$\frac{dv}{dt} = g - \frac{gv^2}{c^2} \tag{42d}$$

applies. | It agrees with the differential equation that one uses in classical mechanics [500]
for the fall in air under the assumption of a law of friction proportional to the square [of the speed]; therefore its integration poses no difficulties and becomes, if one sets the initial velocity equal to zero,

$$v = c\tanh\left(\frac{gt}{c}\right). \tag{42e}$$

If the field extended sufficiently far, the speed would asymptotically approach the speed of light.

M. Behacker(25) has treated horizontal projection [of bodies] and planetary motion on the basis of Nordström's first theory. The deviations from the laws of classical mechanics are here too of second order in $v : c$.

The theory of G. Mie,(21) though developed independently, is closely related to Nordström's first theory. It forms part of an extensive investigation of the "foundations of a theory of matter," for whose presentation this is not the appropriate place. Within matter, Mie differentiates between two gravitational four-vectors, which, however, coincide in the case of an "ideal vacuum." Here, as with Abraham, the differential equations of longitudinal waves apply. The gravitational waves emitted by an oscillating mass-particle are treated in more detail by Mie. He further emphasizes that, from the point of view of the theory of relativity, one necessarily arrives at the notion that the weight of a body is not to be set proportional to its energy, but to its Lagrangian function. Accordingly, the kinetic energy of the thermal motion of the molecules makes a negative contribution to the weight. Since, on the other hand, in the theory of relativity the theorem of the inertia of energy is valid, the gravitational mass of a warm body is thus not precisely proportional to its inertial mass, but with increasing temperature the quotient of gravitational and inertial mass decreases. However, the proof of the decrease demanded by Mie's theory would require pendulum measurements with an accuracy of the order of 10^{-11}; this theoretical deviation from the proportionality law is not subjectable to experiment.

Greater difficulties arise for the Mie-Nordström theory from observations by Southerns, which demonstrate the constancy of the quotient of inertial and gravitational mass during radioactive transformations (I C). These observations can only be

brought into agreement with that theory through rather artificial assumptions. One could for example imagine that the entire energy loss during the transformation |
[501] takes place at the expense of only the electric potential energy, so that the kinetic (magnetic) energy has the same value before and after the transformation of the uranium atom. Or, one can assume with Kretschmann[36] that the total energy emitted from the uranium atom in the form of heat, β - and γ -particles does not come from those particles that later form the lead atom, but from the eight escaped helium atoms. Then, however, helium would have to show a correspondingly greater deviation from the proportionality law.

G. Mie also investigated the relation between the gravitational potential and electromagnetic processes. An electromagnetic field possesses a gravitational mass, which is proportional to its Lagrangian function, i.e. proportional to the difference of its electric and magnetic energy. A plane electromagnetic wave, in which the electric and the magnetic energy density is known to have the same value, accordingly has no weight in this theory (corresponding to the rectilinear propagation of light). An electrostatic field, however, possesses a positive weight and a magnetostatic field a negative weight. On the other hand, the gravitational potential enters exponentially into the electromagnetic quantities, as well as into the mechanical ones. Nevertheless, the value of the potential in a system escapes detection by observers belonging to that system. This is expressed by Mie's theorem of the "*relativity of the gravitational potential*": If two physical systems differ merely by the value of the gravitational potential, then this does not have the least effect on the size and the form of the electrons and of the other material elementary particles, on their charge, their laws of oscillation and motion, and on the velocity of light, indeed on all physical relations and processes.

E. Nordström's Second Theory

With the intention of satisfying the law of the proportionality of the gravitational and inertial mass, as far as it is possible within the framework of the theory of relativity, G. Nordström[26] later made a change in his theory. Before we turn to a discussion of this change, we must briefly return to the properties of the world tensors (I B), and mention several theorems of von Laue regarding "complete static systems."

We form the "*diagonal sum*" of a world tensor T

$$D = T_{xx} + T_{yy} + T_{zz} + T_{uu}; \tag{43}$$

[502] | since the components entering this sum transform like the square of the four coordinates, the diagonal sum is an invariant; it is frequently called the "Laue scalar."

The diagonal sum of the *electromagnetic world tensor* T^e is identically equal to zero:

$$D^e = T^e_{xx} + T^e_{yy} + T^e_{zz} + T^e_{uu} = 0. \tag{43a}$$

In contrast, our *gravitational tensor* T^g yields, according to (16), a diagonal sum different from zero:

$$D^g = \left(\frac{\partial\varphi}{\partial x}\right)^2 + \left(\frac{\partial\varphi}{\partial y}\right)^2 + \left(\frac{\partial\varphi}{\partial z}\right)^2 + \left(\frac{\partial\varphi}{\partial u}\right)^2. \tag{43b}$$

The matter is the carrier of a third world tensor T^m; its last component is the energy density of matter. For a body consisting of discrete and freely movable mass particles, the momentum density and the kinetic stresses determine the remaining components of the "*material tensor*" T^m. In a continuously connected body, elastic stresses enter in addition into the consideration, as they also cause an energy flux and thus a momentum density. It appears useful to include in T^m the "stress tensor." The diagonal sum of the material tensor

$$D^m = T^m_{xx} + T^m_{yy} + T^m_{zz} + T^m_{uu} \tag{43c}$$

then equals the sum of the three (kinetic and elastic) principal stresses plus the total energy density (η^m) of the matter.

Now, if

$$T = T^m + T^e + T^g \tag{44}$$

is the *total ten-tensor, resultant of the material, the electromagnetic and of the gravitational tensor*, then the *laws of conservation of momentum and energy* can be summarized as (cf. (5))

$$0 = \operatorname{div} T. \tag{44a}$$

Namely the momentum and energy extracted from the electromagnetic field and from the gravitational field are transformed into momentum and energy of matter.

By a "*complete static system*," one understands, according to M. Laue, an isolated physical system that is in equilibrium in an appropriately chosen system of reference Σ_0. Therefore, in Σ_0 the components

$$T^0_{xu}, \quad T^0_{yu}, \quad T^0_{zu}$$

| of the resultant tensor, which determine the energy flux and the momentum density, [503]
are all zero; one can derive from this[9] that in Σ_0 the volume integrals of the resulting normal stresses are equal to zero:

$$\int dV_0 T^0_{xx} = \int dV_0 T^0_{yy} = \int dV_0 T^0_{zz} = 0. \tag{45}$$

If one is only concerned about the temporal mean values, then a system in static equilibrium can also be taken as a completely static system. It suffices that the tempo-

---

9 M. Laue, *Das Relativitätsprinzip*, 2nd. ed., p. 209.

ral mean value of the resultant energy flux vanish everywhere. Thus, for example, a hot gas in static equilibrium, together with its container, forms a complete static system. For the measurable stresses, the eqs. (45) apply; the negative contributions arising from the kinetic pressure of the gas, and the positive, arising from the stresses in the walls of the container, cancel each other in the integrals extended over the entire system.

Because it is an invariant, the diagonal sum (43) of the resultant world tensor does not depend on the choice of the system of reference. Its value in a complete static system can thus be determined by referring it to Σ_0:

$$D = D^0 = T^0_{xx} + T^0_{yy} + T^0_{zz} + T^0_{uu}. \tag{45a}$$

In view of (45), the volume integral yields

$$\int dV_0 D = \int dV_0 T^0_{uu} = \int dV_0 \eta_0 = E_0, \tag{45b}$$

where E_0 denotes the *total rest-energy* of the system.

After these preparations, we turn to Nordström's second theory of gravitation. It too is based on the tensor-scheme of (16); however, the following takes the place of the field equation (18),

$$\varphi \Box \varphi = D^m. \tag{46}$$

Since the diagonal sum D^m of the material tensor, as well as φ and the operator $\Box$, are four-dimensional scalars, this approach corresponds to the relativistic scheme. The gravitational force per unit volume now follows from (17b):

$$\mu \mathfrak{F}^g = -\frac{D^m}{\varphi} \cdot \operatorname{grad} \varphi. \tag{46a}$$

The integration over the volume of a body moving with respect to Σ_0, taking into account the Lorentz contraction (38), yields the gravitational force acting on the
[504] body. |

$$\mathfrak{K}^g = -M\sqrt{c^2 - v^2} \cdot \operatorname{grad} \varphi, \tag{46b}$$

where

$$Mc = \int \frac{dV_0 D^m}{\varphi} = \int \frac{dV_0 D^m_0}{\varphi_0} \tag{46c}$$

denotes the gravitational mass of the body for the case of rest $(v = 0)$. Indeed, on the one hand, according to (46b), the gravitational force on the resting body is equal to the negative gradient of the potential φ multiplied by Mc. On the other hand, when rest reigns in Σ_0, it follows from (46) that

$$\Delta\varphi_0 = \text{div grad}\varphi_0 = \frac{\overset{m}{D_0}}{\varphi_0}.$$

Therefore, according to Gauss' law, the force flux through a surface f_0 enclosing the body, which determines the attracting mass, becomes:

$$\int df_0 \frac{\partial\varphi_0}{\partial n} = \int \frac{dV_0 \overset{m}{D_0}}{\varphi_0} = Mc. \tag{46d}$$

The gravitating mass Mc, and thus also M, is considered to be a *constant*, i.e. as independent of φ, in Nordström's second theory.

Now, what relation exists between the gravitational mass and the rest-energy for a complete static system? Equation (45b) yields:

$$E_0 = \int dV_0 D_0. \tag{47}$$

Taking into consideration (43a, b), one obtains for the diagonal sum of the resultant world tensor, since equilibrium reigns in Σ_0,

$$D_0 = \overset{m}{D_0} + (\text{grad}\varphi_0)^2. \tag{47a}$$

And since, furthermore,

$$\varphi_0\Delta\varphi_0 + (\text{grad}\varphi_0)^2 = \text{div}(\varphi_0\text{grad}\varphi_0)$$

is an identity, the field equation (46) can then be brought into the form

$$\text{div}(\varphi_0\text{grad}\varphi_0) = D_0 \tag{47b}$$

and from (47) it follows, on the basis of Gauss' theorem, that

$$E_0 = \int dV_0 D_0 = \int df_0 \varphi_0 \frac{\partial\varphi_0}{\partial n}, \tag{47c}$$

where the integration is to be carried out over a surface wholly enclosing the complete static system and its gravitational field. However, on such a surface, far removed from the system, one has to set for φ_0, the potential φ_a of the external masses not belonging to the system, | which is constant there. Namely, if φ_a were not constant [505]
on f_0, the external field would superpose on the system's own; the energy of its own field could then not be separated from the external field, and the system could thus not be considered isolated (i.e., enclosed by the surface f_0). Now, it follows from (47c) and (46d) that

$$E_0 = \varphi_a \int df_0 \frac{\partial\varphi_0}{\partial n} = \varphi_a Mc. \tag{48}$$

Accordingly, the total *rest-energy of the complete static system is equal to its gravitational mass multiplied by the external gravitational potential*. On the other hand, according to the theorem of the inertia of energy (8a),

$$m_0 = \frac{E_0}{c^2} = \varphi_a \frac{M}{c} \tag{48a}$$

is the *inertial rest-mass*.

In Nordström's second theory, the law of proportionality between inertial and gravitational mass applies for a complete static system, at least for the rest-masses. Of course, for moving bodies, the postulate of the weight of energy is not satisfied in Nordström's second theory any more than in his first, since, in the expression (46b) for the gravitational force, the root factor characteristic of the theory of relativity, and entering into the denominator of the expression for the energy, appears again in the numerator. However, admittedly one cannot determine the weight of a moving body with a scale; thus, as far as weighing is concerned, the theorem of the weight of energy therefore applies in Nordström's theory, and accordingly also the theorem that weight is conserved. However, the strict validity of these theorems, and thus also that of the law of proportionality of inertial and gravitational mass, fails when the masses are in motion. For example, in pendulum measurements, one would expect deviations of the order of $(v/c)^2$, which are however unmeasurable of course.

But what is now the situation in an isolated system in a state of statistical equilibrium, e.g., in a hot gas? The thermal motion of the molecules causes here, too, a reduction of weight corresponding to the negative kinetic stresses entering into D^m, and thus also into the expression for the gravitational mass (cf. (46d)). This, however, is compensated by the positive contributions which the stresses in the wall of the vessel contribute to D^m, and thus to the weight. Thus, for Nordström, the proportionality of weight and energy comes about merely through a compensation of the contributions of the individual parts of the complete static system. | [506] It is in no way based on a fundamental property of matter or energy.

A complete system of sufficiently small dimensions can be considered as equivalent to a material point. The motion in a given gravitational field is then obtained from Lagrange's equations (40a)

$$\frac{d\mathfrak{G}}{dt} = \frac{d\mathfrak{L}}{\partial\varphi}\mathrm{grad}\varphi = \mathfrak{K}^g. \tag{49}$$

As a comparison with (46b) reveals, one has to set for the *Lagrangian function*

$$\mathfrak{L} = -M\varphi\sqrt{c^2 - v^2}. \tag{49a}$$

This Lagrangian of Nordström's second theory turns into the one (41a) of his first theory, if φ is replaced by the exponential function $e^{\frac{\varphi}{c^2}}$. However, through this substitution, the field equation (46) does not change into agreeing with (18) of the first theory; therefore, the two theories are not equivalent.

The momentum and energy can be derived from the Lagrangian function (49a) in the usual manner:

$$\mathfrak{G} = \frac{\partial \mathfrak{L}}{\partial v}\frac{\mathfrak{v}}{v} = \frac{M\varphi}{\sqrt{c^2 - v^2}} \cdot \mathfrak{v}, \tag{49b}$$

$$E = v\frac{\partial \mathfrak{L}}{\partial v} - \mathfrak{L} = \frac{Mc^2\varphi}{\sqrt{c^2 - v^2}}. \tag{49c}$$

Through substitution of (49b) into (49) one obtains the *equations of motion of material points* and of the equivalent complete systems

$$\frac{d}{dt}\left(\frac{\varphi \mathrm{v}}{\sqrt{c^2 - v^2}}\right) = -\sqrt{c^2 - v^2} \cdot \operatorname{grad}\varphi. \tag{50}$$

In a static gravitational field, the energy (49c) is constant:

$$\frac{\varphi}{\sqrt{c^2 - v^2}} = \text{const.} \tag{50a}$$

Therefore, the equation of motion can be brought into the form[31]

$$\frac{d\mathfrak{v}}{dt} = -\left(\frac{c^2 - v^2}{\varphi}\right)\operatorname{grad}\varphi. \tag{50b}$$

For a homogeneous mass-free gravitational field, whose lines of force are parallel to the negative z-axis, one must set, corresponding to the field equation (46),

$$\varphi = \varphi_0 + gz. \tag{50c}$$

| The free fall from the rest position $z = 0$ in this field can be treated on the basis of the energy equation (50a), which yields [507]

$$\varphi = \frac{\varphi_0}{c} \cdot \sqrt{c^2 - v^2}. \tag{50d}$$

It is satisfied by

$$z = -\frac{\varphi_0}{g}\left(1 - \cos\left(\frac{gct}{\varphi_0}\right)\right); \tag{51}$$

because according to (50c)

$$\varphi = \varphi_0 \cos\left(\frac{gct}{\varphi_0}\right), \tag{51a}$$

and, on the other hand, the velocity of fall is

$$\upsilon = -\frac{dz}{dt} = c\sin\left(\frac{gct}{\varphi_0}\right), \tag{51b}$$

so that (50d) is indeed satisfied.

G. Nordström,[31] incidentally, also treats oblique projection [of bodies] and planetary motion; the area law proves to be valid. The deviations from the laws of classical mechanics are minute.

In Nordström's theory, the rest-energy is proportional to the external gravitational potential (cf. (48)). In Abraham's theory, a similar behavior was found; the rest-energy (cf. (21a)) was proportional to the speed of light, which there played the role of the gravitational potential. Thus, in either theory, the rest-energy of a body decreases as a result of the approach of external masses. The rest-mass, in contrast, exhibits a different behavior in the two theories; for Abraham (cf. (21b)) the rest-mass is inversely proportional to c, for Nordström (cf. (48a)) it is proportional to φ_a. Hence, for Abraham, the inertial mass of a body increases upon the approach of an external body, while, for Nordström, it decreases.

Incidentally, according to Nordström, the units of time (τ) and length (λ) should depend in the following way on the gravitational potential:

$$\tau\varphi = \text{const.}, \tag{52}$$

$$\lambda\varphi = \text{const.}, \tag{52a}$$

i.e., the rate of a portable clock and the length of a portable measuring rod should be inversely proportional to the gravitational potential. Thus, local spatial and temporal measurements do not allow the construction of the universal system of reference in which light should propagate rectilinearly. This conception of space and time can
[508] hardly be made compatible with the theory of relativity of Minkowski. | However, it can be readily fitted into the Einstein-Grossmann "generalized theory of relativity," as has been shown by A. Einstein and A. D. Fokker.[34]

We must refrain here from entering into Nordström's five-dimensional interpretation of his theory.[33] [37]

F. Kretschmann's Theory

Among the newest relativistic theories of gravitation, the one by E. Kretschmann[36] deserves to be mentioned even if only briefly. This theory is to be counted among the scalar theories, as its assumes that the gravitational force is determined by the gradient of an "aether pressure" p, which, in a vacuum, satisfies the equation

$$\Box p = 0.$$

Matter is assumed to consist of elementary drops, which obey the equations of state of "ideal fluids of smallest compressibility and least expendability" developed by E. Lamla in his inaugural dissertation "Über die Hydrodynamik des Relativitätsprinzips." These drops are assumed to move irrotationally and each to carry a positive electric elementary charge. The divergence of the gradient of p in the interior of the drops, which determines the attracting mass, turns out to be proportional to the square of their acceleration. On the other hand, the attracted mass of a body is proportional to the sum over the volume of its elementary drops. In order to obtain the proportionality of the attracted, the attracting and of the inertial mass, the author assumes not only that the drops are all of the same kind, but that they also move with the same mean velocity and acceleration; then, of course, all three masses simply become proportional to the number of the particles. Following A. Korn and H. A. Lorentz, the author explains, that such an energy equilibrium of the elementary drops is produced by means of a universal radiation state, or state of oscillation, which is supposed to immediately smooth out disturbances in the energy equilibrium.

The contrast between this theory and the other theories of gravitation discussed here is strikingly clear. The latter introduce merely mechanical and energetic quantities, since the speed of light, too, is a quantity of relativistic mechanics. Kretschmann's theory, in contrast, is based on quite particular concepts, which are virtually without significance for the final result; for not the particular properties, but only the number of | particles is to be considered in the derivation of the proportional- [509]
ity of inertial and gravitational mass.

III. TENSOR THEORIES

A. Einstein's Theory of the Static Gravitational Field

In I C we discussed the relation between gravity,[2] inertia, and energy, and the necessity of assigning to them a place in the worldview of modern physics. These relations also formed the starting point of the investigations of A. Einstein.

His first paper(5) is based on the so called *equivalence hypothesis*: "Equivalence exists between two systems of reference of which one is at rest in a homogeneous gravitational field and the other is uniformly accelerated in a field free of gravitation." Indeed, in classical mechanics, such an equality exists because the inertial force under uniform acceleration is equivalent to a constant gravitational force; the equivalence hypothesis results here in the identity of inertial and gravitational mass. But it is questionable whether the equivalence hypothesis can be maintained for other physical processes, namely for those that satisfy the principle of relativity of 1905; if this were the case, then from this and from the theorem of the inertia of energy, valid in that theory, the theorem of the weight of energy would follow immediately. Without a critical examination of whether his earlier principle of relativity is compatible with the new equivalence hypothesis, Einstein connects these two trains of thought and arrives thus at the following remarkable result:

The gravitational potential influences the speed of light; if the values of the speeds of light, c and c_0, correspond to the potentials, φ and φ_0, then

$$c = c_0\left(1 + \frac{\varphi - \varphi_0}{c^2}\right). \tag{53}$$

By application of Huygens' principle this implies a curvature of the light rays grazing the surface of the Sun; the total deviation of such a ray would amount to 0.83 seconds of arc, by which the angular distance of a star appears to be increased with respect to the center of the Sun. Einstein considers it possible to observe this deviation on the occasion of an eclipse of the Sun. |
[510]

Furthermore, the gravitational potential changes the frequency of periodic processes as determined by the following formula

$$\frac{\nu_0 - \nu}{\nu_0} = \frac{\varphi_0 - \varphi}{c^2}. \tag{53a}$$

If φ refers to the surface of the Sun and φ_0 to the Earth, then the right-hand side is equal to $2 \cdot 10^{-6}$. Thus, the frequency ν of the oscillations of light should be somewhat smaller on the surface of the Sun than the frequency ν_0 of the corresponding spectral line of a terrestrial light source: i.e., compared to the terrestrial lines, the solar ones should be displaced to the right-hand side of the spectrum. Astrophysicists have indeed found displacements of the Fraunhofer lines in this sense, but have mostly ascribed them to pressure effects.

In that first study, Einstein had investigated the effect of the gravitational field on radiation phenomena without, however, making use of the connection found between the gravitational potential and the speed of light for the theory of the gravitational field itself. Only after the first publication by Abraham did Einstein also turn to the problem of gravitation.(11) There, again, he started from the equivalence hypothesis and derived the equations of motion ((29), (29b)) of material points in static gravitational fields. He also concluded from that hypothesis that the speed of light c in mass-free static fields must satisfy Laplace's equation. Shortly thereafter,(12) however, he convinced himself that in order to avoid contradictions with the law of conservation of momentum, one has to begin from Laplace's equation for $\sqrt{c}$ rather than for c, and that, hence, the equivalence hypothesis does not form a firm foundation for the theory of the gravitational field. This led to the necessity for a new foundation for the theory. M. Abraham (II C) made the theorem of the weight of energy the starting point of his second theory, and derived from it the field equations and the equations of motion. For static fields, this theory coincides with Einstein's. The essential difference appears only upon the extension to time varying fields. Abraham's theory considers such fields as still being determined solely by the four derivatives of the velocity of light, whereas Einstein's theory derives the dynamic field from a tensor potential.

B. The Generalized Theory of Relativity of A. Einstein and M. Grossmann

The basic idea of the tensor theory of the gravitational field can | be understood as follows. The energy density, which in a static field is determined by the divergence of the gradient of the gravitational potential, plays in the theory of relativity merely the role of one component of the resulting world tensor T; it is joined by nine other tensor components which characterize the energy flux and the stresses. The tensor theory assumes that, like the energy density (T_{44}), the remaining nine components $T_{\mu\nu}$ $(\mu, \nu = 1, 2, 3, 4)$ generate gravitational fields whose potentials $g_{\mu\nu}$ form a ten-tensor themselves. With the introduction of such a tensor potential it aims to give the differential equations of the gravitational field a form satisfying the scheme of the principle of relativity. That it appears possible to fit tensor theories of gravitation into the framework of Minkowski's conception of spacetime has been shown by G. Mie.(28) However, the theory of gravitation sketched by A. Einstein and M. Grossmann(24) supersedes the framework of the earlier theory of relativity by closely relating these tensor potentials $g_{\mu\nu}$ to a generalized relativistic spacetime doctrine. [511]

In Minkowski's four-dimensional representation of the theory of relativity, the square of the four-dimensional distance between two neighboring spacetime points was given by the following quadratic differential form:

$$ds^2 = dx^2 + dy^2 + dz^2 - c^2dt^2, \tag{54}$$

in which c, the speed of light, was constant. In a static gravitational field, the very expression (54) should, according to Einstein, express the "natural distance" ds of two neighboring world points, but where c is now a function of x, y, z. Its application to the general case of a dynamic gravitational field would, however, imply a preferential treatment of the time coordinate $(x_4 = t)$ over the spatial coordinates $(x_1 = x, x_2 = y, \ x_3 = z)$, which would not be reconcilable with the relativistic ideas about space and time. It appears natural to replace the special quadratic form (54) by the most general homogeneous function of second degree in the coordinate differentials:

$$ds^2 = \sum_{\mu\nu} g_{\mu\nu}dx_\mu dx_\nu \qquad (\mu, \nu = 1, 2, 3, 4). \tag{55}$$

This form was used by E. and G. for the natural distance between two neighboring world points. The system of the ten coefficients $g_{\mu\nu}$ in the form (55) form a world tensor, and it is said to be identical to the tensor potential of the dynamic gravitational field, similar to the way c determines the gravitational potential in the static case. | [512]

According to (55) the natural length of a portable, infinitely small measuring rod is not determined solely by the coordinate differentials dx_1, dx_2, dx_3 of its end points; rather, the six potentials

$$g_{11}, \qquad g_{22}, \qquad g_{33}, \qquad g_{12} = g_{21}, \qquad g_{23} = g_{32}, \qquad g_{31} = g_{13}$$

enter into it, and, in particular, in a way that depends on the direction of the measuring rod. That means that apparently rigid bodies are stretched and twisted in a gravitational field. The measure of time too is influenced by the gravitational potential g_{44} (this occurs already in the static field) in such a way that the natural distance of two neighboring points in time, measured with the aid of a portable clock, is different from the differential dx_4 of the time coordinate. Thus, the coordinates x_1, x_2, x_3, x_4 have no direct physical meaning. In order to determine from their differentials the natural distance between neighboring spacetime points, the values of the ten gravitational potentials $g_{\mu\nu}$ must be known. Regarding, by the way, the three potentials,

$$g_{14} = g_{41}, \qquad g_{24} = g_{42}, \qquad g_{34} = g_{43},$$

they arise, for example, if the system of reference rotates in a static gravitational field, and then determine the velocity of the point in question. The complete system of the $g_{\mu\nu}$ characterizes the state of deformation of four-dimensional space.

In the same way as the invariant differential form (54) is related to the group of rotations in four dimensions (Lorentz transformations), and hence to the vector calculus of Minkowski's theory of relativity, so the invariance of the more general differential form (55) is related to a more general group of transformations. The study of the behavior of the different geometric objects (vectors, tensors) with respect to these transformations forms the mathematical basis of the "generalized theory of relativity." The authors of the *Entwurf* found the required mathematical tools already fully developed in the "absolute differential calculus" of G. Ricci and T. Levi-Civita. Based on this, in the second part of the *Entwurf*, M. Grossmann outlines the fundamental concepts of the four-dimensional vector calculus, which corresponds to that general group of transformations. This is not the place to enter into the mathematical aspects of the *Entwurf*. It may, however, be noted that Grossmann designates the "four-vectors" of the usual theory of relativity as "first rank tensors," whereas by "second rank tensors" he understands the quantities which by us are called simply
[513] "tensors." |

In the first, physical, part of the *Entwurf*, written by A. Einstein, the equations of motion of material points are derived in a generally covariant form from Hamilton's principle

$$\delta\int ds = 0. \tag{56}$$

The gravitational force appears as dependent on the forty derivatives of the ten potentials $g_{\mu\nu}$; here, however, the so called "fictitious forces of relative motion" are counted among the gravitational forces.

Just like the equations of motion, the differential equations of the electromagnetic field are also brought to a generally covariant form. Yet, the attempt to develop generally covariant equations for the gravitational field fails. The *field equations* put down by Einstein read as follows:

$$\sum_{\alpha\beta\mu} \frac{\partial}{\partial x_\alpha}\left(\sqrt{-g}\gamma_{\alpha\beta} g_{\sigma\mu} \frac{\partial \gamma_{\mu\nu}}{\partial x_\beta}\right) = S^m_{\sigma\nu} + S^g_{\sigma\nu}. \tag{57}$$

Here, g is the determinant of the $g_{\mu\nu}$, $\gamma_{\mu\nu}$ is the cofactor (*adjungierte Unterdeterminante*) of $g_{\mu\nu}$ divided by g. On the right-hand side are quantities that are linear functions of the components of the material tensor T^m and of the gravitational tensor T^g:

$$S^m_{\sigma\nu} = \sqrt{-g} \cdot \sum_\mu g_{\sigma\mu} T^m_{\mu\nu}, \tag{57a}$$

$$S^g_{\sigma\nu} = \sqrt{-g} \cdot \sum_\mu g_{\sigma\mu} T^g_{\mu\nu}. \tag{57b}$$

If one places the quantities $S^g_{\sigma\nu}$:

$$S^g_{\sigma\nu} = \sqrt{-g} \cdot \left\{\sum_{\beta\tau\rho} \gamma_{\beta\nu} \frac{\partial g_{\tau\rho}}{\partial x_\sigma} \frac{\partial \gamma_{\tau\rho}}{\partial x_\beta} - \frac{1}{2} \sum_{\alpha\beta\tau\rho} \delta_{\sigma\nu} \gamma_{\alpha\beta} \frac{\partial g_{\tau\rho}}{\partial x_\alpha} \frac{\partial \gamma_{\tau\rho}}{\partial x_\beta}\right\}, \tag{57c}$$

(in which one has to set $\delta_{\sigma\nu} = 1$, for $\sigma = \nu$, $\delta_{\sigma\nu} = 0$ for $\sigma \neq \nu$) on the left hand side of (57), then there arise ten second order partial differential equations for the ten quantities $\gamma_{\mu\nu}$, respectively $g_{\mu\nu}$, in the right-hand side of which the components of the material tensor T^m enter as field generating quantities. If an electromagnetic field is present, then the components of the electromagnetic tensor T^e are of course to be introduced in the same manner.

With regard to the invariance properties of his gravitational field equations, Einstein still appears to have hoped, during the writing of the *Entwurf*, to obtain covariance, if not for the more general group of transformations associated with the form (55), but at least for a group encompassing acceleration transformations; thereby, his earlier | "equivalence hypothesis" was to be supported mathematically. In his Vienna [514] lecture,(27) however, he states that it is probable, and later(29) that it is certain, that these field equations are only covariant with respect to linear transformations. Lately however, with the aid of M. Grossmann, he was able to prove(35) that only the laws of conservation of momentum and energy are responsible for restricting the diversity of "allowed" transformations. These conservation laws, which we expressed symbolically by

$$\operatorname{div}(T^m + T^g) = 0$$

become in the generalized theory of relativity, upon the introduction of the quantities (57a, b):

$$\sum_\nu \frac{\partial}{\partial x_\nu}(S^m_{\sigma\nu} + S^g_{\sigma\nu}) = 0. \tag{58}$$

Those four equations, together with the field equations (57), yield four third order partial differential equation

$$\sum_{\alpha\beta\mu\nu} \frac{\partial^2}{\partial x_\nu \partial x_\alpha}\left(\sqrt{-g}\gamma_{\alpha\beta} g_{\sigma\mu} \frac{\partial \gamma_{\mu\nu}}{\partial x_\beta}\right) = 0. \tag{58a}$$

Only such coordinate systems are allowed that are characterized by $g_{\mu\nu}$ satisfying this system of equations; only transformations that transform such "adapted" systems of reference into each other are "allowed." For all allowed transformations, the field equations (57) now turn out[(35)] to be covariant. A covariance in a further sense can hardly be demanded.

Although from a mathematical point of view the question appears to have been settled, the physicist still wishes to obtain an insight into the extent of the group of "allowed" transformations. Do these contain, besides the linear ones, still other real transformations? And do these have a physical significance, perhaps as transformations describing acceleration or rotation? Only in this case would one be justified in speaking of a "generalized theory of relativity" in the sense that the equivalence of different systems of reference, which the principle of relativity of 1905 postulates for systems in translatory motion with respect to each other, is now extended to such systems that are in accelerated or rotational motion with respect to one another. Such an extension of the relativity of motion does not appear to be achievable.

Also, the physical significance of covariance with respect to linear transformations is not sufficiently discussed in the *Entwurf*. [515] | According to G. Mie,[(28)] it expresses the following theorem: "The observable laws of nature do not depend on the absolute values of the gravitational potentials $g_{\mu\nu}$ at the location of the observer." G. Mie showed already earlier[(21)] that this "*theorem of the relativity of the gravitational potential*" is valid in his own theory (II D). However, in Mie's theory, based on the spacetime doctrine of Minkowski, the space and time measurements as well as the dimensions and periods of oscillation of elementary particles remain unchanged, even in a gravitational field. Einstein, in contrast, and also Nordström (II E), achieve the independence of the physical processes from the values of the potentials only because, together with the units of length and time, the dimensions and the periods of oscillation of the atoms and electrons depend on the gravitational potentials $g_{\mu\nu}$.

A. Einstein places such an importance on the circumstance that his gravitational equations are invariant with respect to linear orthogonal substitutions that, in his Vienna lecture,[(27)] he believed he was allowed to leave theories unmentioned whose field equations do not have such a covariance. The significance of this transformation group rests on the fact that it contains the transformations; in the early theory of relativity the covariance with respect to this group gave expression to the equivalence of systems of reference in translatory motion with respect to each other. Is this presently also the case in the "generalized theory of relativity"? Does the covariance of the field equations with respect to linear orthogonal transformations imply that in a finite system of mutually gravitating bodies the course of the relative motions is not altered by a uniform translation of the entire system? That this is so has so far not been proven. Such a proof may already be impossible to construct for the reason that the concept of "uniform motion" of a finite system is completely up in the air in the new theory of

relativity. Since the "natural" space and time measurement is influenced by the values of the local potential, observers at different locations in the gravitational field will ascribe different velocities to the same material point. Only for an infinitesimally small region of four-dimensional space, i.e. for one in which the potentials $g_{\mu\nu}$ can be considered a constant—is "velocity" defined at all. Presumably, only within such an infinitesimal region may the covariance of the gravitational equations with respect to linear orthogonal transformations be interpreted in the sense of an | equivalence of [516]
systems of reference moving with respect to each other. But if relativity of motion no longer exists for finite systems of gravitating masses in Einstein's theory, with what right does he assign such great importance to the formal connection to the earlier theory of relativity?

As one can see, the new theory of relativity indeed generalizes the concept of space and time, but thereby restricts the relativity of motion to infinitesimally small spacetime regions. Therefore, its endeavors to achieve relativity with respect to rotational or accelerated motion also do not appear to be promising. Since the velocities of the system of reference enter into the $g_{\mu\nu}$, a spacetime region is only to be considered as infinitesimally small, if the velocity is spatially and temporally constant in it. Therefore it makes no sense to speak of a rotational or accelerated motion of an "infinitesimally small region," and, for instance, to claim that the "equivalence hypothesis is valid in the infinitesimally small."

To what extent is the theorem of the weight of energy valid in the Einstein-Grossmann theory of gravitation? The field equations (57) make it apparent that the quantities $S_{\sigma\nu}$, according to (57a, b) linear functions of the components $T_{\mu\nu}$ of the resultant world tensor, are the ones that generate the gravitational field. The gravitational forces, which a body exerts and experiences, are accordingly defined by the volume integral of these tensor components:

$$\int dV T_{\mu\nu} = m_{\mu\nu}. \tag{59}$$

Therefore, when dealing with a moving and stressed body, one generally has to differentiate between ten "masses" $m_{\mu\nu}$. There also exist ten "gravitational forces," which are derived from the ten potentials $g_{\mu\nu}$. If one transports bodies with differing stress states and different states of motion to the same world point of a variable gravitational field, then the resultant gravitational forces acting on them have in general different directions, and their magnitudes do not depend solely on the energy of the bodies, but also on all the ten $m_{\mu\nu}$.

This intricate mechanism fortunately simplifies since, in relation to the gravitational force derived from g_{44}, the other nine gravitational forces are very small as all the remaining nine masses in relation to the mass m_{44} determined by the energy. Especially for "static complete systems" (II E), all | masses (59) except for m_{44} are [517]
equal to zero, and only the gravitational force acting on it remains effective. *Thus, the total energy alone is the factor determining the gravitational mass of complete static systems*. Similarly as in Nordström's second theory, this result is caused by the com-

pensation of the forces that act on the different parts of the system, and as there it can be applied to closed systems in static equilibrium.

Let us consider for example a hot gas which is enclosed in a cylinder. Forces act on the individual molecules of the gas, whose magnitudes are neither proportional to the energy of the molecules, nor have directions agreeing exactly with the vertical. Nevertheless, the gravitational forces acting on the molecules and on the volume elements of the cylinder sum to a resultant, whose temporal mean value is proportional to the total energy and has a vertical direction. The relation between gravity[2] and energy (respectively, inertia) are accordingly represented in Einstein's tensor theory to the same degree as in Nordström's scalar theory. However, one will have to agree with G. Mie,(28) when he denies that these theories manifest an "physical unity of essence" of gravitational and inertial mass.

In his detailed critique of Einstein's theory, G. Mie(28) attempts to prove that the proportionality of gravitational mass and energy for complete static systems occurs there only as the result of certain specific assumptions, which, incidentally, are said to contradict one another. In his reply,(29) Einstein states the belief—without refuting Mie's objections one by one—that he can trace those contradictions to Mie's demand of covariance only with respect to Lorentz transformations, thereby introducing "preferred systems of reference." That "special assumptions of any kind are not used" in the establishment of his theory, as Einstein asserts there, does not appear credible. In his just-published summary presentation(38) of the "general theory of relativity" Einstein derives the differential equations expressing his theory of gravitation from a variational principle, on the basis of certain specializations whose physical significance is however not elucidated.

The *integration of the field equations* (57) is extraordinarily difficult. Only the method of successive approximations promises success. In this one will usually take
[518] as a first approximation the solution that treats the field | as static. Here, Einstein's theory becomes identical with Abraham's theory; therefore the solution of the problem of the sphere, for example, given in (II C), remains valid here as well.

In his Vienna lecture,(27) A. Einstein takes the normal values of the $g_{\mu\nu}$ as the first approximation:

$$g_{11} = g_{22} = g_{33} = 1, \qquad g_{44} = -c^2, \qquad g_{\mu\nu} = 0 \text{ for } \mu \neq \nu;$$

he considers the deviations $\overset{*}{g}_{\mu\nu}$ from these normal values as quantities of first order, and arrives, by neglecting quantities of second order, at the following differential equations:

$$\Box \overset{*}{g}_{\mu\nu} = \overset{m}{T}_{\mu\nu}. \tag{60}$$

For incoherent motions of masses, the last $(\overset{m}{T}_{44})$ among the components of the material tensor T^m is the most important; it determines the potential $\overset{*}{g}_{44} = \Phi^g$. Then follow the components $\overset{m}{T}_{14}, \overset{m}{T}_{24}, \overset{m}{T}_{34}$, which are of first order in v/c; these determine the potentials $\overset{*}{g}_{14}, \overset{*}{g}_{24}, \overset{*}{g}_{34}$, which can be viewed as the components of a space vector $-(1/c)\mathfrak{A}^g$. The remaining components of T^m are of second order in v/c. If one neglects quantities of this order, then one only needs to consider those four potentials, and obtains for them the differential equations

$$\Box\Phi^g = c^2\mu, \tag{60a}$$

$$\Box\mathfrak{A}^g = c^2\mu\cdot\frac{\mathfrak{v}}{c}, \tag{60b}$$

where μ is the mass density.

Here the analogy to electrodynamics catches one's eye. Except for the sign, the field equations (60a, b) agree with those that must be satisfied in the theory of electrons by the "electromagnetic potentials", the scalar one (Φ) and the vectorial one (A). In this approximation, the Einstein-Grossmann tensor theory of the gravitational field leads to the same results as the vector theory sketched in (I A). Also concerning the expression for the *gravitational force per unit mass*, there exists a far reaching analogy. Einstein obtains:

$$\mathfrak{F}^g = -\frac{1}{2}\mathrm{grad}\Phi^g - \frac{1}{c}\frac{\partial\mathfrak{A}^g}{\partial t} + \frac{1}{c}[\mathrm{v}, \mathrm{curl}\mathfrak{A}^g]. \tag{61}$$

This expression agrees with the one (1) from the vector theory:

$$\mathfrak{F}^g = \mathfrak{E}^g + \frac{1}{c}[\mathfrak{v}\mathfrak{H}^g], \tag{61a}$$

| if one writes: [519]

$$\mathfrak{E}^g = -\frac{1}{2}\mathrm{grad}\Phi^g - \frac{1}{c}\frac{\partial\mathfrak{A}^g}{\partial t}, \tag{61b}$$

$$\mathfrak{H}^g = \mathrm{curl}\mathfrak{A}^g. \tag{61c}$$

Apart from the factor $1/2$ in the gradient of Φ^g, the formulas (61b, c) are identical with the well known formulas which derive the two vectors of the electromagnetic field from the electromagnetic potentials.

The Einstein-Grossmann theory of gravitation is related to the vector theory to the extent that the approximation applied is satisfactory. As in the vector theory, induced gravitational forces are generated on neighboring bodies by the acceleration of a body, forces that act in the direction of the acceleration. Einstein attaches great importance to the existence of these induced gravitational forces in connection with the so called "*hypothesis of the relativity of inertia*."

This hypothesis, already advocated by E. Mach, states that the inertia of a body is only a consequence of the relative acceleration with respect to the totality of the remaining bodies. If this hypothesis applies, then the inertial mass of a body will depend on its position with respect to the remaining bodies. This is now the case in the Einstein's theory because the rest-mass

$$m_0 = \frac{M}{c} \qquad (M \text{ mass constant})$$

is inversely proportional to the speed of light c, whose value in a static gravitational field is reduced upon the approach of external masses. The inertia of a body is thus increased by the accumulation of mass in its vicinity; through this circumstance, Einstein believes himself justified in considering inertia as a consequence of the presence of the remaining masses in the context of that relativity hypothesis.

This consideration looses cogency however, if one calculates by how much c is decreased, and hence m_0 increased, by the masses of celestial objects. The mass of the Sun, at its surface, engenders a relative change in c of the order of 10^{-6}, which then decreases approximately inversely proportional to the distance from the center of the Sun. But in order for the mass of a material point to be considered as being a consequence of the presence of the remaining matter, it mass should vanish upon the removal of that matter. The mass to be removed from the vicinity of that point for that
[520] purpose would now be a millionfold larger than the | mass of the Sun. Accordingly, the hypothesis of the relativity of inertia can only be justified from the point of view of Einstein's theory, if, in addition to the masses of the visible bodies, one assumes also hidden masses. But with this, that hypothesis looses all concrete physical meaning. In the end it does not matter if one anchors the system of reference in "hidden masses," in the "aether" or in a "body α."

Upon the mutual approach of two bodies A and B, their inertial masses grow. But according to the hypothesis of the relativity of inertia, this increase in mass should not occur if A and B are accelerated in unison. Indeed, according to the theorem of the inertia of energy, the inertial resistance of the system of the two bodies should even be smaller than the sum of the inertial masses of the two bodies while separated, because the energy of the system decreases upon the approach of the two bodies. Here, the gravitational forces, induced by the temporal variation of the vector potential $\mathfrak{A}^g$, now come into play, forces that the bodies A and B exert on each other as a result of their acceleration. Through them, the additional inertial forces are overcompensated, so that the inertial mass of the system $A + B$ becomes smaller than the sum of the inertial masses of the bodies A and B.

Einstein sees a significant advantage of this tensor theory over the scalar theories in that the hypothesis of the relativity of inertia fits into his theory, because the scalar theories lack the vector potential $\mathfrak{A}^g$, from which the induced forces are derived. However, we saw that the theory of Einstein and Grossmann also fails to provide a satisfactory quantitative basis for Mach's bold hypothesis, unless one appeals to hidden masses. In view of the enormous complication engendered by the ten-fold multiplication of the gravitational potentials and by the distortion of the four-dimensional world, one is likely, for reasons of the Machian "economy of thought," to prefer the scalar theories, as long as the supposition, that there exist ten gravitational forces instead of one, is not supported by experience.

APPENDIX

Notation: [472]

t time, c the speed of light.

$x, y, z, u = ict$, the coordinates of four-dimensional space.

$\Box \equiv \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} + \frac{\partial^2}{\partial u^2}$.

$\mathfrak{v}$ velocity, v its magnitude.

$\mathfrak{E}$, $\mathfrak{H}$ electric and magnetic vector.

$\mathfrak{F}^e$ force per unit charge.

ρ charge density.

μ mass density, ν rest-mass density.

M mass constant, m_0 rest-mass.

V volume.

$\mathfrak{E}^g$, $\mathfrak{H}^g$ gravitational vectors in the vector theory.

$\mathfrak{F}^g$ gravitational force per unit mass.

$\mathfrak{K}^g$ gravitational force, φ gravitational potential.

$\mathfrak{L}$ Lagrangian function. |

$\mathfrak{G}$ momentum. [473]

E energy, E_0 rest-energy.

E^m, E^e, E^g energy of matter, of the electromagnetic and of the gravitational field.

T^m, T^e, T^g world tensors of matter, of the electromagnetic and of the gravitational field.

D^m, D^e, D^g sum of their diagonal components.

T, D resultant of world tensors and of diagonal sums respectively.

η^m, η^e, η^g energy density of matter, of the electromagnetic and of the gravitational field.

g^m, g^e, g^g the momentum density of matter, of the electromagnetic and of the gravitational field.

$\mathfrak{S}^m$, $\mathfrak{S}^e$, $\mathfrak{S}^g$ the energy flux of matter, of the electromagnetic and of the gravitational field.

$g_{\mu\nu}$ Einstein's gravitational potentials.

[470] REFERENCES

1. V. Volterra, "Sul flusso di energia meccanica." *Nuovo Cimento* (4), X, 337, 1899.
2. H. Poincaré, "Sur la dynamique de l'électron." *Rendiconti del circolo matematico di Palermo,* XXI, 129, 1906 I.
3. H. Minkowski, "Die Grundgleichungen für die elektromagnetischen Vorgänge in bewegten Körpern." *Göttinger Nachrichten*, 1908, p. 53.
4. A. Sommerfeld, "Zur Relativitätstheorie." *Ann. d. Phys.*, 32, 749; 33, 649, 1910. |

[471] 5. A. Einstein, "Über den Einfluß der Schwerkraft auf die Ausbreitung des Lichtes." *Ann. d. Phys.*, 35, 898, 1911.
6. M. Abraham, "Sulla teoria della gravitazione." *Rendiconti della R. Acc. dei Lincei*, XX^2, 678, 1911; XXI^1, 27, 1912. German: *Physik. Zeitschr.,* 13, 1, 1912.
7. M. Abraham, "Sulla legge elementare della gravitazione." *Rendic. d. R. Acc. dei Lincei*, XXI^1, 94, 1912. German: *Physik. Zeitschr.*, 13, 4, 1912.
8. M. Abraham, "Sulla conservazione dell' energia e della materia nel campo gravitazionale." *Rendic. d. R. Acc. d. Lincei*, XXI^1, 432, 1912. German: *Physik. Zeitschr.*, 13, 311, 1912.
9. M. Abraham, "Sulla caduta libera." *Rondic d. R. Instituto Lombardo* (2), XLV, 290, 1912. German: *Physik. Zeitschr.*, 13, 310, 1912.
10. M. Abraham, "Sulle onde luminose e gravitazionali." *Nuovo Cimento* (6), III, 211, 1912.
11. A. Einstein, "Lichtgeschwindigkeit und Statik des Gravitationsfeldes." *Ann. d. Phys.*, 38, 355, 1912.
12. A. Einstein, "Zur Theorie des statischen Gravitationsfeldes." *Ann. d. Phys.*, 38, 443, 1912.
13. M. Abraham, "Relativität und Gravitation." *Ann. d. Phys.*, 38, 1056, 1912.
14. A. Einstein, "Relativität und Gravitation." *Ann d. Phys.*, 38, 1059, 1912.
15. M. Abraham, "Nochmals Relativität und Gravitation." *Ann. d. Phys.*, 39, 444, 1912.
16. M. Abraham, "Das Gravitationsfeld." *Physik. Zeitsch.*, 13, 793, 1912.
17. M. Abraham, "Una nuova teoria della gravitazione." *Nuovo Cimento* (6), IV, 459, 1912. German: *Archiv der Mathematik u. Physik* (3), XX, 193, 1912.
18. G. Pavanini, "Prime conseguence d'una recente teoria della gravitazione." *Rendic. d. R. Acc. dei Lincei*, XXI^2, 648, 1912; XXII^1, 369, 1913.
19. I. Ishiwara, "Zur Theorie der Gravitation." *Physik. Zeitschr.*, 13, 1189, 1912.
20. G. Nordström, "Relativitätsprinzip und Gravitation." *Physik. Zeitschr.*, 13, 1126, 1912.

21. G. Mie, “Grundlagen einer Theorie der Materie.” Chapter V. *Ann. d. Phys.*, 40, 25, 1913.

22. G. Nordström, “Träge und schwere Masse in der Relativitätsmechanik.” *Ann. d. Phys.*, 40, 856, 1913.

23. B. Caldonazzo, “Traiettorie dei raggi luminosi e dei punti materiali nel campo gravitazionale.” *Nuovo Cimento* (6), V, 267, 1913.

24. A. Einstein and M. Grossmann, *Entwurf einer verallgemeinerten Relativitätstheorie und einer Theorie der Gravitation*. Leipzig, B. G. Teubner 1913.

25. M. Behacker, “Der freie Fall und die Planetenbewegung in Nordströms Gravitationstheorie.” *Physik. Zeitschr.*, 14, 989, 1913.

26. G. Nordström, “Zur Theorie der Gravitation vom Standpunkt des Relativitätsprinzips.” *Ann. d. Phys.*, 42, 533, 1913.

27. A. Einstein, “Zum gegenwärtigen Stande des Gravitationsproblems.” *Physik. Zeitschr.*, 14, 1249, 1913. |

28. G. Mie, “Bemerkungen zu der Einsteinschen Gravitationstheorie.” *Physik. Zeitschr.*, 15, 115 and 169, 1914. [472]

29. A. Einstein, “Prinzipielles zur verallgemeinerten Relativitätstheorie und Gravitationstheorie.” *Physik. Zeitschr.*, 15, 176, 1914.

30. I. Ishiwara, “Grundlagen einer relativistischen elektromagnetischen Gravitationstheorie.” *Physik. Zeitschr.*, 15, 294, 1914.

31. G. Nordström, “Die Fallgesetze und Planetenbewegungen in der Relativitätstheorie.” *Ann. d. Phys.*, 43, 1101, 1914.

32. G. Nordström, “Über den Energiesatz in der Gravitationstheorie.” *Physik. Zeitschr.*, 15, 375, 1914.

33. G. Nordström, “Über die Möglichkeit, das elektromagnetische Feld und das Gravitationsfeld zu vereinigen.” *Physik. Zeitschr.*, 15, 504, 1914.

34. A. Einstein and A. D. Fokker, “Die Nordströmsche Gravitationstheorie vom Standpunkt des absoluten Differentialkalküls.” *Ann. d. Phys.*, 44, 321, 1914.

35. A. Einstein and M. Grossmann, “Kovarianzeigenschaften der Feldgleichungen der auf die verallgemeinerte Relativitätstheorie gegründeten Gravitationstheorie.” *Zeitschr. f. Mathem. u. Phys.*, 63, 215, 1914.

36. E. Kretschmann, *Eine Theorie der Schwerkraft im Rahmen der ursprünglichen Einsteinschen Relativitätstheorie*. Doctoral Thesis. Berlin 1914.

37. G. Nordström, “Zur Elektrizitäts- und Gravitationstheorie.” *Finska V. S. Förhandlingar*, LVII. A. No. 4. 1914.

38. A. Einstein, “Die formale Grundlage der allgemeinen Relativitätstheorie.” *Sitzungsber. d. Berl. Ak.*, Nov. 1914.

EDITORIAL NOTES

[1] Here "*Schwere*" has been translated as "gravity," but elsewhere it has been translated as "weight," since in these other occurrences Abraham apparently uses "*Schwere*" and "*Gewicht*" interchangeably.

[2] Raised numbers in parentheses refer to the numbers of the above chronological list of references.

[3] Note that the wording Abraham uses here differs slightly from that of the earlier work to which he refers.

[4] The superscript ^{g} in T^{g} in eq. (36b) is missing in the original.

A FIELD THEORY OF GRAVITATION
IN THE FRAMEWORK
OF SPECIAL RELATIVITY

JOHN D. NORTON

EINSTEIN, NORDSTRÖM, AND THE EARLY DEMISE OF SCALAR, LORENTZ COVARIANT THEORIES OF GRAVITATION

1. INTRODUCTION

The advent of the special theory of relativity in 1905 brought many problems for the physics community. One, it seemed, would not be a great source of trouble. It was the problem of reconciling Newtonian gravitation theory with the new theory of space and time. Indeed it seemed that Newtonian theory could be rendered compatible with special relativity by any number of small modifications, each of which would be unlikely to lead to any significant deviations from the empirically testable consequences of Newtonian theory.[1] Einstein's response to this problem is now legend. He decided almost immediately to abandon the search for a Lorentz covariant gravitation theory, for he had failed to construct such a theory that was compatible with the equality of inertial and gravitational mass. Positing what he later called the principle of equivalence, he decided that gravitation theory held the key to repairing what he perceived as the defect of the special theory of relativity—its relativity principle failed to apply to accelerated motion. He advanced a novel gravitation theory in which the gravitational potential was the now variable speed of light and in which special relativity held only as a limiting case.

It is almost impossible for modern readers to view this story with their vision unclouded by the knowledge that Einstein's fantastic 1907 speculations would lead to his greatest scientific success, the general theory of relativity. Yet, as we shall see, in

1 In the historical period under consideration, there was no single label for a gravitation theory compatible with special relativity. The Einstein of 1907 would have talked of the compatibility of gravitation and the *principle* of relativity, since he then tended to use the term "principle of relativity" where we would now use "theory of relativity". See (CPAE 2, 254). Minkowski (1908, 90) however, talked of reform "in accordance with the world postulate." Nordström (1912, 1126), like Einstein, spoke of "adapting ... the theory of gravitation to the principle of relativity" or (Nordström 1913, 872) of "treating gravitational phenomena from the standpoint of the theory of relativity," emphasizing in both cases that he planned to do so retaining the constancy of the speed of light in order to distinguish his work from Einstein's and Abraham's. For clarity I shall describe gravitation theories compatible with special relativity by the old-fashioned but still anachronistic label "Lorentz covariant." It describes exactly the goal of research, a gravitation theory whose equations are covariant under Lorentz transformation. For a simplified presentation of the material in this chapter, see also (Norton 1993).

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 3
Gravitation in the Twilight of Classical Physics: Between Mechanics, Field Theory, and Astronomy.

1907 Einstein had only the slenderest of grounds for judging all Lorentz covariant gravitation theories unacceptable. His 1907 judgement was clearly overly hasty. It was found quite soon that one could construct Lorentz covariant gravitation theories satisfying the equality of inertial and gravitational mass without great difficulty. Nonetheless we now do believe that Einstein was right in so far as a thorough pursuit of Lorentz covariant gravitation theories does lead us inexorably to abandon special relativity. In the picturesque wording of Misner et al. (1973, Ch.7) "gravity bursts out of special relativity."

These facts raise some interesting questions. As Einstein sped towards his general theory of relativity in the period 1907–1915 did he reassess his original, hasty 1907 judgement of the inadequacy of Lorentz covariant gravitation theories? In particular, what of the most naturally suggested Lorentz covariant gravitation theory, one in which the gravitational field was represented by a scalar field and the differential operators of the Newtonian theory were replaced by their Lorentz covariant counterparts? Where does this theory lead? Did the Einstein of the early 1910s have good reason to expect that developing this theory would lead outside special relativity?

This paper provides the answers to these questions. They arise in circumstances surrounding a gravitation theory, developed in 1912–1914, by the Finnish physicist Gunnar Nordström. It was one of a number of more conservative gravitation theories advanced during this period. Nordström advanced this most conservative scalar, Lorentz covariant gravitation theory and developed it so that it incorporated the equality of inertial and gravitation mass. It turned out that even in this most conservative approach, odd things happened to space and time. In particular, the lengths of rods and the rates of clocks turn out to be affected by the gravitational field, so that the spaces and times of the theory's background Minkowski spacetime ceased to be directly measurable. The *dénouement* of the story came in early 1914. It was shown that this conservative path led to the same sort of gravitation theory as did Einstein's more extravagant speculations on generalizing the principle of relativity. It lead to a theory, akin to general relativity, in which gravitation was incorporated into a dynamical spacetime background. If one abandoned the inaccessible background of Minkowski spacetime and simply assumed that the spacetime of the theory was the one revealed by idealized rod and clock measurements, then it turned out that the gravitation theory was actually the theory of a spacetime that was only conformally flat—gravitation had burst out of special relativity. Most strikingly the theory's gravitational field equation was an equation strongly reminiscent to modern readers of the field equations of general relativity:

$$R = kT$$

where R is the Riemann curvature scalar and T the trace of the stress-energy tensor. This equation was revealed before Einstein had advanced the generally covariant field equations of general relativity, at a time in which he believed that no such field equations could be physically acceptable.

What makes the story especially interesting are the two leading players other than Nordström. The first was Einstein himself. He was in continued contact with Nordström during the period in which the Nordström theory was developed. We shall see that the theory actually evolved through a continued exchange between them, with Einstein often supplying ideas decisive to the development of the theory. Thus the theory might more accurately be called the "Einstein-Nordström theory." Again it was Einstein in collaboration with Adriaan Fokker who revealed in early 1914 the connection between the theory and conformally flat spacetimes.

The second leading player other than Nordström was not a person but a branch of special relativity, the relativistic mechanics of stressed bodies. This study was under intensive development at this time and had proven to be a locus of remarkably non-classical results. For example it turned out that a moving body would acquire additional energy, inertia and momentum simply by being subjected to stresses, even if the stresses did not elastically deform the body. The latest results of these studies—most notably those of Laue—provided Einstein and Nordström with the means of incorporating the equality of inertial and gravitational mass into their theory. It was also the analysis of stressed bodies within the theory that led directly to the conclusion that even idealized rods and clocks could not measure the background Minkowski spacetime directly but must be affected by the gravitational field. For Einstein and Nordström concluded that a body would also acquire a gravitational mass if subjected to non-deforming stresses and that one had to assume that such a body would alter its size in moving through the gravitational field on pain of violating the law of conservation of energy.

Finally we shall see that the requirement of equality of inertial and gravitational mass is a persistent theme of Einstein's and Nordström's work. However the requirement proves somewhat elastic with both Einstein and Nordström drifting between conflicting versions of it. It will be convenient to prepare the reader by collecting and stating the relevant versions here. On the observational level, the equality could be taken as requiring:

- Uniqueness of free fall:[2] The trajectories of free fall of all bodies are independent of their internal constitution.

Einstein preferred a more restrictive version:

- Independence of vertical acceleration: The vertical acceleration of bodies in free fall is independent of their constitutions and horizontal velocities.

In attempting to devise theories compatible with these observational requirements, Einstein and Nordström considered requiring equality of gravitational mass with

- inertial rest mass
- the inertial mass of closed systems
- the inertial mass of complete static systems
- the inertial mass of a a complete stationary systems[3]

2 This name is drawn from (Misner et al. 1973, 1050).

3 The notions of complete static and complete stationary systems arise in the context of the mechanics of stressed bodies and are discussed in Sections 9 and 12 below.

More often than not these theoretical requirements failed to bring about the desired observational consequences. Unfortunately it is often unclear precisely which requirement is intended when the equality of inertial and gravitational mass was invoked.

2. THE PROBLEM OF GRAVITATION IMMEDIATELY AFTER 1905

In the years immediately following 1905 it was hard to see that there would be any special problem in modifying Newtonian gravitation theory in order to bring it into accord with the special theory of relativity. The problem was not whether it could be done, but how to choose the best of the many possibilities perceived, given the expectation that relativistic corrections to Newtonian theory might not have measurable consequences even in the very sensitive domain of planetary astronomy. Poincaré (1905, 1507–1508;1906, 166–75), for example, had addressed the problem in his celebrated papers on the dynamics of the electron. He limited himself to seeking an expression for the gravitational force of attraction between two masses that would be Lorentz covariant[4] and would yield the Newtonian limit for bodies at rest. Since this failed to specify a unique result he applied further constraints including the requirement[5] of minimal deviations from Newtonian theory for bodies with small velocities, in order to preserve the Newtonian successes in astronomy. The resulting law, Poincaré noted, was not unique and he indicated how variants consistent with its constraints could be derived by modifying the terms of the original law.

Minkowski (1908, 401–404; 1909, 443–4) also sought a relativistic generalization of the Newtonian expression for the gravitational force acting between two bodies. His analysis was simpler than Poincaré's since merely stating his law in terms of the geometric structures of his four dimensional spacetime was sufficient to guarantee automatic compatibility with special relativity. Where Poincaré (1905, 1508; 1906, 175) had merely noted his expectation that the deviations from Newtonian astronomical prediction introduced by relativistic corrections would be small, Minkowski (1908, 404) computed the deviations due to his law for planetary motions and concluded that they were so small that they allowed no decision to be made concerning the law.

Presumably neither Poincaré nor Minkowski were seeking a fundamental theory of gravitation, for they both considered action-at-a-distance laws at a time when field theories were dominant. Rather the point was to make *plausible*[6] the idea that some slight modification of Newtonian gravitational law was all that was necessary to bring it into accord with special relativity, even if precise determination of that modification was beyond the reach of the current state of observational astronomy.

---

4 More precisely he required that the law governing propagation of gravitational action be Lorentz covariant and that the gravitational forces transform in the same way as electromagnetic forces.

5 Also he required that gravitational action propagate forward in time from a given body.

6 The word is Minkowski's. He introduced his treatment of gravitation (Minkowski 1908, 401) with the remark "I would not like to fail to make it plausible that nothing in the phenomena of gravitation can be expected to contradict the assumption of the postulate of relativity."

3. EINSTEIN'S 1907 REJECTION OF LORENTZ COVARIANT GRAVITATION THEORIES

In 1907 Einstein's attention was focussed on the problem of gravitation and relativity theory when he agreed to write a review article on relativity theory for Johannes Stark's *Jahrbuch der Radioaktivität und Elektronik*. The relevant parts of the review article (Einstein 1907a, 414; Section V, 454–62) say nothing of the possibility of a Lorentz covariant gravitation theory. Rather Einstein speculates immediately on the possibility of extending the principle of relativity to accelerated motion. He suggests the relevance of gravitation to this possibility and posits what is later called the principle of equivalence as the first step towards the complete extension of the principle of relativity.

It is only through later reminiscences that we know something of the circumstances leading to these conclusions. The most informative are given over 25 years later in 1933 when Einstein gave a sketch of his pathway to general relativity.[7] In it he wrote (Einstein 1933, 286–87):

> I came a step nearer to the solution of the problem [of extending the principle of relativity] when I attempted to deal with law of gravity within the framework of the special theory of relativity. Like most writers at the time, I tried to frame a *field-law* for gravitation, since it was no longer possible, at least in any natural way, to introduce direct action at a distance owing to the abolition of the notion of absolute simultaneity.
>
> The simplest thing was, of course, to retain the Laplacian scalar potential of gravity, and to complete the equation of Poisson in an obvious way by a term differentiated with respect to time in such a way that the special theory of relativity was satisfied. The law of motion of the mass point in a gravitational field had also to be adapted to the special theory of relativity. The path was not so unmistakably marked out here, since the inert mass of a body might depend on the gravitational potential. In fact this was to be expected on account of the principle of the inertia of energy.

While Einstein's verbal description is brief, the type of gravitation theory he alludes to is not too hard to reconstruct. In Newtonian gravitation theory, with a scalar potential ϕ, mass density ρ and G the gravitation constant, the gravitational field equation—the "equation of Poisson"— is[8]

$$\nabla^2\phi = \left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)\phi = \frac{\partial^2\phi}{\partial x_i \partial x_i} = 4\pi G\rho. \tag{1}$$

The force **f**, with components (f_1, f_2, f_3), on a point mass m is given by $-m\nabla\phi$ so that

$$f_i = m\frac{dv_i}{dt} = -m\left(\frac{\partial}{\partial x_i}\right)\phi \tag{2}$$

7 A similar account is given more briefly in (Einstein 1949, 58–63).

8 $(x, y, z) = (x_1, x_2, x_3)$ are the usual spatial Cartesian coordinates. The index i ranges over 1, 2, 3. Here and henceforth, summation over repeated indices is implied.

is the law of motion of a point mass m with velocity $v_i = dx_i/dt$ in the gravitational field ϕ.

The adaptation of (1) to special relativity is most straightforward. The added term, differentiated with respect to the time coordinate, converts the Laplacian operator ∇^2 into a Lorentz covariant d' Alembertian $\Box^2$. so that the field equation alluded to by Einstein would be

$$\Box^2\phi = \left(\frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \nabla^2\right)\phi = -4\pi G\nu. \tag{3}$$

For consistency ϕ is assumed to be Lorentz invariant and the mass density ρ must be replaced with a Lorentz invariant, such as the rest mass density ν used here.

The modification to the law of motion of a point mass is less clear. The natural Lorentz covariant extension of (2) is most obvious if we adopt the four dimensional spacetime methods introduced by Minkowski (1908). Einstein could not have been using these methods in 1907. However I shall write the natural extension here since Einstein gives us little other guide to the form of the equation he considered, since the properties of this equation fit exactly with Einstein's further remarks and since this equation will lead us directly to Nordström's work. The extension of (2) is

$$F_\mu = m\frac{dU_\mu}{d\tau} = -m\frac{\partial\phi}{\partial x_\mu} \tag{4}$$

where F_μ is the four force on a point mass with rest mass m, $U_\mu = dx_\mu/d\tau$ is its four velocity, τ is the proper time and $\mu = 1, 2, 3, 4$.[9] Following the practice of Nordström's papers, the coordinates are $(x_1, x_2, x_3, x_4) = (x, y, z, u = ict)$, for c the speed of light.

Simple as this extension is, it turns out to be incompatible with the kinematics of a Minkowski spacetime. In a Minkowski spacetime, the constancy of c entails that the four velocity U_μ along a world line is orthogonal to the four acceleration $dU_\mu/d\tau$. For we have $c^2 d\tau^2 = -dx_\mu dx_\mu$, so that $c^2 = -U_\mu U_\mu$ and the orthogonality now follows from the constancy of c

$$-\frac{1}{2}\frac{dc^2}{d\tau} = U_\mu\frac{dU_\mu}{d\tau} = 0 \tag{5}$$

(4) and (5) together entail

$$F_\mu U_\mu = -m\frac{\partial\phi}{\partial x_\mu}\frac{dx_\mu}{d\tau} = -m\frac{d\phi}{d\tau} = 0$$

so that the law (4) can only obtain in a Minkowski spacetime in the extremely narrow case in which the field ϕ is constant along the world line of the particle, i.e.

---

9 Throughout this paper, Latin indices $i, k, \ldots$ range over 1, 2, 3 and Greek indices $\mu, \nu, \ldots$ range over 1, 2, 3, 4.

$$\frac{d\phi}{d\tau} = 0. \tag{6}$$

We shall see below that one escape from this problem published by Nordström involves allowing the rest mass m to be a function of the potential ϕ. Perhaps this is what Einstein referred to above when he noted of the law of motion that the "path was not so unmistakably marked out here, since the inert mass of a body might depend on the gravitational potential."

Whatever the precise form of the modifications Einstein made, he was clearly unhappy with the outcome. Continuing his recollections, he noted:

> These investigations, however, led to a result which raised my strong suspicions. According to classical mechanics, the vertical acceleration of a body in the vertical gravitational field is independent of the horizontal component of its velocity. Hence in such a gravitational field the vertical acceleration of a mechanical system or of its center of gravity works out independently of its internal kinetic energy. But in the theory I advanced, the acceleration of a falling body was not independent of its horizontal velocity or the internal energy of the system.

The result Einstein mentions here is readily recoverable from the law of motion (4) in a special case in which it is compatible with the identity (5). The result has more general applicability, however. The modifications introduced by Nordström to render (4) compatible with (5) vanish in this special case, as would, presumably, other natural modifications that Einstein may have entertained. So this special case is also a special case of these more generally applicable laws.

We consider a coordinate system in which:

(i) the field is time independent $(\partial\phi/\partial t = 0)$ at some event and

(ii) the motion of a point mass m in free fall at that event is such that the "vertical" direction of the field, as given by the acceleration three vector dv_i/dt, is perpendicular to the three velocity v_i, so that

$$v_i \cdot \frac{dv_i}{dt} = 0 \tag{7}$$

and the point's motion is momentarily "horizontal."

Condition (7) greatly simplifies the analysis, since it entails that the t derivative of any function of $v^2 = v_i v_i$ vanishes, so that we have

$$\frac{d}{dt}\frac{dt}{d\tau} = \frac{d}{dt}\left(\frac{1}{\sqrt{1-\frac{v^2}{c^2}}}\right) = 0. \tag{8}$$

Notice also that in this case (7) entails that $v_i\frac{\partial\phi}{\partial x_i} = 0$ so that

$$\frac{d\phi}{d\tau} = \frac{dt}{d\tau}\left(\frac{\partial\phi}{\partial t} + v_i\frac{\partial\phi}{\partial x_i}\right) = 0$$

and (6) and then also (5) are satisfied for this special case. Finally an expression for the acceleration of the point mass now follows directly from (4) and is[10]

$$\frac{dv_i}{dt} = -\left(1 - \frac{v^2}{c^2}\right)\frac{\partial\phi}{\partial x_i}. \tag{9}$$

According to (9), the greater the horizontal velocity v, the less the vertical acceleration, so that this acceleration is dependent on the horizontal velocity as Einstein claimed.

Einstein also claims in his remarks that the vertical acceleration would not be independent of the internal energy of the falling system. This result is suggested by equation (9), which tells us that the vertical acceleration of a point mass diminishes with its kinetic energy if the velocity generating that kinetic energy is horizontally directed. If we apply this result to the particles of a kinetic gas, we infer that in general each individual particle will fall slower the greater its velocity. Presumably this result applies to the whole system of a kinetic gas so that the gas falls slower the greater the kinetic energy of its particles, that is, the greater its internal energy. This example of a kinetic gas was precisely the one given by Einstein in an informal lecture on April 14, 1954 in Princeton according to lecture notes taken by J. A. Wheeler.[11]

Einstein continued his recollections by explaining that he felt these results so contradicted experience that he abandoned the search for a Lorentz covariant gravitation theory.

> This did not fit with the old experimental fact that all bodies have the same acceleration in a gravitational field. This law, which may also be formulated as the law of the equality of inertial and gravitational mass, was now brought home to me in all its significance. I was in the highest degree amazed at its existence and guessed that in it must lie the key to a deeper understanding of inertia and gravitation. I had no serious doubts about its strict validity even without knowing the results of the admirable experiments of Eötvös, which—if my memory is right—I only came to know later. I now abandoned as inadequate the attempt to treat the problem of gravitation, in the manner outlined above, within the framework of the special theory of relativity. It clearly failed to do justice to the most fundamental property of gravitation.

Einstein then recounted briefly the introduction of the principle of equivalence, upon which would be based his continued work on gravitation and relativity, and concluded

10 Since $\frac{dU_i}{d\tau} = \frac{dt}{d\tau}\frac{d}{dt}\left(v_i\frac{dt}{d\tau}\right)$, (9) follows directly from (4) using (8).

11 Wheeler's notes read "I had to write a paper about the content of special relativity. Then I came to the question how to handle gravity. The object falls with a different acceleration if it is moving than if it is not moving. ... Thus a gas falls with another acceleration if heated than if not heated. I felt this is not true ..." (Wheeler 1979, 188).

> Such reflections kept me busy from 1908 to 1911, and I attempted to draw special conclusions from them, of which I do not propose to speak here. For the moment the one important thing was the discovery that a reasonable theory of gravitation could only be hoped for from an extension of the principle of relativity.

Our sources concerning Einstein's 1907 renunciation of Lorentz covariant gravitation theories are largely later recollections so we should be somewhat wary of them. Nonetheless they all agree in the essential details:[12] Einstein began his attempts to discover a Lorentz covariant theory of gravitation as a part of his work on his 1907 *Jahrbuch* review article. He found an inconsistency between these attempts and the exact equality of inertial and gravitational mass, which he found sufficiently disturbing to lead him to abandon the search for such theories.

We shall see shortly that Einstein's 1907 evaluation and dismissal of the prospects of a Lorentz covariant gravitation theory—as reconstructed above—was far too hasty. Within a few years Einstein himself would play a role in showing that one could construct a Lorentz covariant gravitation theory that was fully compatible with the exact equality of inertial and gravitational mass. We can understand why Einstein's 1907 analysis would be hurried, however, once we realize that the he could have devoted very little time to contemplation of the prospects of a Lorentz covariant gravitation theory. He accepted the commission of the *Jahrbuch*'s editor, Stark, to write the review in a letter of September 25, 1907 (EA 22 333) and the lengthy and completed article was submitted to the journal on December 4, 1907, a little over two months later. This period must have been a very busy one for Einstein. As he explained to Stark in the September 25 letter, he was not well read in the current literature pertinent to relativity theory, since the library was closed during his free time. He asked Stark to send him relevant publications that he might not have seen.[13] During this period, whatever time Einstein could have spent privately contemplating the prospects of a Lorentz covariant gravitation theory would have been multiply diluted. There were the attractions of the principle of equivalence, whose advent so dazzled him that he called it the "happiest thought of [his] life".[14] Its exploitation attracted all the pages of the review article which concern gravitation and in which the prospects of a Lorentz covariant gravitation theory are not even mentioned. Further diluting his time would be the demands of the remaining sections of the review article. The section devoted to gravitation filled only nine of the article's fifty two pages. Finally, of

---

12 See also the 1920 recollections of Einstein on p. 23, "Grundgedanken und Methoden der Relativitätstheorie in ihrer Entwicklung dargestellt," unpublished manuscript, control number 2 070, Duplicate Einstein Archive, Mudd Manuscript Library, Princeton, NJ. (Henceforth "EA 2 070".) Einstein recalls:
"When, in the year 1907, I was working on a summary essay concerning the special theory of relativity for the *Jahrbuch für Radioaktivität und Elektronik* [sic], I had to try to modify Newton's theory of gravitation in such a way that it would fit into the theory [of relativity]. Attempts in this direction showed the possibility of carrying out this enterprise, but they did not satisfy me because they had to be supported by hypotheses without physical basis." Translation from (Holton 1975, 369–71).

13 Einstein thanked him for sending papers in a letter of October 4, 1907 (EA, 22 320).

14 In Einstein's 1920 manuscript (EA 2 070, 23–25).

course, there were the obligations of his job at the patent office. It is no wonder that he lamented to Stark in a letter of November 1, 1907, that he worked on the article in his "unfortunately truly meagerly measured free time" (EA, 22 335).

4. EINSTEIN'S ARGUMENT OF JULY 1912

If the Einstein of 1907 had not probed deeply the prospects of Lorentz covariant gravitation theories, we might well wonder if he returned to give the problem more thorough treatment in the years following. We have good reason to believe that as late as July 1912, Einstein had made no significant advance on his deliberations of 1907.[15] Our source is an acrimonious dispute raging at this time between Einstein and Max Abraham. In language that rarely appeared in the unpolluted pages of *Annalen der Physik*, Abraham (1912c, 1056) accused Einstein's theory of relativity of having "exerted an hypnotic influence especially on the youngest mathematical physicists which threatened to hamper the healthy development of theoretical physics." He rejoiced especially in what he saw as major retractions in Einstein's latest papers on relativity and gravitation. Einstein (1911) involved a theory of gravitation which gave up the constancy of the velocity of light and Einstein (1912a, 1912b) even dispensed with the requirement of the invariance of the equations of motion under Lorentz transformation. These concessions, concluded Abraham triumphantly, were the "death blow" for relativity theory.

Einstein took this attack very seriously. His correspondence from this time, a simple gauge of the focus of his thoughts, was filled with remarks on Abraham. He repeatedly condemned Abraham's (1912a,1912b) new theory of gravitation, which had adopted Einstein's idea of a variable speed of light as the gravitational potential. "A stately beast that lacks three legs," he wrote scathingly of the theory to Ludwig Hopf.[16] He anticipated the dispute with Abraham with some relish, writing to Hopf earlier of the coming "difficult ink duel."[17] The public dispute ended fairly quickly, however, with Einstein publishing a measured and detailed reply (Einstein 1912d) and then refusing to reply to Abraham's rejoinder (Abraham 1912d). Instead Einstein published a short note (Einstein 1912e) indicating that both parties had stated their views and asking readers not to interpret Einstein's silence as agreement. Nonetheless Einstein continued to hold a high opinion of Abraham as a physicist, lamenting in a letter to Hopf that Abraham's theory was "truly superficial, contrary to his [Abraham's] usual practice."[18]

15 This is a little surprising. Einstein had neglected gravitation in 1908–1911, possibly because of his preoccupation with the problem of quanta. (See Pais (1983, 187–90.)) However he had returned to gravitation with vigor with his June 1911 submission of Einstein (1911) and by July 1912, the time of his dispute with Abraham, he had completed at least two more novel papers on the subject, (Einstein 1912a, 1912b), and possibly a third, (Einstein 1912c).

16 Einstein to Ludwig Hopf, 16 August 1912, (EA 13 288).

17 Einstein to Ludwig Hopf, December 1911 (?), (EA 13 282).

Under these circumstances, Einstein had every incentive to make the best case for his new work on gravitation. In particular, we would expect Einstein to advance the best arguments available to him to justify his 1907 judgement of the untenability of Lorentz covariant gravitation theories, for it was this conclusion that necessitated the consideration of gravitation theories that went beyond special relativity. What he included in his response shows us that as late as July 4, 1912—the date of submission of his response (Einstein 1912d)—his grounds for this judgement had advanced very little beyond those he recalled having in 1907. He wrote (pp. 1062–63)

> One of the most important results of the theory of relativity is the realization that every energy E possesses an inertia (E/c^2) proportional to it. Since each inertial mass is at the same time a gravitational mass, as far as our experience goes, we cannot help but ascribe to each energy E a gravitational mass E/c^2.[19] From this it follows immediately that gravitation acts more strongly on a moving body than on the same body in case it is at rest.
>
> If the gravitational field is to be interpreted in the sense of our current theory of relativity, this can happen only in two ways. One can conceive of the gravitation vector either as a four-vector or a six-vector. For each of these two cases there are transformation formulae for the transition to a uniformly moving reference system. By means of these transformation formulae and the transformation formulae for ponderomotive forces one can find for both cases the forces acting on moving material points in a static gravitational field. However from this one arrives at results which conflict with the consequences mentioned of the law of the gravitational mass of energy. Therefore it seems that the gravitation vector cannot be incorporated without contradiction in the scheme of the current theory of relativity.

Einstein's argument is a fairly minor embellishment of the reflections summarized in Section 3 above. Einstein has replaced a single theory, embodied in equations such as (3) and (4), with two general classes of gravitation theory, the four-vector and six-vector theory. In both classes of gravitation theory, in the case of moving masses, Einstein claims that the gravitational field fails to act on them in proportion to their total energy, in effect violating the requirement of equality of inertial and gravitational mass.

---

18 Einstein to Ludwig Hopf, 12 June 1912, (EA 13 286). Einstein retained his high opinion of Abraham as a physicist. Late the following year, after his work had advanced into the first sketch of the general theory of relativity, Einstein conceded to his confidant Besso that "Abraham has the most understanding [of the new theory]." Einstein to Michele Besso, end of 1913, in (Speziali 1972, 50). For further mention of Abraham in correspondence from this period see Einstein to Heinrich Zangger, 27 January 1912, (EA 39 644); Einstein to Wilhelm Wien, 27 January 1912, (EA 23 548); Einstein to Heinrich Zangger, 29 February 1912, (EA 39 653); Einstein to Wilhelm Wien, 24 February 1912, (EA 23 550); Einstein to Heinrich Zangger, 20 May 1912, (EA 39 655); Einstein to Michele Besso, 26 March 1912, (EA 7 066); Einstein to Heinrich Zangger, summer 1912, (EA 39 657); Einstein to Arnold Sommerfeld, 29 October 1912, (EA 21 380). See also (Pais 1982, 231–32).

19 At this point, Einstein inserts the footnote:
Hr. Langevin has orally called my attention to the fact that one comes to a contradiction with experience if one does not make this assumption. That is, in radioactive decay large quantities of energy are given off, so that the *inertial* mass of the matter must diminish. If the gravitational mass were not to diminish proportionally, then the gravitational acceleration of bodies made out of different elements would have to be demonstrably different in the same gravitational field.

Einstein does not give a full derivation of the result claimed. However we can reconstruct what he intended from the derivation sketch given. The two types of force fields correspond to the "spacetime vectors type I and II" introduced by Minkowski (1908, §5), which soon came to be known as four- and six-vector fields, respectively (Sommerfeld 1910, 750). They represented the two types of force fields then examined routinely in physics. The four-vector corresponds to the modern vector of a four-dimensional manifold. The gravitational four-force F_μ acting on a body with rest mass m in a four-vector theory is

$$F_\mu = mG_\mu. \tag{10a}$$

An example of such a theory is given by (4) above in which the gravitation four-vector G_μ is set equal to $-\partial\phi/\partial x_\mu$. The six-vector corresponds to our modern antisymmetric second rank tensor which has six independent components. The classic example of a six-vector is what Sommerfeld called "the six-vector ... of the electromagnetic field" (Sommerfeld 1910, 754). We would now identify it as the Maxwell field tensor. Presumably Einstein intended a six-vector gravitation theory to be modelled after electrodynamics, so that the gravitational four-force F_μ acting on a body with rest mass m and four-velocity U_ν in such a theory would be given by

$$F_\mu = mG_{\mu\nu}U_\nu. \tag{11a}$$

The gravitation six-vector, $G_{\mu\nu}$, satisfies the antisymmetry condition $G_{\mu\nu} = -G_{\nu\mu}$. This antisymmetry guarantees compatibility with the identity (5) since it forces $F_\mu U_\mu = 0$.

Einstein claims that one needs only the transformation formulae for four and six-vectors and for ponderomotive forces. to arrive at the results. However, since both (10a) and (11a) are Lorentz covariant, application of the transformation formulae to these equations simply returns equations of identical form—an uninformative outcome. We do recover results of the type Einstein claims, however, if we apply these transformation formulae to non-covariant specializations of (10a) and (11a).

We consider arbitrary four and six-vector gravitational fields G_μ and $G_{\mu\nu}$. In each there is a body of mass m in free fall. In each case, select and orient a coordinate system $S'(x', y', z', u' = ict')$ in such a way that each mass is instantaneously at rest and is accelerating only in $y' = x'_2$ direction. For these coordinate systems, the three spatial components of the four-force, F'_i , are equal to the three components of the three force, f'_i , acting on the masses. In particular the $x'_2 = y'$ component f'_2 of the three force is given in each case by

$$f'_2 = mG'_2 , \tag{10b}$$

$$f'_2 = mG'_{2\nu}U'_\nu = mG'_{24}\, ic \tag{11b}$$

since $U'_\mu = (0, 0, 0, ic)$. If we now transform from S' to a reference system $S(x, y, z, u = ict)$ moving at velocity v in the $x'_1 = x'$ direction S', then the relevant Lorentz transformation formulae are

$$f'_2 = \frac{f_2}{\sqrt{1-\frac{v^2}{c^2}}}, \qquad G'_{24} = \frac{G_{24}-(iv/\mathrm{c})G_{21}}{\sqrt{1-\frac{v^2}{c^2}}}.$$

Thus far we have not restricted the choice of four or six-vector fields G_μ and $G_{\mu\nu}$. The considerations that follow are simplified if we consider a special case of the six-vector field $G_{\mu\nu}$ in which $G_{21} = 0$.[20] Substituting with these transformation formulae for f'_2 and G'_{24} in this special in (10b) and (11b), we recover

$$f_2 = m\sqrt{1-\frac{v^2}{c^2}}G_2 \tag{10c}$$

$$f_2 = mG_{24}ic. \tag{11c}$$

These two equations describe the component of gravitational three-force, f_2, in the "vertical" $x_2 = y$ direction on a mass m moving with velocity v in the "horizontal" $x_1 = x$ direction. In his 1912 argument, Einstein noted that the inertia of energy and the equality of inertial and gravitational mass leads us to expect that "gravitation acts more strongly on a moving body than on the same body in case it is at rest." We read directly from equations (10c) and (11c) that both four and six-vector theories fail to satisfy this condition. The gravitational force is independent of velocity in the six-vector case and actually decreases with velocity in the four-vector case. To meet Einstein's requirements, the gravitational force would need to increase with velocity, in direct proportion to the mass's energy $mc^2/\sqrt{1-(v^2/\mathrm{c}^2)}$.

We can also confirm that (10c) and (11c) lead to the result that the vertical acceleration of the masses is not independent of their horizontal velocities. To see this, note that, were the masses of (10c) and (11c) instantaneously at rest, the vertical forces exerted by the two fields would be respectively

$$f_2^{\mathrm{rest}} = mG_2, \quad f_2^{\mathrm{rest}} = mG_{24}ic.$$

In all cases, the three velocity and three-accelerations are perpendicular, so that condition (8) holds. Therefore we have

---

20 This restriction does not compromise the generality of Einstein's claim. If a Lorentz covariant theory proves inadequate in a special case, that is sufficient to demonstrate its general inadequacy. A natural instance of a six-vector field $G_{\mu\nu}$ in which $G_{21} = 0$ is easy to construct. Following the model of electromagnetism, we assume that $G_{\mu\nu}$ is generated by a vector potential A_μ, according to

$$G_{\mu\nu} = \frac{\partial A_\mu}{\partial x_\nu} - \frac{\partial A_\nu}{\partial x_\mu}.$$

We choose a "gravito-static" field in $S(x, y, z, u)$, that is, one that is analogous to the electrostatic field, by setting $A_\mu = (0, 0, 0, A_4)$. Since A_1 and A_2 are everywhere vanishing, $G_{21} = 0$. Finally note that Einstein does explicitly restrict his 1912 claim to static gravitational fields. Perhaps he also considered simplifying special examples of this type.

$$f_i = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}} m \frac{dv_i}{dt}, \quad f_i^{\text{rest}} = m\left(\frac{dv_i}{dt}\right)^{\text{rest}}.$$

Combining these results with (10c) and (11c) we recover expressions for the vertical acceleration dv_2/dt of the masses in terms of the acceleration $(dv_2/dt)^{\text{rest}}$ they would have had if they had no horizontal velocity

$$\frac{dv_2}{dt} = \left(1-\frac{v^2}{c^2}\right)\left(\frac{dv_2}{dt}\right)^{\text{rest}} \tag{10d}$$

$$\frac{dv_2}{dt} = \sqrt{1-\frac{v^2}{c^2}}\left(\frac{dv_2}{dt}\right)^{\text{rest}} \tag{11d}$$

We see that in both four and six-vector cases the vertical acceleration decreases with horizontal velocity, with equation (10d) generalizing the result in equation (9).

5. A GRAVITATION THEORY MODELLED AFTER MAXWELL'S ELECTROMAGNETISM?

Einstein's mention of a six-vector theory of gravitation in his 1912 response to Abraham raises the question of Einstein's attitude to a very obvious strategy of relativization of Newtonian gravitation theory. With hindsight one can view the transition from the theory of Coulomb electrostatic fields to full Maxwell electromagnetism as the first successful relativization of a field theory. Now Newtonian gravitation theory is formally identical to the theory of electrostatic fields excepting a change of sign needed to ensure that gravitational masses attract where like electric charges repel. This suggests that one can relativize Newtonian gravitation theory by augmenting it to a theory formally identical to Maxwell theory excepting this same change of sign.

While it is only with hindsight that one sees the transition from electrostatics to electromagnetism as a relativization, Einstein had certainly developed this hindsight by 1913. In his (Einstein 1913, 1250) he noted that Newtonian theory has sufficed so far for celestial mechanics because of the smallness of the speeds and accelerations of the heavenly bodies. Were these motions to be governed instead by electric forces of similar magnitude, one would need only Coulomb' s law to calculate these motions with great accuracy. Maxwell's theory would not be required. The problem of relativizing gravitation theory, Einstein continued, corresponded exactly to this problem: if we knew only experimentally of electrostatics but that electrical action could not propagate faster than light, would we be able to develop Maxwell electromagnetics? In the same paper Einstein proceeded to show (p. 1261) that his early 1913 version of general relativity reduced in suitable weak field approximation to a theory with a four-vector field potential that was formally analogous to electrodynamics. It was this approximation that yielded the weak field effects we now label as "Machian." The previous year, when seeking similar effects in his 1912 theory of static gravitational fields, Einstein demonstrated that he then expected a relativized gravitation theory to be formally analogous to electrodynamics at some level. For then he wrote a paper with the revealing title "Is there a gravitational effect that is analogous to electrodynamic induction?" (Einstein 1912c).

The celebrated defect of a theory of gravitation modelled after Maxwell electromagnetism was first pointed out by Maxwell himself (Maxwell 1864, 571). In such a theory, due to the change of signs, the energy density of the gravitational field is negative and becomes more negative as the field becomes stronger. In order not to introduce net negative energies into the theory, one must then suppose that space, in the absence of gravitational forces, must contain a positive energy density sufficiently great to offset the negative energy of any possible field strength. Maxwell professed himself baffled by the question of how a medium could possess such properties and renounced further work on the problem. As it turns out it was Einstein's foe, Abraham, shortly after his exchange with Einstein, who refined Maxwell's concern into a more telling objection. In a lecture of October 19, 1912, he reviewed his own gravitation theory based on Einstein's idea of using the speed of light as a gravitational potential. (Abraham 1912e) He first reflected (pp. 193–94), however, on a gravitation theory modelled after Maxwell electromagnetism. In such a theory, a mass, set into oscillation, would emit waves analogous to light waves. However, because of the change of sign, the energy flow would not be away from the mass but towards it, so that the energy of oscillation would increase. In other words such an oscillating mass would have no stable equilibrium. Similar difficulties were reported by him for gravitation theories of Maxwellian form due to H.A. Lorentz and R. Gans.

What was Einstein's attitude to such a theory of gravitation? He was clearly aware of the formal possibility of such a theory in 1912 and 1913. From his failure to exploit such a theory, we can only assume that he did not think it an adequate means of relativizing gravitation.[21] Unfortunately I know of no source from that period through which Einstein states a definite view on the matter beyond the brief remarks in his exchange with Abraham. We shall see that Einstein is about to renounce the conclusion of his reply to Abraham, that a Lorentz covariant theory cannot capture the equality of inertial and gravitational mass, at least for the case of Nordström's theory of gravitation. Did Einstein have other reservations about six-vector theories of gravitation? How seriously, for example, did he regard the negative field energy problem in such a theory?

The idea of an analogy between a relativized gravitation theory and electrodynamics seems to play no significant role in the methods Einstein used to generate relativized gravitation theories. The effect analogous to electrodynamic induction of

21 Notice these reservations must have amounted to more than the observation that such a theory fails to extend the relativity of motion to acceleration. In (Einstein 1913), immediately after his remarks on the similarity between the problems of relativizing gravitation and electrostatics, he considers Lorentz covariant gravitation theories. The only theory taken seriously in this category is a version of Nordström's theory of gravitation.

(Einstein 1912c), for example, was derived fully within Einstein's 1912 theory of static gravitational fields and the analogy to electrodynamics appeared only in the description of the final result. In general, the mention of an analogy to electrodynamics seems intended solely to aid Einstein's readers in understanding the enterprise and physical effects appearing in the relativized theories of gravitation by relating them to an example familiar to his readers. That we have any surviving, written remarks by Einstein directly on this matter we owe to J.W. Killian. Some thirty years later, in a letter of June 9, 1943 (EA 14 261) to Einstein, Killian proposed a gravitation theory modelled after Maxwell electromagnetism.[22] Einstein's reply of June 28, 1943, gives a fairly thorough statement of his attitude at that time to this theory.[23]

Because there was no question of experimental support for the theory, Einstein proposed to speak only to its formal properties. To begin, he noted, Maxwell's equations only form a complete theory for parts of space free of source charges, for the theory cannot determine the velocity field of the charge distribution without further assumption. After Lorentz, to form a complete theory, it was assumed that charges were carried by ponderable masses whose motions followed from Newton's laws. What Einstein called "real difficulties" arise only in explaining inertia. These difficulties result from the negative gravitational field energy density in the theory. Assuming, apparently, that the energy of a mass in the theory would reside in its gravitational field, Einstein pointed out that the kinetic energy of a moving mass point would be negative. This negativity would have to be overcome by a device entirely arbitrary from the perspective of the theory's equations, the introduction of a compensating positive energy density located within the masses. This difficulty is more serious for the gravitational version of the theory, for, in the electromagnetic theory, the positivity of electromagnetic field energy density allows one to locate all the energy of a charge in its electromagnetic field.

Calling the preceding difficulty "the fundamental problem of the wrong sign," Einstein closed his letter with brief treatment of two further and, by suggestion, lesser difficulties. The proposed theory could not account, Einstein continued, for the proportionality of inertial and gravitational mass. Here we finally see the concern that drove Einstein's work on Lorentz covariant gravitation theory in the decade following 1907. Yet Einstein does not use the transformation arguments of this early period to establish the failure of the proposed theory to yield this proportionality. Instead he continues to imagine that the energy and therefore inertia of a mass resides in its gravitational field. Some fixed quantity of gravitational mass could be configured in many different ways. Thus it follows that the one quantity of gravitational mass could be associated with many different gravitational fields and thus many different inertial masses, in contradiction with the proportionality sought.[24] Finally Einstein remarked

---

22 Einstein addresses his reply to "Mr. J.W. Killian, Dept. of Physics, Rockerfeller Hall, Ithaca N.Y."

23 EA 14 265 is an autograph draft of the letter in German. EA 14 264 is an unsigned typescript of the English translation. There are some significant differences of content between the two, indicating further editing of content presumably by Einstein between the draft and typescript.

that the proposed theory allows no interaction between electromagnetic and gravitational fields other than through charged, ponderable masses. Thus it could not explain the bending of starlight in a gravitational field.

6. NORDSTRÖM'S FIRST THEORY OF GRAVITATION

The dispute between Einstein and Abraham was observed with interest by a Finnish physicist, Gunnar Nordström.[25] In a paper submitted to *Physikalische Zeitschrift* in October 1912 (Nordström 1912), he explained that Einstein's hypothesis that the speed of light c depends on the gravitational potential led to considerable problems such as revealed in the Einstein-Abraham dispute. Nordström announced (p. 1126) that he believed he had found an alternative to Einstein's hypothesis which would

> ... leave c constant and still adapt the theory of gravitation to the relativity principle in such a way that gravitational and inertial masses are equal.

The theory of gravitation which Nordström developed was a slight modification of the theory embodied in equations (3) and (4) above. Selecting the commonly used coordinates $x, y, z, u = ict$, Nordström gave his version of the field equation (3):

$$\frac{\partial^2\Phi}{\partial x^2}+\frac{\partial^2\Phi}{\partial y^2}+\frac{\partial^2\Phi}{\partial z^2}+\frac{\partial^2\Phi}{\partial u^2} = 4\pi f\gamma \qquad (12)$$

where Φ is the gravitational potential, γ the rest density of matter and f the gravitational constant. As he noted, this field equation was identical to the one advanced by Abraham (1912a, equation (1)) in the latter's gravitation theory.

Where Nordström differed from Abraham, however, was in the treatment of the force equation (4). This equation, as Nordström pointed out, is incompatible with the constancy of c. We saw above that force equation (4), in conjunction with the constancy of c in equation (5) entails the unphysical condition (6). Abraham had resolved the problem by invoking Einstein's hypothesis that c not be constant but vary with gravitational potential so that condition (5) no longer obtains. Thus Abraham's gravitation theory was no longer a special relativistic theory. Nordström, determined to preserve special relativity and the constancy of c, offered a choice of two modified versions of (4).

First, one could allow the rest mass m of a body in a gravitational field to vary with gravitational potential. Defining

---

24 In the German autograph draft (EA 14 265), Einstein imagines some fixed quantity of gravitational mass distributed between two bodies. The field strength they generate, and therefore their energy and inertial mass, would increase as the bodies were concentrated into smaller regions of space. (We may conjecture here that Einstein is ignoring the fact the field energy becomes more *negative* as the field strength increases.) The translated typescript (EA 14 264) simplifies the example by imagining that the gravitational mass is located in a single corpuscle, whose field and thence inertia varies with the radius of the corpuscle.

25 For a brief account of Nordström's life and his contribution to gravitation theory see (Isaksson 1985).

$$\mathfrak{F}_\mu = -\frac{\partial\Phi}{\partial x_\mu},$$

the four force on a body of mass m is

$$m\mathfrak{F}_\mu = -m\frac{\partial\Phi}{\partial x_\mu} = \frac{d}{d\tau}(m\mathfrak{a}_\mu) = m\frac{d\mathfrak{a}_\mu}{d\tau} + \mathfrak{a}_\mu\frac{dm}{d\tau} \tag{13}$$

where $\mathfrak{a}_\mu$ is the mass' four velocity and τ proper time.[26] The dependence of m on Φ introduces the additional, final term in $dm/d\tau$, which prevents the derivation of the disastrous condition (6). In its place, by contracting (13) with $\mathfrak{a}_\mu = \frac{dx_\mu}{dt}$ and noting that $\mathfrak{a}_\mu\mathfrak{a}_\mu = -c^2$, Nordström recovered the condition

$$m\frac{d\Phi}{d\tau} = c^2\frac{dm}{d\tau} \tag{14}$$

which yields an expression for the Φ dependence of m upon integration

$$m = m_0\exp\left(\frac{\Phi}{c^2}\right) \tag{15}$$

where m_0 is the value of m when $\Phi = 0$. Using (14) to substitute $\frac{dm}{d\tau}$ in (13), Nordström then recovered an equation of motion for a mass point independent of m

$$-\frac{\partial\Phi}{\partial x_\mu} = \frac{d\mathfrak{a}_\mu}{d\tau} + \frac{\mathfrak{a}_\mu}{c^2}\frac{d\Phi}{d\tau}. \tag{16}$$

---

26 *Note on notation*: The notation used in the sequence of papers discussed here varies. I shall follow the notation of the original papers as it changes, with one exception for brevity. Where the components of an equation such as (13) were written out explicitly as four equations

$$-m\frac{\partial\Phi}{\partial x} = \frac{d}{d\tau}(m\mathfrak{a}_x) = m\frac{d\mathfrak{a}_x}{d\tau} + \mathfrak{a}_x\frac{dm}{d\tau},$$
$$-m\frac{\partial\Phi}{\partial y} = \frac{d}{d\tau}(m\mathfrak{a}_y) = m\frac{d\mathfrak{a}_y}{d\tau} + \mathfrak{a}_y\frac{dm}{\tau},$$
etc.

I silently introduce the coordinates $x_\mu = (x_1, x_2, x_3, x_4) = (x, y, z, u) = ict$ and corresponding index notation as in equation (13) above.

Nordström's second alternative to force equation (4) preserved the independence of m from the potential. The quantity $m\mathfrak{F}_\mu = -m\frac{\partial\Phi}{\partial x_\mu}$ could not be set equal to the gravitational four-force on a mass m, $\frac{d}{d\tau}(m\mathfrak{a}_\mu)$, for that would be incompatible with the orthogonality (5) of four velocity and four acceleration. However one can retain compatibility with this orthogonality if one selects as the four force only that part of $m\mathfrak{F}_\mu$ which is orthogonal to the four-velocity $\mathfrak{a}_\mu$. This yields the second alternative for the force equation

$$\frac{d}{d\tau}(m\mathfrak{a}_\mu) = m\mathfrak{F}_\mu + m\frac{\mathfrak{a}_\mu}{c^2}(\mathfrak{F}_\nu\mathfrak{a}_\nu) = \begin{pmatrix}\text{part of } m\mathfrak{F}_\mu \\ \text{orthogonal} \\ \text{to } \mathfrak{a}_\mu\end{pmatrix}, \tag{17}$$

Nordström somewhat casually noted that he would use the first alternative, since it corresponded to "the position of most researchers in the domain of relativity theory." (p. 1126) Indeed Nordström proceeded to show that both force equations lead to exactly the same equation of motion (16) for a point mass, planting the suggestion that the choice between alternatives could be made arbitrarily.

Regular readers of *Physikalische Zeitschrift*, however, would know that Nordström's decision between the two alternatives could not have been made so casually by him. For in late 1909 and early 1910, Nordström had engaged in a lively public dispute with none other than Abraham on a problem in relativistic electrodynamics that was in formal terms virtually the twin of the choice between the force laws (13) and (17). (Nordström 1909, 1910; Abraham 1909, 1910.) The problem centered on the correct expression for the four force density on a matter distribution in the case of Joule heating. The usual formula for the four force density $\mathfrak{K}_\mu$ on a mass distribution with rest mass density v is, in the notation of Abraham (1910),

$$\mathfrak{K}_\mu = \nu\frac{d}{d\tau}\left(\frac{dx_\mu}{d\tau}\right)$$

with τ proper time and coordinates $x_\mu = (x, y, z, u = ict)$. In the case of Joule heating, it turns out that this expression leads to a contradiction with the orthogonality condition (5). The two escapes from this problem at issue in the dispute are formally the same as the two alternative gravitational force laws. Nordström defended Minkowski's approach, which took the four force density to be that part of $\mathfrak{K}_\mu$ orthogonal to the matter four velocity $\left(\frac{dx_\mu}{d\tau}\right)$ —the counterpart of force law (17).

Abraham concluded that the rest mass density v would increases in response to the energy of Joule heat generated. He showed a consistent system could be achieved if one now imported this variable v into the scope of the $d/d\tau$ operator in the expression for $\mathfrak{K}_\mu$ —an escape that is the counterpart of (13). Abraham's escape was judged the only tenable one when he was able to show that it yielded the then standard Lorentz transformation formula for heat whereas the Nordström-Minkowski formula did not.[27]

The connection between Nordström's 1912 gravitation theory and this earlier dispute surfaced only in extremely abbreviated form in Nordström (1912). In a brief sentence in the body of the paper, Nordström noted gingerly that (p. 1127)

> the latter way of thinking [alternative (17)] corresponds to Minkowski's original, that treated first [alternative (13)] to that held by Laue and Abraham.

That Nordström had any stake in the differing viewpoints is only revealed in a footnote to this sentence in which the reader is invited to consult "the discussion between Abraham and the author," followed by a citation to the four papers forming the dispute. Nordström then closes with the remark[28]

> I now take the position then taken up by Abraham.

Nordström now continued his treatment of gravitation by extending the discussion from isolated point masses to the case of continuous matter distributions—an area in which he had some interest and expertise (Nordström 1911). He derived a series of results in a straightforward manner. They included expressions for gravitational four force density on a continuous mass distributions and the corresponding equations of motion, expressions for the energy density and flux due to both gravitational field and matter distribution, the gravitational field stress-energy tensor and the laws of conservation of energy and momentum.

The last result Nordström derives concerns point masses. He notes that the field equation (12) admits the familiar retarded potential as a solution for a matter distribution with rest density γ

$$\Phi(x_0, y_0, z_0, t) = -f\int\frac{dxdydz}{r}\gamma_{t-r/c} + \text{constant}$$

where $\gamma_{t-r/c}$ is γ evaluated at time $t-r/c$, the integration extends over all of three dimensional space and $r = \sqrt{(x-x_0)^2+(y-y_0)^2+(z-z_0)^2}$. It follows from the factor of $1/r$ in the integral that the potential Φ at a true point mass would be $-\infty$. Allowing for the dependence of mass on potential given in (15), it follows that the mass of such a point would have to be zero so that true point masses cannot exist.

27 Thus the authoritative judgement of Pauli's Teubner Encyklopädie article (Pauli 1921, 108) is that "Nordström's objections cannot be upheld." For a lengthy discussion of this debate and an indication that the issues were not so simple, see (Liu 1991).

28 His earlier work (Nordström 1911) had explicitly employed Abraham's "force concept," although Nordström had then noted very evasively that, in using it, he "wish[es] to assert no definite opinion on the correctness of one or other of the two concepts" (p. 854).

Nordström concluded with confidence, however, that he could see no contradictions arising from this result.

We may wonder at Nordström's lack of concern over this result. It would be thoroughly intelligible, however, if Nordström were to agree—as Nordström's later (Nordström 1913a, 856) suggests —with Laue's view on the relation between the theory of point masses and of continua. Laue had urged that the former ought to be derived from the latter (Laue 1911a, 525). Under this view, the properties of extended masses are derived from consideration of discrete volumes in a continuous matter distribution, not from the accumulated behavior of many point masses. So the impossibility of point masses in Nordström's theory would present no obstacle in his generation of the behavior of extended bodies.

7. EINSTEIN REPLIES

In advancing his theory, Nordström had claimed to do precisely what Einstein had claimed impossible: the construction of a Lorentz covariant theory of gravitation in which the equality of inertial and gravitational mass held. We need not guess whether Einstein communicated his displeasure to Nordström, for Einstein's missive was sufficiently swift for Nordström to acknowledge it in an addendum (p. 1129) to his paper which read

> *Addendum to proofs*. From a letter from Herr Prof. Dr. A. Einstein I learn that he had already earlier concerned himself with the possibility used above by me for treating gravitational phenomena in a simple way. He however came to the conviction that the consequences of such a theory cannot correspond with reality. In a simple example he shows that, according to this theory, a rotating system in a gravitational field will acquire a smaller acceleration than a non-rotating system.

Einstein's objection to Nordström is clearly an instance of his then standard objection to Lorentz covariant theories of gravitation: in such theories the acceleration of fall is not independent of a body's energy so that the equality of inertial and gravitational mass is violated. It is not hard to guess how Einstein would establish this result for a spinning body in Nordström's theory. It would seem to follow directly from the familiar equation (9) which holds in Nordström's theory and which says, loosely speaking, that a body falls slower if it has a greater horizontal velocity. Indeed, as we shall see below, this is precisely how Nordström shortly establishes the result in his next paper on gravitation theory.

Nordström continued and completed his addendum with a somewhat casual dismissal of Einstein's objection.

> I do not find this result dubious in itself, for the difference is too small to yield a contradiction with experience. Of course, the result under discussion shows that my theory is not compatible with Einstein's principle of equivalence, according to which an unaccelerated reference system in a homogeneous gravitational field is equivalent to an accelerated reference system in a gravitation free space.
>
> In this circumstance, however, I do not see a sufficient reason to reject the theory. For, even though Einstein's hypothesis is extraordinarily ingenious, on the other hand it still

> provides great difficulties. Therefore other attempts at treating gravitation are also desirable and I want to provide a contribution to them with my communication.

Nordström's reply is thoroughly reasonable. The requirement of exact equality of inertial and gravitational mass was clearly an obsession of Einstein's thinking at this time and not shared by Einstein's contemporaries. The now celebrated Eötvös experiment had not yet been mentioned in the publications cited up to this point. We can also see from equation (9) that the failure of the equality of inertial and gravitational mass implied by Nordström's theory would reside in a second order effect in v/c. Nordström clearly believed it to be beyond recovery from then available experiments.

Finally we should note that Einstein and Nordström are using quite different versions of the requirement of the equality of inertial and gravitational mass. At this time, Einstein presumed that the total inertial mass would enter into the equality with the expectation that it would yield the independence of the vertical acceleration of a body in free fall from its horizontal velocity. We can only conjecture the precise sense Nordström had in mind, when he promised his theory would satisfy the equality, for he does not explain how the equality is expressed in his theory. My guess is that he took the rest mass to represent the body's inertial mass in the equality, for Nordström's equation (16) clearly shows that the motion of a massive particle in free fall is independent of its rest mass m. Under this reading Nordström's version of the equality entails the weaker observational requirement of the "uniqueness of free fall" defined above in Section 1.[29]

8. NORDSTRÖM'S FIRST THEORY ELABORATED

Nordström's first paper on his gravitation theory was followed fairly quickly by another (Nordström 1913a), submitted to *Annalen der Physik* in January 1913. This new paper largely ignored Einstein's objection although the paper bore the title "Inertial and gravitational mass in relativistic mechanics." The closing sections of this paper recapitulated the basic results of (Nordström 1912) with essentially notational differences only. In Section 6, Nordström's original field equation (12) was rewritten as

$$\frac{\partial^2\Phi}{\partial x^2}+\frac{\partial^2\Phi}{\partial y^2}+\frac{\partial^2\Phi}{\partial z^2}+\frac{\partial^2\Phi}{\partial u^2} = g\nu. \tag{12'}$$

As before, Φ was the gravitational potential. The rest density of matter was now represented by ν and Nordström explicitly named the new constant g the "gravitation factor." The force equation was presented as a density in terms of the gravitational force per unit volume of matter $\mathfrak{K}^g_\mu$

29 The equality of inertial and gravitational mass and the uniqueness of free fall are distinct from the principle of equivalence. Einstein's version of the principle has been routinely misrepresented since about 1920 in virtually all literatures. See (Norton 1985). It is stated correctly, however, in Nordström's addendum.

$$\mathfrak{K}_{\mu}^{g} = -g\nu\frac{\partial\Phi}{\partial x_{\mu}}. \tag{18}$$

The only difference between this expression and the analogous one offered in the previous paper was the presence of the gravitation factor g. Since this factor was a constant and thus did not materially alter the physical content of either of the basic equations, Nordström might well have anticipated his readers' puzzlement over its use. He hastened to explain that, while g was a constant here, nothing ruled out the assumption that g might vary with the inner constitution of matter. The paper continued to derive the Φ dependence of rest mass m of equation (15). The new version of the relation now contained the gravitation factor g and read

$$m = m_0 \exp\left(\frac{g\Phi}{c^2}\right).$$

Nordström no longer even mentioned the possibility of avoiding this dependence of m on Φ by positing the alternative force equation (17). The section continued with a brief treatment of the gravitational field stress-energy tensor and related quantities. It closed with a statement of the retarded potential solution of the field equation.

The final section 7 of the paper analyzed the motion of a point mass in free fall in an arbitrary static gravitational field. The analysis was qualified by repetition of his earlier observation that true point masses are impossible in his theory (Nordström 1912, 1129). In addition he noted that the particle's own field must be assumed to be vanishingly weak in relation to the external field. The bulk of the section is given over to a tedious but straightforward derivation of the analog of equation (9). Nordström considered a static field, that is one in which $\partial\Phi/\partial t = 0$, where t is the time coordinate of the coordinate system $(x, y, z, u = ict)$. He assumed the field homogeneous and acting only in the z-direction of the coordinate system. A point mass in free fall moves according to

$$\frac{d\mathfrak{v}_z}{dt} = -\left(1-\frac{\mathfrak{v}^2}{c^2}\right)g\frac{\partial\phi}{\partial z}, \quad \frac{d\mathfrak{v}_x}{dt} = 0, \quad \frac{d\mathfrak{v}_y}{dt} = 0, \tag{19}$$

where $\mathfrak{v}_x$, $\mathfrak{v}_y$ and $\mathfrak{v}_z$ are the components of the mass' velocity $\mathfrak{v}$.[30] At this point in the paper, readers of (Nordström 1912) might well suspect that the entire purpose of developing equation (19) was to enable statement of the objection of Einstein reported in that last paper's addendum. For, after observing that this result (19) tells us that a body with horizontal velocity falls slower than one without, he concluded immediately that a rotating body must fall slower than a non-rotating body.

30 Notice that this result is more general than result (9), since it is not restricted to masses with vanishing vertical velocity, that is, masses whose motion satisfies the condition (7). Curiously Nordström's condition that the field be homogeneous, so that $\partial\Phi/\partial z =$ constant, is invoked nowhere in the derivation or discussion of the result.

Because this example will be reappraised shortly, it is worth inserting the steps that Nordström must have assumed to arrive at this conclusion. In the simplest case, the axis of rotation of the body is aligned vertically in the static field. Each small element of the spinning body has a horizontal motion, due to the rotation. If each such element were independent, then the vertical acceleration of each would be given by the equation (19), so that each element would fall slower because of the horizontal velocity imparted by the rotation. If this result holds for each element, it seems unproblematic to conclude that it obtains for the whole, so that the vertical acceleration of fall of the body is diminished by its rotation.

While Nordström urged that this effect is much too small to be accessible to observation, he was more sanguine about the analogous effect on the acceleration of fall of a body by the independent motions of its molecules. Its possibility could not be denied, he said. However, in the penultimate paragraph of the paper, he anticipated that such an effect could be incorporated into his theory by allowing the gravitation factor g to depend on the molecular motion of the body. He pointed out that the rest energy of a body would also be influenced by this molecular motion.

The results Nordström recapitulated in Section 6 and 7 were not the major novelties of the paper. In fact the paper was intended to address a quite precise problem. The field equation (12') contained a density term ν. Nordström's problem was to identify what this term should be. The term—or, more precisely, $g\nu$—represented the gravitational field source density. According to Nordström's understanding of the equality of inertial and gravitational mass, ν must also represent the inertial properties of the source matter. The selection of such a term was not straightforward. For, drawing upon his own work and that of Laue and others, he knew that stressed bodies would exhibit inertial properties that were not reducible to the inertial properties of any individual masses that may compose them. Thus Nordström recognized that his gravitation theory must be developed by means of the theory of relativistic continua, in which stresses were treated. This had clearly been his program from the start. In a footnote to the first paragraph of his first paper on gravitation, Nordström (1912), at the mention of the equality inertial and gravitational mass, he foreshadowed his next paper

> By the equality of inertial and gravitational mass, I do not understand, however, that every inertial phenomenon is caused by an inertial and gravitational mass. For elastically stressed bodies, according to Laue ..., one recovers a quantity of motion [momentum] that cannot at all be reduced back to a mass. I will return to this question in a future communication.

The special behavior of stressed bodies proved to be of decisive importance for the development of Nordström's theory. Therefore, in the following section, I review the understanding of this behavior at the time of Nordström's work on gravitation. I will then return to (Nordström 1913a).

9. LAUE AND THE BEHAVIOR OF STRESSED BODIES

By 1911 it was apparent that a range of problems in the theory of relativity had a common core—they all involved the behavior of stressed bodies—and that a general theory of stressed bodies should be able to handle all of these problems in a unified format. The development of this general theory was largely the work of Laue and came from a synthesis and generalization of the work of many of his predecessors, including Einstein, Lorentz, Minkowski and Planck. The fullest expression of this general theory came in Laue (1911a) and was also incorporated into Laue (1911b), the first text book published on the new theory of relativity.[31] Three problems treated in Laue' s work give us a sense of the range of problems that Laue's work addressed.

9.1 Three Problems for Relativity Theory

In 1909, in a remarkably prescient paper, Lewis and Tolman (1909) set out to develop relativistic mechanics in a manner that was independent of electromagnetic theory using simple and vivid arguments. At this time relativity theory was almost invariably coupled with Lorentzian electrodynamics and its content was accessible essentially only to those with significant expertise in electrodynamics. Their exposition was marred, however, by an error in its closing pages (pp. 520–21). By this point, they had established the Lorentz transformation for forces transverse to the direction of motion. Specifically, if the force is f^0_{trans} in the rest frame, then the force f_{trans} measured in a frame moving at a fraction β of the speed of light is

$$f_{\text{trans}} = \sqrt{1-\beta^2}\, f^{\,0}_{\text{trans}}. \tag{20}$$

To recover the transformation formula for forces parallel to the direction of motion, the "longitudinal" direction, they considered the rigid, right angled lever of Fig. 1. The arms ab and bc are of equal length and pivot about point b. In its rest frame two equal forces f act at points a and c, the first in direction bc, the second in direction ba. The level will not turn since there is no net turning couple about its pivot b. They then imagined the whole system in motion in the direction bc. They conclude—presumably directly from the principle of relativity—that the system must remain in equilibrium. Therefore the net turning couple about b must continue to vanish for the moving system, so that

$$\begin{pmatrix}\text{force}\\ \text{at } c\end{pmatrix}\begin{pmatrix}\text{length}\\ bc\end{pmatrix}+\begin{pmatrix}\text{force}\\ \text{at } a\end{pmatrix}\begin{pmatrix}\text{length}\\ ab\end{pmatrix} = 0.$$

Now, according to (20), the transverse force at c is diminished by the factor $\sqrt{1-\beta^2}$. The length of its lever arm bc is also contracted by the same factor,

31 Presumably the two works were prepared together. (Laue 1911a) was submitted on 30 April, 1911. The introduction to (Laue 1911b) is dated May, 1911.

whereas the arm ab, being transverse to the motion, is uncontracted. Lewis and Tolman now concluded that equilibrium can only be maintained if the longitudinal force f_{long} at a transforms according to

$$f_{\text{long}} = (1-\beta^2)f^0_{\text{long}}. \tag{21}$$

This conclusion comes from an argument so simple that one would hardly suspect it. What they did not point out, however, was that its conclusion (21) contradicted the then standard expositions of relativity theory (e.g. Einstein 1907a, 448) according to which (20) is correct but (21) should be replaced by

$$f_{\text{long}} = f^0_{\text{long}}. \tag{22}$$

We now see the problem in its starkest form. If we apply the standard transformation formulae (20) and (22) to the case of Lewis and Tolman's bent lever we seem driven to a curious conclusion. We have a system at equilibrium in its rest frame which now forfeits that equilibrium in a moving frame through the appearance of a non-vanishing turning couple. Indeed we seem to have a violation of the principle of relativity, for the presence of this turning couple should yield an experimental indication of the motion of the system.

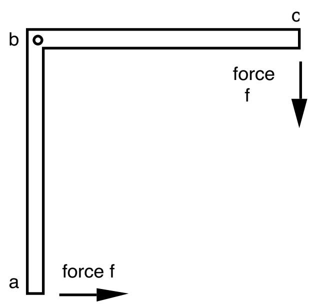


Figure 1: Lewis and Tolman's Bent Lever

The second problem is, at first glance, quite unrelated to the Lewis and Tolman bent lever. Under a classical analysis, one expects that a charged, parallel plate condenser can experience a net turning couple if it is set in motion through the aether. In a classic experiment, Trouton and Noble (1903) sought to detect the turning couple acting on a charged condenser due to its motion with the Earth. Their null result is celebrated. Just as in the case of the Lewis and Tolman bent lever, the problem is to see how relativity theory allows one to predict this null result, which otherwise would contradict the principle of relativity. In fact, as Laue (1911c, 517) and others soon pointed out, the two problems were closely connected. In its rest frame, the Trouton-Noble condenser was simply a rigid system of two parallel plates with an electric force acting on each plate in such a way that the entire system was in equilibrium. If that equilibrium system was set into motion, under either a classical or relativistic analysis, the electric forces would transform according to the Lorentz transformation (20) and (22). Unless the direction of motion imparted was parallel or exactly perpendicular to the plates, the net effect would be exactly the same as the Lewis and Tolman bent lever. A non-vanishing turning couple is predicted which deprives the system of equilibrium. The couple ought to be detectable in violation of the principle of relativity.[32]

The third problem concerns the theory of electrons. The decade preceding 1911 had seen considerable work on the problem of providing a model for the electron. Best known of these were the models of Lorentz and Abraham, which depicted electrons as electrically charged spheres with varying properties. The general problem was to show that the relativistic dynamics of an acceptable model of the electron would coincide with the relativistic dynamics of a point mass. There were a range of difficulties to be addressed here. In introducing his (Laue 1911a, 524–25), Laue recalled a brief exchange between Ehrenfest and Einstein. In a short note, Ehrenfest (1907) had drawn on work of Abraham that raised the possibility of troubling behavior by an electron of non-spherical or non-ellipsoidal shape when at rest. It was suggested that such an electron cannot persist in uniform translational motion unless forces are applied to it.[33] We might note that such a result would violate not only the principle of inertia in the dynamics of point masses but also the principle of relativity. Einstein's reply (1907b) was more a promise than resolution, although he ultimately proved correct. He pointed out that Ehrenfest's model of the electron was incomplete. One must also posit that the electron's charge was carried by a rigid frame, stressed to counteract the forces of self repulsion of the charge distribution. Ehrenfest's problem could not be solved until a theory of such frames was developed.

Finally another aspect of the problem of the relativistic dynamics of electrons was the notorious question of electromagnetic mass. If one computed the total momentum and energy of the electromagnetic field of an electron, the result universally accepted at this time was the one reported in (Laue 1911b, 98):

$$\begin{pmatrix}\text{Total field}\\\text{momentum}\end{pmatrix} = \frac{4}{3}\frac{1}{c^2}\begin{pmatrix}\text{Total field}\\\text{energy}\end{pmatrix}\begin{pmatrix}\text{electron}\\\text{velocity}\end{pmatrix}. \tag{23}$$

The conflict with the relativistic dynamics of point masses arose if one now posited that all the energy and momentum of the electron resides in its electromagnetic field. For one must then identify $\frac{1}{c^2}\begin{pmatrix}\text{Total field}\\\text{energy}\end{pmatrix}$, the electromagnetic mass of the electron,

32 For further extensive discussion of the Trouton-Noble experiment and its aether theoretic treatment by Lorentz, see (Janssen 1995).

33 In a footnote, Ehrenfest pointed out the analogy to the turning couple induced on a charged condenser and reviewed the then current explanation of it absence in terms of molecular forces.

as the total inertial mass of the electron, so that equation (23) tells us that the momentum of an electron is $4/3$ the product of its mass and velocity. The canonical resolution of this difficulty, as stated for example in (Pauli 1921, 185–86), is that such a purely electromagnetic account of the dynamics of the electron is inadmissible. As Einstein (1907b) urged, there must be also stresses of a non-electromagnetic character within the electron.[34] The puzzle Laue addressed in 1911 was to find very general circumstances under which the dynamics of such an electron would agree with the relativistic dynamics of point masses.

9.2 The General-Stress Energy Tensor

The focus of Laue's treatment of stressed bodies in his (1911a) and (1911b) lay in a general stress-energy tensor.[35] While Minkowski (1908, §13) had introduced the four dimensional stress-energy tensor at the birth of four dimensional methods in relativity theory, his use of the tensor was restricted to the special case of the electromagnetic field. Laue's 1911 work concentrated on extending the use of this tensor to the most general domain. The properties of this tensor and its behavior under Lorentz transformation summarized a great deal of the then current knowledge of the behavior of stressed bodies. Laue (1911a) uses a coordinate system $(x, y, z, l = ict)$ so that the components of the stress energy tensor $T_{\mu\nu}$ have the following interpretations:

$$\begin{pmatrix} T_{xx} = p_{xx} & T_{xy} = p_{xy} & T_{xz} = p_{xz} & T_{xl} = -ic\mathfrak{g}_x \\ T_{yx} = p_{yx} & T_{yy} = p_{yy} & T_{yz} = p_{yz} & T_{yl} = -ic\mathfrak{g}_y \\ T_{zx} = p_{zx} & T_{zy} = p_{zy} & T_{zz} = p_{zz} & T_{zl} = -ic\mathfrak{g}_z \\ T_{lx} = \frac{i}{c}\mathfrak{S}_x & T_{ly} = \frac{i}{c}\mathfrak{S}_y & T_{lz} = \frac{i}{c}\mathfrak{S}_z & T_{ll} = -W \end{pmatrix}.$$

The three dimensional tensor $p_{ik}(i, k = 1, 2, 3)$ is the familiar stress tensor. The vector $\mathfrak{g} = (\mathfrak{g}_x, \mathfrak{g}_y, \mathfrak{g}_z)$ represents the momentum density. The vector $\mathfrak{S} = (\mathfrak{S}_x, \mathfrak{S}_y, \mathfrak{S}_z)$ is the energy flux. W is the energy density.

The most fundamental result of relativistic dynamics is Einstein's celebrated inertia of energy according to which every quantity of energy E is associated with an inertial mass (E/c^2). The symmetry of Laue's tensor entails a result closely con-

---

34 While these stresses are needed to preserve the mechanical equilibrium of the electron, Rohrlich (1960) showed that they were not needed to eliminate the extraneous factor of 4/3 in equation (23). He showed that the standard derivation of (23) was erroneous and that the correct derivation did not yield the troubling factor of 4/3.

35 The label "stress-energy tensor" is anachronistic. Laue had no special name for the tensor other than the generic "world tensor," which, according to the text book exposition of Laue (1911b, §13) described any structure which transformed as what we would now call a second rank, symmetric tensor. Notice that the term "tensor" was still restricted at this time to what we would now call second rank tensors and even then usually to symmetric, second rank tensors. See (Norton 1992a, Appendix).

nected with Einstein's inertia of energy and attributed to Planck by Laue (1911a, 530). We have $(T_{xl}, T_{yl}, T_{zl}) = (T_{lx}, T_{ly}, T_{lz})$ which immediately leads to

$$g = \frac{1}{c^2}\mathfrak{S} \tag{24}$$

This tells us that whenever there is an energy flux $\mathfrak{S}$ present in a body then there is an associated momentum density g.

As emphasized in (Laue 1911c), this result is already sufficient to resolve the first of the three problems described above in Section 9.1, the Lewis and Tolman bent lever. Notice first that Fig. 1 does not display all the forces present. There must be reaction forces present at the pivot point b to preserve equilibrium in the rest frame. See Fig. 2, which also includes the effect of the motion of the system at velocity v in the direction bc. When the lever moves in the direction bc, then work is done by the force f at point a which acts in the direction bc. The energy of this work is transmitted along the arm ab as an energy current of magnitude fv and is lost at the pivot point b as work done against the reaction force that acts in the direction cb. This energy current fv in the arm ab must be associated with a momentum fv/c^2, according to (24) when integrated over the volume of the arm ab, and this momentum will be directed from a towards b. As Laue (1911c) showed, a short calculation reveals that this momentum provides precisely the additional turning couple needed to return the moving system to equilibrium. Notice that the force at c and its associated reaction force at b are directed transverse to the motion so no work is done by them.[36]

The essential and entirely non-classical part of the analysis resides in the result that the there is an additional momentum present in the moving arm ab because it is under the influence of a shear stress due to the force f at a and the corresponding reaction force at b. As Laue (1911c, 517) and Pauli (1921, 128–29) point out, exactly this same relativistic effect explains the absence of net turning couple in the Trouton-Noble condenser. The condenser's dielectric must be stressed in reaction to the attractive forces between the oppositely charged plates. The additional momentum associated with these non-electromagnetic stresses provides the additional turning couple required to preserve the equilibrium of the moving condenser.

---

36 Assume that the level arms are of unit length at rest so that the arm bc contracts to length $\sqrt{1 - v^2/c^2}$ when the system moves at velocity v in direction bc. The turning couple about a point b' at rest and instantaneously coincident with the moving pivot point b, due to the applied forces alone is $f(1 - v^2/c^2) - f = -f(v^2/c^2)$ where a positive couple is in the clockwise direction. The relativistic momentum fv/c^2 generates angular momentum about the point b'. Since the distance of the arm ab from b' is growing at the rate of v, this angular momentum is increasing at a rate $v(fv/c^2) = f(v^2/c^2)$, which is exactly the turning couple needed to balance the couple due to the applied forces.

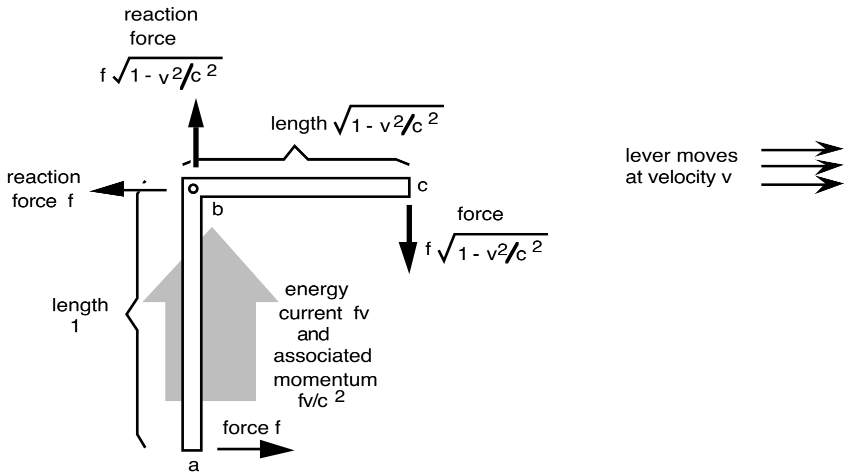


Figure 2: Lewis and Tolman's Bent Lever Showing Reaction Forces and Effects of Motion

While this analysis satisfactorily resolves at least the case of the Lewis and Tolman lever, more complicated cases will require a clearer statement of the relationship between the stresses in a moving body and the momentum associated with it. These results were derived directly by Laue (1911a, 531–32) from the Lorentz transformation of the components of the stress-energy tensor. In the rest frame of the stressed matter distribution, the matter has energy density W^0 and a stress tensor p^0_{ik}, for $i, k = 1, 2, 3$. The momentum density $\mathfrak{g}$ and energy flux $\mathfrak{S}$ vanish. Transforming to a frame of reference moving at velocity q in the x direction, from a direct application of the Lorentz transformation formula for a tensor, Laue recovered results that included

$$\mathfrak{g}_x = \frac{q}{c^2 - q^2}(p^0_{xx} + W^0), \quad \mathfrak{g}_y = \frac{q}{c\sqrt{c^2 - q^2}}p^0_{xy}, \quad W = \frac{c^2 W^0 + q^2 p^0_{xx}}{c^2 - q^2}. \qquad (25)$$

The first two equations show that, in a moving body, there is a momentum density associated with both normal stresses (p^0_{xx}) and with shear stresses (p^0_{xy}). The Lewis and Tolman lever is a case in which a momentum density is associated with shear stresses in a moving body in accord with (25). The third equation shows that there is an energy density associated with normal stresses in moving body.

This last result, in a form integrated over a whole stressed body, had already been investigated and clearly stated by Einstein (1907c, §1), as a part of his continuing analysis of the inertia of energy. He gave the result a very plausible, intuitive basis, relating it directly to the relativity of simultaneity. He imagined a rigid body in uniform translational motion. At some single instant in the body's rest frame, an equilibrium state of stress appears in the body. Since it appears at a single instant, the new forces do not alter the state of motion of the body. However in the frame in which the body moves, because of the relativity of simultaneity, these new forces do not appear simultaneously over the entire body. Thus there is a brief period of disequilibrium of forces during which net work is done on the body. This new work is exactly the energy associated with the stresses p_{xx}^0 in the equation (25). Einstein (1907a, §2) continued with an example similar in structure to the Trouton-Noble condenser—a rigid body, in uniform motion, carrying an electric charge distribution. The forces between the charges carried stress the rigid body, so that there is an energy associated with these stresses. Einstein showed that this latter energy was essential. Otherwise the energy of the moving body would depend on the direction of its motion which would lead to a contradiction.[37]

Laue (1911, §2) continued his treatment of the transformation formulae (25) by restating them for an extended body. In particular, integration of (25) over such a body revealed relationships between the rest energy E^0 of the body and its energy E and momentum $\mathfrak{G}$ in the frame of reference in which the body moves at velocity q. Writing the three components of the body's ordinary velocity as (q_1, q_2, q_3), he recovered[38]

$$\begin{aligned} E &= \frac{c}{\sqrt{c^2-q^2}}\left\{E^0 + \frac{1}{c^2}q_i q_k \int p_{ik}^0 dV^0\right\} \\ \mathfrak{G}_i &= \frac{q_i}{c\sqrt{c^2-q^2}}\left\{E^0 + \frac{1}{q^2}q_j q_k \int p_{jk}^0 dV^0\right\} \\ &\quad + \frac{1}{c^2}\left\{q_k \int p_{ik}^0 dV^0 - \frac{q_i}{q^2} q_j q_k \int p_{jk}^0 dV^0\right\}. \end{aligned} \tag{26}$$

The expression for momentum had an immediate and important consequence. In general, whenever the body was stressed so that the stress tensor p_{ik}^0 does not vanish, the momentum $\mathfrak{G}_i$ of the body will not be in the same direction as its velocity q_i. This was exemplified in the Lewis and Tolman lever. Although it was set in motion in the direction bc, the presence of stresses in the arm ab led to a momentum in that arm

37 Specifically, in the body's rest frame, the body can rotate infinitely slowly without application of any forces. By the principle of relativity, this same motion will be possible if the body is in uniform translational motion as well. However in this latter case the kinetic energy of the body would alter according to its orientation as it rotates. Since no forces were applied, this would violate the "energy principle," the law of conservation of energy. Notice that the rotation is infinitely slow, so that it does not contribute to the body's kinetic energy.

38 V^0 is the rest volume of the body and $i, j, k = 1, 2, 3$. I have simplified Laue's opaque notation by introducing an index notation, where Laue used round and square brackets to represent various products. For example, where I would write $q_i q_k p_{ik}^0$, he would write "$(q[qp^0])$."

directed transverse to the motion. If this momentum is added vectorially to the momentum of the inertial mass of the lever, the resultant total momentum vector will not be parallel to the direction of motion.

Laue was now in a position to restate the analysis given for the Lewis and Tolman lever in a way that would apply to general systems. This was the principle of burden of (Laue 1911a, §3 and §4). To begin, Laue introduced a new three dimensional stress tensor. In a body at rest, the time rate of change of momentum density $\partial \mathfrak{g}_i / \partial t$ is given by the negative divergence of the tensor p_{ik}:

$$\frac{\partial \mathfrak{g}_i}{\partial t} = -\frac{\partial p_{ik}}{\partial x_k}.$$

However if one wishes to investigate the time rate of change of momentum density in a moving body, one must replace the partial time derivative $\frac{\partial}{\partial t}$ with a total time derivative coordinated to the motion, $\frac{d}{dt} = \frac{\partial}{\partial t} + q_x \frac{\partial}{\partial x} + q_y \frac{\partial}{\partial y} + q_z \frac{\partial}{\partial z}$. Laue was able to show that the relevant time rate of change of momentum was given as the negative divergence of a new tensor t_{ik}

$$\frac{d \mathfrak{g}_i}{dt} = -\frac{\partial t_{ik}}{\partial x_k}.$$

where this "tensor of elastic stresses" was defined by

$$t_{ik} = p_{ik} + \mathfrak{g}_i q_k.$$

Note in particular that t_{ik} will not in general be symmetric since the momentum density $\mathfrak{g}_i$ will not in general be parallel to the velocity q_i.

The lack of symmetry of t_{ik} is a cause of momentary concern, for it is exactly the symmetry of p_{ik} that enables recovery, in effect, of the law of conservation of angular momentum. More precisely, the symmetry of the stress tensor is needed for the standard derivation of the result that the time rate of change of angular momentum of a body is equal to the total turning couple impressed on its surface. Laue proceeds to show, however, that this asymmetry does not threaten recovery of this law and is, in fact essential for it.[39] He writes the time rate of change of total angular momentum $\mathfrak{L}_i$ of a moving body as[40]

---

39 Laue calls §4, which contains this discussion, "the area law." I presume this is a reference to Kepler's second law of planetary motion, which amount to a statement of the conservation of angular momentum for planetary motion.

40 dV is a volume element of the body. I make no apology at this point for shielding the reader from Laue's notation, which has become more than opaque. Laue now uses square brackets to represent vector products, where earlier they represented an inner product of vector and tensor. ε_{ikl} is the fully antisymmetric Levi-Civita tensor, so that $\varepsilon_{ikl} A_k B_l$ is the vector product of two vectors A_k and B_l.

$$\frac{d\mathfrak{L}_i}{dt} = \int \varepsilon_{ikl}\mathfrak{r}_k\frac{d\mathfrak{g}_l}{dt} + \varepsilon_{ikl}q_k\mathfrak{g}_l dV = \int -\varepsilon_{ikl}\mathfrak{r}_k\frac{\partial t_{lm}}{\partial x_m} + \varepsilon_{ikl}q_k\mathfrak{g}_l dV. \tag{27}$$

Were the tensor t_{ik} symmetric, then the integration of the first term alone could be carried out using a version of Gauss' theorem. One could then arrive at the result that the total rate of change of momentum of the body is given by the turning couple applied to its surface. Thus if there is no applied couple, angular momentum would be conserved. However the tensor t_{ik} is not symmetric, and so an integral over the first term leaves a residual rate of change of angular momentum even when no turning couple is applied to the body. Fortunately there is a second term in the integrals of (27) that results from allowing for the use of the total time derivative. It is the vector product of the velocity q_i and momentum density $\mathfrak{g}_i$. This term would not be present in a classical analysis since these two vectors would then be parallel so that their vector product would vanish. In the relativistic context, this is not the case. This term corresponds exactly to the stress induced momentum in the arm ab of the Lewis and Tolman lever. This extra term exactly cancels the residual rate of change of angular momentum of the first term, restoring the desired result, the rate of change of angular momentum equals the externally applied turning couple.

9.3 Laue's "Complete Static Systems"

The last of the group of results developed in (Laue 1911a, §5) proved to be the most important for the longer term development of Nordström's theory of gravitation. Laue had shown clearly just how different the behavior of stressed and unstressed bodies in relativity theory could be. He now sought to delineate circumstances in which the presence of stresses within a body would not affect its overall dynamics. Such was the case of a "complete static system," which Laue defined as follows:

> We understand by this term such a system which is in static equilibrium in any justified reference system K^0, without sustaining an interaction with other bodies.

This definition is somewhat elusive and the corresponding definition in (Laue 1911b, 168–69) is similar but even briefer. In both cases, however, Laue immediately gave the same example of such a system, "an electrostatic field including all its charge carriers." This example and the definition leaves open the question of whether a body spinning at constant speed and not interacting with any other bodies is a complete static system. Such a body is in equilibrium and static in the sense that its properties are not changing with time, especially if the body spins around an axis of rotational symmetry. Tolman (1934, 81) gave a clearer definition:

> And in general we shall understand by a complete static system, an entire structure which can remain in a permanent state of rest with respect to a set of proper coordinates S^0 without the necessity for any forces from the outside.

He clearly understood this definition to rule out rotating bodies, for he noted a few lines later "the velocity of all parts of the system is zero in these coordinates $[S^0]$".

Tolman used this to justify the condition that the momentum density in the rest frame vanishes at every point

$$\mathfrak{g}^0 = 0. \tag{28}$$

Presumably Laue agreed for he also invoked this condition. From it, both Laue and Tolman derived the fundamental result characteristic of complete static systems:

$$\int p_{ik}^0 dV^0 = 0, \tag{29}$$

where the integral extends over the rest volume V^0 of the whole body. Laue allowed, in effect, that his conception of a complete static system could be relaxed without compromising the recovery of (29). For in a footnote (Laue 1911a, 540) to the example of an electrostatic field with its charge carriers, he noted that one could also consider the case of electrostatic-magnetostatic fields. Even though (28) failed to obtain for this case, the time derivative of $\mathfrak{g}^0$ did vanish which still allowed the derivation of (29).

This fundamental property (29) of complete static systems greatly simplified the expression (26) for the energy and momentum of a stressed body. Through (29) all the terms explicitly dependent on stresses vanish so that

$$\begin{aligned} E &= \frac{c}{\sqrt{c^2-q^2}}E^0, \\ \mathfrak{G}_i &= \frac{q_i}{c\sqrt{c^2-q^2}}E^0. \end{aligned} \tag{30}$$

As Laue pointed out, these expressions coincide precisely with those of a point mass with rest mass $m^0 = (E^0/c^2)$. Moreover under quasi-stationary acceleration—that is acceleration in which "the inner state $(E^0, \mathbf{p}^0)$ is not noticeably changed"—a complete static system will behave exactly like a point mass.

Laue could now offer a full resolution of the remaining problems described above in Section 9.1. An electron together with its field is a complete static system, he noted, no matter how it may be formed. As a result it will behave like a point mass, as long as its acceleration is quasi-stationary. In particular it will sustain inertial motion without the need for impressed forces. While Laue did not explicitly mention the problem of relating the electron's total field momentum to its inertial mass, Laue's result (30) resolves whatever difficulty might arise for the overall behavior of an electron. For however the electron may be constructed, as long as it forms a complete static system, equation (30) shows that the extraneous factor of 4/3 in equation (23) cannot appear. Finally, the Trouton-Noble condenser is a complete static system. While neither the momentum of its electromagnetic field or of its stressed mechanical structure will lie in the direction of its motion, equation (30) shows that the combined momentum $\mathfrak{G}_i$ will lie parallel to the velocity q_i, so that there is no net turning couple acting on the condenser.[41]

10. THE DEFINITION OF INERTIAL MASS IN NORDSTRÖM'S FIRST THEORY

What had emerged clearly from Laue's work was that the inertial properties of bodies could not be explained solely in terms of their rest masses and velocities, if the bodies were stressed. For Laue's equation (26) showed that the momentum of a moving body would be changed merely by the imposition of a stress, even though that stress need not deform the body or perform net work on it. Nordström clearly had results such as these in mind when he laid out the project of his (Nordström 1913a, 856–57). Laue and Herglotz, he reported, had constructed the entire mechanics of extended bodies without exploiting the concept of inertial mass. That concept, he continued, was neither necessary nor sufficient to represent the inertial properties of stressed matter. This now seems to overstate the difficulty, for Laue's entire system depended upon Einstein's result of the inertia of energy. Nonetheless nowhere did Laue's mechanics of stressed bodies provide a single quantity that represented *the* inertial mass of a stressed body.

It was to this last omission that Nordström planned to direct his paper. It was important, he urged, to develop a notion of the inertial mass of matter for the development of a gravitation theory. Such a theory must be based on the "unity of essence"[42] of inertia and gravity. He promised to treat the relativistic mechanics of deformable bodies in such a way that it would reveal a concept of inertial mass suitable for use in a theory of gravitation.

Nordström's analysis was embedded in a lengthy treatment of the mechanics of deformable bodies whose details will not be recapitulated here. Its basic supposition, however, was that the stress energy tensor $T_{\mu\nu}$ of a body with an arbitrary state of motion and stress would be given as the sum of two symmetric tensors (p. 858)

$$T_{\mu\nu} = p_{\mu\nu} + \nu\mathfrak{B}_\mu\mathfrak{B}_\nu. \tag{31}$$

The second tensor, $\nu\mathfrak{B}_\mu\mathfrak{B}_\nu$, he called the "material tensor." It represented the contribution to the total stress tensor from a matter distribution with rest mass density ν and four velocity $\mathfrak{B}_\mu$. The first tensor, $p_{\mu\nu}$, he called the "elastic stress tensor." It represented the stresses in the matter distribution. In the rest frame of the matter distribution, Nordström wrote the elastic stress tensor as (p. 863)

---

41 However this result did not end Laue's analysis of the Trouton-Noble experiment. See (Laue 1912).

42 *Wesenseinheit*. The term is sufficiently strong and idiosyncratic for it to be noteworthy that, so far as I know, Einstein was the only other figure from this period who used even a related term in connection with inertia and gravitation. In a paper cited earlier in (Nordström 1912, 1126), Einstein (1912d, 1063) had talked of the "equality of essence" (*Wesensgleichheit*) of inertial and gravitational mass. Einstein used the term again twice in later discussion. See (Norton 1985, 233).

$$\begin{bmatrix} p^0_{xx} & p^0_{xy} & p^0_{xz} & 0 \\ p^0_{yx} & p^0_{yy} & p^0_{yz} & 0 \\ p^0_{zx} & p^0_{zy} & p^0_{zz} & 0 \\ 0 & 0 & 0 & p^0_{uu}. \end{bmatrix}$$

The six zero-valued components in this matrix represent the momentum density and energy current due to the presence of stresses. They must vanish, Nordström pointed out, since stresses cannot be responsible for a momentum or energy current in the rest frame.[43] In particular, Nordström identified the component p^0_{uu} as a Lorentz invariant.[44]

In the crucial Section 4, "Definition of Inertial Mass," Nordström turned his attention to the (uu) component of equation (31) in the rest frame. This equation gave an expression for the Lorentz invariant rest energy density Ψ in terms of the sum of two invariant quantities[45]

$$\Psi = -p^0_{uu} + c^2\nu. \tag{32}$$

This equation gave simplest expression to the quantity of fundamental interest to Nordström's whole paper, the density ν, which would provide the source for the gravitational field equation. This density would be determined once Ψ and p^0_{uu} were fixed. However, while the rest energy density Ψ was a "defined quantity," it was not so clear how p^0_{uu} was to be determined. It represented an energy density associated with the stresses. Clearly if there were no stresses in the material, then this energy would have to be zero. But what if there were stresses?

To proceed Nordström considered a special case, a body in which there is an isotropic, normal pressure. In this case, Nordström continued, it is possible to fix the value of p^0_{uu} in such a way that the density ν can be determined. The elastic stress tensor could be generated out of a single scalar invariant, which I will write here as p, so that the elastic stress tensor in the rest frame is given by

---

43 In Section 5, Nordström augmented his analysis by considering the effect of heat conduction. This was represented by a third symmetric tensor, $w_{\mu\nu}$, whose *only* non-zero components in the rest frame were exactly these six components. Thus heat conduction was represented by an energy current and associated momentum density which did not arise from stresses and which had no associated energy density in the rest frame.

44 This followed easily from the fact that the tensor $p_{\mu\nu}$ twice contracted with the four velocity $\mathfrak{B}_\mu$ yields a Lorentz invariant, $p_{\mu\nu}\mathfrak{B}_\mu\mathfrak{B}_\nu$, which can be evaluated in the rest frame, where $\mathfrak{B}_\mu = (0, 0, 0, ic)$, and turns out to be $-c^2 p^0_{uu}$.

45 The presence of the negative sign follows from the use of a coordinate system in which the fourth coordinate is $x_4 = u = ict$. Therefore $T_{44} = T_{uu} = -$ (energy density).

$$\begin{bmatrix} p & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{bmatrix} \tag{33}$$

With this particular choice of stress tensor, Nordström pointed out, there is no momentum density associated with the stresses when the body is in motion. We can confirm this conclusion merely by inspecting the matrix (33). Since p is an invariant, the matrix will transform back into itself under Lorentz transformation. Therefore in all frames of reference, the six components (p_{14}, p_{24}, p_{34}) and (p_{41}, p_{42}, p_{43}) will remain zero. But these six components between them represent the momentum density and energy current due the stresses. Thus any momentum density present in the body will be due to the density ν.[46]

At this point, the reader might expect Nordström to recommend that one set p^0_{uu} in the general case in such a way that there are no momentum densities associated with stresses. Nordström informs us, however, that he could find no natural way of doing this. As a result, he urged that the "simplest and most expedient definition" lies in setting

$$c^2\nu = \Psi \tag{34}$$

so that $p^0_{uu} = 0$. To quell any concern that this choice had been made with undue haste, Nordström continued by asserting that the factual content of relativistic mechanics is unaffected by the choice of quantity that represents inertial mass. It is only when weight is assigned to inertial masses, such as in his gravitation theory, that the choice becomes important.

The reader who has followed the development of Nordström's argument up to this point cannot fail to be perplexed at the indirectness of what is the core of the entire paper! There are three problems. First, the choice of ν as given in (34) seems unchallengeable as the correct expression for the rest density of inertial mass. It merely sets this density equal to $1/c^2$ times the rest energy density—exactly as one would expect from Einstein's celebrated result of the inertia of energy. Indeed, any other division of total rest energy Ψ between the two terms of (32) would force us to say that ν does not represent the total inertial rest mass density, for there would be another part of the body's energy it does not embrace. Second, no argument is given for the claim that the choice of ν has no effect on the factual content of relativistic mechanics.[47] Finally, even if this second point is correct, it hardly seems worth much attention since the choice of expression for ν does significantly affect the factual content of gravitation theory.

The clue that explains the vagaries of Nordström's analysis lies in his citation of his own (Nordström 1911). There one finds an elaborate analysis of the relativistic

46 Unless, of course, heat conduction is present.

mechanics of the special case of a body with isotropic normal stresses—exactly the special case considered above. The analysis began by representing the stresses through a tensor of form (33). Nordström then showed the effect of arbitrarily resetting the value of p^0_{uu}. Reverting to the notation of (Nordström 1911), Nordström imagined that the p^0_{uu} term of (33) is replaced by some arbitrary U_0. He showed that the effect of this substitution is simply to replace the rest mass density γ (the analog of v in the 1913 paper) in the equations of the theory with an augmented

$$\gamma' = \gamma + \frac{p + U_0}{c^2}$$

without otherwise altering the theory's relations. He was able to conclude that setting $U_0 = p$ "is not a specialization of the theory, but only a specialization of concepts." In the introduction to (Nordström 1911), he had announced his plan to extend this analysis to the more general case with tangential stresses in another paper. Presumably the discussion of Section 4 in (Nordström 1913a) was intended to inform his readers that he was now unable to make good on his earlier plan. Indeed the remarks that seemed puzzling are merely a synopsis of some of the major points of (Nordström 1911). That the choice of (34) does not affect the factual content of relativistic mechanics is merely an extension of the result developed in detail in (Nordström 1911). It had become something of a moot point, however, in the context of gravitation theory.

To sum up, Nordström's choice of source density v was given by equation (34) and it was this result that gave meaning to the quantity v in the final sections of the paper in which his gravitation theory was recapitulated. We can give this quantity more transparent form by writing it in a manifestly covariant manner[48]

$$\nu = -\frac{1}{c^2} T_{\mu\nu} \mathfrak{B}_\mu \mathfrak{B}_\nu. \tag{35}$$

Natural as this choice seemed to Nordström, it was Einstein who shortly proclaimed that another term derived from the stress energy tensor was the only viable candidate and that this unique candidate led to disastrous results.

47 On reflection, however, I think the result not surprising. Barring special routes such as might be provided through gravitation theory, we have no independent access to the energy represented by the term p^0_{uu}. For example, in so far as this energy is able to generate inertial effects, such as through generation of a momentum density, it is only through its contribution to the sum $T_{uu} = -p^0_{uu} + c^2\nu$. The momentum density follows from the Lorentz transformation of the tensor $T_{\mu\nu}$. How we envisage the energy divided between the two terms of this sum will be immaterial to the final density yielded.

48 While $-c^2\nu$ is the trace of the material tensor $\nu\mathfrak{B}_\mu\mathfrak{B}_\nu$, this quantity $-c^2\nu$ is not the trace of the full tensor $T_{\mu\nu}$ as given in (31). This latter trace would contain terms in p^0_{xx}, etc.

11. EINSTEIN OBJECTS AGAIN

By early 1913, Einstein's work on his own gravitation theory had taken a dramatic turn. With his return to Zurich in August 1912, he had begun a collaboration with his old friend Marcel Grossmann. It culminated in the first sketch of his general theory of relativity, (Einstein and Grossmann 1913), the so-called "*Entwurf*" paper. This work furnished his colleagues all the essential elements of the completed theory of 1915, excepting generally covariant gravitational field equations.[49] While we now know that this work would soon be Einstein' s most celebrated achievement, the Einstein of 1913 could not count on such a jubilant reception for his new theory. He had already survived a bitter dispute with Abraham over the variability of the speed of light in his earlier theory of gravitation. And Einstein sensed that the lack of general covariance of his gravitational field equations was a serious defect of the theory which would attract justifiable criticism.

There was one aspect of the theory which dogged it for many years, its very great complexity compared with other gravitation theories. In particular, in representing gravitation by a metric tensor, Einstein had, in effect, decided to replace the single scalar potential of gravitation theories such as Newton's and Nordström's, with ten gravitational potentials, the components of the metric tensor. This concern was addressed squarely by Einstein in Section 7 of his part of the *Entwurf* paper. It was entitled "Can the gravitational field be reduced to a scalar?" Einstein believed he could answer this question decisively in the negative, thereby, of course, ruling out not just Nordström's theory of gravitation, but any relativistic gravitation theory which represented the gravitational field by a scalar potential.

Einstein's analysis revealed that he agreed with Nordström's assessment of the importance of Laue's work for gravitation theory. However he felt that Laue's work, in conjunction with the requirement of the equality of inertial and gravitational mass, pointed unambiguously to a different quantity as the gravitational source density. That was the trace of the stress-energy tensor. He proposed that a scalar theory would be based on the equation of motion for a point mass

$$\delta\left\{\int \Phi ds\right\} = 0 \tag{36}$$

where Φ is the gravitational potential, ds is the spacetime line element of special relativity and δ represents a variation of the mass' world line. He continued, tacitly comparing the scalar theory with his new *Entwurf* theory:

> Here also material processes of arbitrary kind are characterized by a stress-energy tensor $T_{\mu\nu}$. However in this approach a scalar determines the interaction between the gravitational field and material processes. This scalar, as Herr Laue has made me aware, can only be

49 For an account of this episode, see (Norton 1984).

$$\sum_{\mu} T_{\mu\mu} = P$$

> I want to call it "Laue's scalar." Then one can do justice to the law of the equivalence of inertial and gravitational mass here also up to a certain degree. That is, Herr Laue has pointed my attention to the fact that, for a closed system,
>
> $$\int P dV = \int T_{44} d\tau.$$
>
> From this one sees that the weight of a closed system is determined by its total energy according to this approach as well.

Recall that Einstein's version of the requirement of the equality of inertial and gravitational mass seeks to use the total energy of a system as a measure of it as a gravitational source. The selection of P, the trace of the stress energy tensor, does this for the special case of one of Laue's complete static systems. For such a system, the integral of the trace P over the spatial volume V of the system is equal to the negative value of the total energy of the system[50] $\int T_{44} dV$ since

$$\int P dV = \int T_{11} + T_{22} + T_{33} + T_{44} dV = \int T_{44} dV. \tag{37}$$

The three terms in T_{11}, T_{22} and T_{33} in the integral vanish because of the fundamental property (29) of complete static systems. Notice that Einstein can only say he does justice to the equality of inertial and gravitational mass "up to a certain degree," since this result is known to hold only for complete static systems and then only in their rest frames.

Einstein's wording indicates direct personal communication from Laue. concerning the stress-energy tensor and complete static systems. Such personal communication is entirely compatible with the fact that both Einstein and Laue were then in Zurich, with Einstein at the ETH and Laue at the University of Zurich. Below, in Section 15, I will argue that there is evidence that, prior to his move to Zurich, Einstein was unaware of the particular application of Laue's work discussed here by him.

Einstein continued his analysis by arguing that this choice of gravitational source density was disastrous. It leads to a violation of the law of conservation of energy. He wrote:

> The weight of a system that is not closed would depend however on the orthogonal stresses T_{11} etc. to which the system is subjected. From this there arise consequences which seem to me unacceptable as will be shown in the example of cavity radiation.[51]

---

50 Presumably Einstein's "$d\tau$" is a misprint and should read dV. In a coordinate system in which $x_4 = ict$, $T_{44} = -$(energy density), so that $\int T_{44} dV$ is the negative value of the total energy.

51 [JDN] One might well think that only Einstein could seriously ask after the gravitational mass of such an oddity in gravitation theory as radiation enclosed in a massless, mirrored chamber. Yet Planck (1908, 4) had already asked exactly this question.

> For radiation in a vacuum it is well known that the scalar P vanishes. If the radiation is enclosed in a massless, mirrored box, then its walls experience tensile stresses that cause the system, taken as a whole, to be accorded a gravitational mass $\int P d\tau$ which corresponds with the energy E of the radiation.
>
> Now instead of the radiation being enclosed in an empty box, I imagine it bounded
> 1. by the mirrored walls of a fixed shaft S,
> 2. by two vertically moveable, mirrored walls W_1 and W_2, which are firmly fixed to one another by a rod. (See Fig. 3.)
>
> In this case the gravitational mass $\int P d\tau$ of the moving system amounts to only a third part of the value which arises for a box moving as a whole. Therefore, in raising the radiation against a gravitational field, one would have to expend only a third part of the work as in the case considered before, in which the radiation is enclosed in a box. This seems unacceptable to me.

Einstein's objection bears a little expansion. He has devised two means of raising and lowering some fixed quantity of radiation in a gravitational field. Notice that in either case the radiation by itself has no gravitational mass, since the trace of the stress-energy tensor of pure electromagnetic radiation vanishes. What introduces such a mass is the fact that the radiation is held within an enclosure upon which it exerts a pressure, so that the enclosure is stressed. Even though the members of the enclosure are assumed massless, it turns out that a gravitational mass must still be ascribed to them simply because they are stressed. The beauty of Einstein's argument is that the gravitational masses ascribed in each of the two cases can be inferred essentially without calculation.

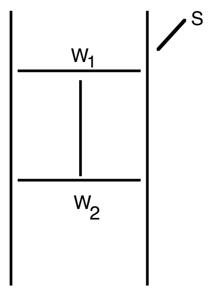


Figure 3: Rendering of Figure in Einstein's Text

In the first case, the radiation is moved in a mirrored box. The radiation and enclosing box form a complete static system. Therefore the gravitational mass of the box together with the radiation is given by the total energy of the radiation, $\int T_{44}^{em} dV$, in its rest frame, where I now write the stress-energy tensor of the electromagnetic radiation as $T_{\mu\nu}^{\text{em}}$. For ease of transition to the second case, it is convenient to imagine that each of the three pairs of opposing wall of the box is held in place *only* by a connecting rod, that the faces of the box are aligned with the x, y and z axes and that the disposition of the system is identical in these three directions. Each connecting rod will be stressed in reaction to the radiation pressure. I write $T^X_{\mu\nu}$ for the stress-energy tensor of the rod aligned in the x direction and the two stressed walls that this rod connects. $T^Y_{\mu\nu}$ and $T^Z_{\mu\nu}$ represent the other two corresponding systems. We can then infer directly from (37) that the gravitational mass of the entire system in its rest frame is proportional to

$$\int T^{\mathrm{em}}_{44} dV = \int T^{\mathrm{em}}_{\mu\mu} dV + \int T^X_{\mu\mu} dV + \int T^Y_{\mu\mu} dV + \int T^Z_{\mu\mu} dV = 3\int T^Z_{\mu\mu} dV.$$

The second equality follows from $T^{\mathrm{em}}_{\mu\mu} = 0$ and from the symmetry of the three axes, which entails

$$\int T^X_{\mu\mu} dV = \int T^Y_{\mu\mu} dV = \int T^Z_{\mu\mu} dV .$$

In Einstein's second case the radiation is trapped between sliding, mirrored baffles in a mirrored shaft aligned, let us say, in the z direction. The only component of the moveable system carrying a gravitation mass in this case will be the stressed rod and the stressed baffles it connects. Its gravitational mass in its rest frame will simply be proportional to the volume integral of the trace of its stress energy tensor. This integral is equal to $\int T^Z_{\mu\mu} dV$ and is one third of the corresponding integral for the first case, as Einstein claimed.

We now combine the two cases into a cycle. We lower the radiation inside the cube into the gravitational field, recovering some work, since the system has a gravitational mass. We then transfer the radiation into the baffle system and raise it. Only one third of the work released in the first step is needed to elevate the radiation because of the baffle system's reduced gravitational mass.[52] The mirrored cube and baffles are weightless once they are unstressed by the release of the radiation so they can be returned to their original positions. The cycle is complete with a net gain of energy.

That a theory should violate the conservation of energy is one of the most serious objections that Einstein could raise against it. Notice that he did not mention another possible objection that would derive directly from the vanishing of the trace of the stress-energy tensor of pure radiation. This vanishing entails that light cannot be deflected by a gravitational field. However, in 1913, prior to the experimental determination of this effect, Einstein could hardly have expected this last objection to have any force.

As devastating and spectacular as Einstein's objection was, he had at this time developed the unfortunate habit of advancing devastating arguments to prove conclusion he later wished to retract, see (Norton 1984, §5). This objection to all theories of

52 Since the estimates of gravitational mass are made in the system's rest frames, these motions would have to be carried out infinitely slowly.

gravitation with a scalar potential proved to be another instance of his habit. Within a few months Einstein had endorsed (if not initiated) a most interesting escape.

12. NORDSTRÖM'S SECOND THEORY

On July 24, 1913, Nordström submitted another version of his gravitation theory to *Annalen der Physik*, (Nordström 1913b). This version of the theory finally took proper notice of what Einstein had presented as obvious in his *Entwurf* paper. The only possible scalar that can represent gravitational source density is the trace of the stress-energy tensor and this choice, in conjunction with Laue's work on complete static systems, enables satisfaction of the requirement of equality of inertial and gravitational mass. Moreover the version of the requirement satisfied is an Einsteinian version in which the quantity of gravitational source is proportional to the total energy. This differs from the version embodied in Nordström's equation (16) in which the motion of a body in free fall is merely independent of its rest mass. Finally the theory offered an ingenious escape from Einstein's *Entwurf* objection. It turned out that the objection failed if one assumed that the proper length of a body would vary with the gravitational potential. This new version of this theory is sufficiently changed that it is now customarily known as Nordström's second theory.

There is room for interesting speculation on the circumstances under which Nordström came to modify his theory. In the introduction (p. 533) he thanked Laue and Einstein for identifying the correct gravitational source density. As we shall see, at two places in the paper (p. 544, 554), he also attributed arguments and results directly to Einstein without citation. Since I know of no place in which Einstein published these results, it seems a reasonable conjecture that Nordström learned of them either by correspondence or personal contact. That it was personal contact during a visit to Zurich at this time is strongly suggested by the penultimate line of the paper, which, in standard *Annalen der Physik* style, gives a place and date. It reads "Zurich, July 1913."[53] Since Nordström does thank both Laue and Einstein directly and in that order and since the wording of Einstein's *Entwurf* suggests a personal communication directly from Laue, we might conjecture also that there was similar direct contact between Laue and Nordström. Laue was also in Zurich at this time at the University of Zurich.

One cannot help but sense a somewhat sheepish tone in the introduction to (Nordström 1913b, 533), when he announced that this earlier presentation (Nordström 1913a) was "not completely unique" and that "the rest density of matter was defined in a fairly arbitrary way." In effect he was conceding that he had bungled the basic idea of his earlier (Nordström 1913a), that one had to take notice of the mechanics of stressed bodies in defining the gravitational source density and that this ought to be

53 An entry in Ehrenfest's Diary ("I", NeLR, Ehrenfest Archive, Scientific Correspondence, ENB: 4–15) reveals a visit to Zurich by Nordström in late June. See (CPAE 4, 294–301), "Einstein on Gravitation and Relativity: the Collaboration with Marcel Grossmann."

done in a way that preserved the equality of inertial and gravitational mass. Laue and Einstein were now telling him how he ought to have written that paper.

12.1 The Identification of the Gravitational Source

The first task of Nordström (1913b) was to incorporate the new source density into his theory and this was tackled in its first section. The final result would be to define this density in terms of what he called the "elastic-material tensor" $T_{\mu\nu}$, which corresponded to the sum of Nordström's (1913a) material tensor and elastic stress tensor as given above in (31). Following Einstein and Laue, he ended up selecting $1/c^2$ times the negative[54] trace D of $T_{\mu\nu}$ as his source density ν

$$\nu = \frac{1}{c^2}D = -\frac{1}{c^2}(T_{xx} + T_{yy} + T_{zz} + T_{uu}). \tag{38}$$

However unlike Laue and Einstein, that selection came at the conclusion of a fairly lengthy derivation. Nordström would show that the requirement of equality of inertial and gravitational mass in the case of a complete static system would force this choice of source density.

To begin, Nordström chose essentially the same field equation for the potential Φ and gravitational force density equations as in (Nordström 1913a):

$$\frac{\partial^2\Phi}{\partial x^2} + \frac{\partial^2\Phi}{\partial y^2} + \frac{\partial^2\Phi}{\partial z^2} + \frac{\partial^2\Phi}{\partial u^2} = g(\Phi)\nu \ , \tag{12''}$$

$$\mathfrak{K}^g_\mu = -g(\Phi)\nu\frac{\partial\Phi}{\partial x_\mu} \ . \tag{18'}$$

Here ν remained the as yet undetermined gravitational source density. The important innovation was that the gravitation factor g was now allowed to vary as a function of the potential Φ. In an attempt to bring some continuity to the development of (Nordström 1913b) from (1913a), he recalled that in the former paper (p. 873, 878) he had foreshadowed the possibility that g might be a function of the inner constitution of bodies. Indeed the paper had closed with the speculation that such a dependence might enable the molecular motions of a falling body to influence its acceleration of fall, presumably as part of a possible escape from Einstein's original objection to his theory.

As it happened, however, the Φ dependence of g was introduced for an entirely different purpose in (Nordström 1913b). Einstein's version of the equality of inertial and gravitational mass required that the *total* energy of a system would be the measure of its gravitational source strength. This total energy would include the energy of the gravitational field itself. This requirement, familiar to us from Einstein's treat-

54 Since he retained his standard coordinate system of $x, y, z, u = ict$, the negative sign is needed to preserve the positive sign of ν.

ment of general relativity, leads to non-linearity of field equations. In Nordström's theory, this non-linearity would be expressed as a Φ dependence of g.

Nordström now turned to complete static systems, whose properties would yield not just the relation (38) but also a quite specific function for $g(\Phi)$. At this point Nordström seemed able to give Laue a little taste of his own medicine. As I pointed out above, Laue's 1911 definition of a complete static system had excluded such systems as bodies rotating uniformly about their axis of symmetry. Nordström now made the obvious extension, defining what he called a "complete *stationary*" system. Curiously he made no explicit statement that his was a more general concept. Readers simply had to guess that his replacement of Laue's "static" by his "stationary" was no accident. Or perhaps they had to wait until Laue's (1917, 273) own concession that Nordström was first to point out the extension.

Nordström's complete stationary system had the following defining characteristics: it was a system of finite bodies for which a "justified"[55] reference system existed in which the gravitational field was static, that is, $\partial\Phi/\partial t = 0$. In particular in the relevant reference system, instead of Laue' s condition (28) which required the vanishing everywhere of the momentum density $\mathfrak{g}$, Nordström required merely that the *total* momentum $\mathfrak{G}$ vanished,

$$\mathfrak{G} = \int \mathfrak{g} dv = 0$$

where v is the volume of the body in its rest frame. The two illustrations Nordström gave—surely not coincidentally—were exactly two systems that Laue's earlier definition did not admit: a body rotating about its axis of symmetry and a fluid in stationary flow. Of course the first example was one of great importance to Nordström. It was precisely the example discussed in the final paragraphs of both (Nordström 1912 and 1913a). That such a system would fall more slowly than a non-rotating system was the substance of Einstein's original objection. Now able to apply the machinery of Laue's complete static systems to this example, Nordström could try to show that these rotating bodies did not fall slower in the new theory.

Nordström proceeded to identify the three stress-energy tensors which could contribute to the total energy of a complete stationary system. They were the "elastic-material tensor" $T_{\mu\nu}$, mentioned above; the "electromagnetic tensor" $L_{\mu\nu}$, which we would otherwise know as the stress-energy tensor of the electromagnetic field; and finally the "gravitation tensor" $G_{\mu\nu}$. This last tensor was the stress-energy tensor of the gravitational field itself. It had been identified routinely in (Nordström (1912, 1128; 1913a, 875). It was given by[56]

$$G_{\mu\nu} = \frac{\partial\Phi}{\partial x_\mu}\frac{\partial\Phi}{\partial x_\nu} - \frac{1}{2}\delta_{\mu\nu}\frac{\partial\Phi}{\partial x_\lambda}\frac{\partial\Phi}{\partial x_\lambda} \tag{39}$$

55 In this context I read this to mean "inertial".

where Φ is the gravitational potential. Invoking Laue's basic result (29), which was also used by Einstein for the same end, Nordström could represent the total energy E_0 of a complete stationary system in its rest frame as the integral over all space of the sum of the traces of these three tensors

$$E_0 = -\int T + G + L \, dv \,.$$

However we have $L = 0$ and have written $T = -D$. Finally the integral of the trace G could be written in greatly simplified fashion if one assumed special properties for the complete stationary system. In particular, the gravitational potential Φ must approach the limiting constant value Φ_a at spatial infinity. From this and an application of Gauss' theorem, Nordström inferred that[57]

$$\int G \, dv = \int (\Phi - \Phi_a) g(\Phi) \nu \, dv \,. \tag{40}$$

Combining and noting that the inertial rest mass m of the system is E_0/c^2, we have

$$m = \frac{E_0}{c^2} = \frac{1}{c^2}\int \{D - (\Phi - \Phi_a) g(\Phi)\nu\} dv \,.$$

However we also have the total gravitational mass M_g of the system is

$$M_g = \int g(\Phi) \nu dv \,. \tag{41}$$

Nordström was now finally able to invoke what he calls "Einstein's law of equivalence," the equality of inertial and gravitational mass. Presumably viewing the completely stationary system from a great distance, one sees that it is a system with inertial mass m lying within a potential Φ_a so that we must be able to write

---

56 I continue to compress Nordström's notation. He did not use the Kronecker delta $\delta_{\mu\nu}$ and wrote individual expressions for G_{xx}, G_{xy}, etc. The derivation of (39) is brief and entirely standard. Writing ∂_μ for $\partial/\partial x_\mu$ we have, from substituting (12'') into (18')

$$\begin{aligned}\mathfrak{K}^g_\mu &= -g(\Phi)\nu\partial_\mu\Phi = -(\partial_\nu\partial_\nu\Phi)\partial_\mu\Phi = -\left(\partial_\nu(\partial_\mu\Phi\partial_\nu\Phi) - \frac{1}{2}\partial_\mu(\partial_\lambda\Phi\partial_\lambda\Phi)\right) \\ &= -\partial_\nu\left(\partial_\mu\Phi\partial_\nu\Phi - \frac{1}{2}\delta_{\mu\nu}(\partial_\lambda\Phi\partial_\lambda\Phi)\right) = -\partial_\nu G_{\mu\nu} \,.\end{aligned}$$

57 Nordström's derivation of (40) seems to require the additional assumption that ν is non-zero only in some finite part of space. For a completely stationary system, the field equation reduces to the Newtonian $\partial_i\partial_i\Phi = g(\Phi)\nu$. If ν satisfies this additional assumption one can recover from it Nordström's result (his equation (5)), that $|\nabla\Phi| = |\partial_i\Phi| = \frac{1}{4\pi r^2}\int g(\Phi)\nu dv$, at large distances r from the system. This result seems to be needed to complete the derivation of (40).

$$M_g = g(\Phi_a)m \ .$$

These last three equations form an integral equation which can only be satisfied identically if

$$D = g(\Phi)\nu\left\{\Phi - \Phi_a + \frac{c^2}{g(\Phi_a)}\right\} .$$

Finally Nordström required that this relation between D and ν be independent of Φ_a, from which the two major result of the analysis followed: first, $g(\Phi)$ is given by

$$g(\Phi) = \frac{c^2}{A + \Phi} \tag{42}$$

where A is a universal constant; second, he recovered the anticipated identity of the source density ν

$$\nu = \frac{1}{c^2}D = -\frac{1}{c^2}(T_{xx} + T_{yy} + T_{zz} + T_{uu}) \ . \tag{38}$$

The constant A in equation (42) is taken by Nordström to be an arbitrary additive gauge factor, corresponding to the freedom in Newtonian gravitation theory of setting an arbitrary zero point for the gravitational potential. However, in contrast with the Newtonian case, there is a natural gauge of $A = 0$ in which the equations are greatly simplified. Writing the potential that corresponds to the choice of $A = 0$ as Φ', the expression for $g(\Phi')$ is

$$g(\Phi') = \frac{c^2}{\Phi'} \tag{43}$$

and in this gauge one recovers a beautifully simple expression for the relationship between the total rest energy E_0, inertial rest mass m and gravitational mass M_g of a completely stationary system

$$E_0 = mc^2 = M_g\Phi'_a \ . \tag{44}$$

In particular, it contains exactly the Newtonian result that the energy of a system with gravitational mass M_g in a gravitational field with potential Φ'_a is $M_g\Phi'_a$. This was an improvement over Nordström's first theory were the closest corresponding result was (15), a dependence of mass on an exponential function of the potential.

12.2 Dependence of Lengths and Times on the Gravitational Potential

Satisfactory as these results were, Nordström had not yet answered the objection of Einstein's *Entwurf* paper. Indeed it is nowhere directly mentioned in Nordström's paper, although Nordström (1913b, 544) does cite the relevant part of the *Entwurf* paper to acknowledge Einstein's priority concerning the expression for the gravitational source density in equation (38). However at least the second, third and fourth parts of Nordström (1913b) are devoted implicitly to escaping from Einstein's objection. Nordström there developed a model of a spherical electron within his theory and showed that its size must vary inversely with the gravitational potential Φ'. It only becomes apparent towards the middle of Section 3 that this result must hold for the dimensions of all bodies and that this general result provides the escape from Einstein's objection. The general result is demonstrated by an argument attributed without any citation to Einstein. Since I know of no place where this argument was published by Einstein and since we know that Nordström was visiting Einstein at the time of submission of his paper, it is a reasonable supposition that he had the argument directly from Einstein in person. Since the general result appears only in the context of this argument, it is a plausible conjecture that the result, as well as its proof, is due to Einstein. Of course the successful recourse to an unusual kinematical effect of this type is almost uniquely characteristic of Einstein's work.

Einstein's argument takes the *Entwurf* objection and reduces it to its barest essentials. The violation of energy conservation inferred there depended solely on the behavior of the massless members of the systems that were oriented transverse to the direction of the field in which they moved. The transverse members lowered were stressed and thus were endowed with a gravitational mass so that work was recovered. The transverse members elevated were unstressed so that they had no gravitational mass and no work was required to elevate them. The outcome was a net gain in energy. Einstein's new argument considered this effect in a greatly simplified physical system. Instead of using radiation pressure to stress the transverse members, Einstein now just imagined a single transverse member—a non-deformable rod—tensioned between vertical rails. The gravitational mass of the rod is increased by the presence of these stresses. Since the rod expands on falling into the gravitational field, the rails must diverge and work must be done by the forces that maintain the tension in the rod. It turns out that this work done exactly matches the work released by the fall of the extra gravitational mass of the rod due to its stressed state. The outcome is that no net work is released. In recounting Einstein's argument, Nordström makes no mention of Einstein's *Entwurf* objection. This is puzzling. It is hard to imagine that he wished to avoid publicly correcting Einstein when the history of the whole theory had been a fruitful sequence of objections and correction and this correction was endorsed and even possibly invented by Einstein. In any case it should have been clear to a contemporary reader who understood the mechanism of Einstein's objection that the objection would be blocked by an analogous expansion of the systems as they fell into the gravitational field. Certainly Einstein (1913, 1253), a few months later, reported that his *Entwurf* objection failed because of the tacit assumption of the constancy of dimensions of the systems as they move to regions of different potential.[58]

Einstein's argument actually establishes that the requirement of energy conservation for such cycles necessitates a presumed isotropic expansion of linear dimensions to be in inverse proportion to the gravitational potential Φ'. Nordström's report of the argument reads:

> Herr Einstein has proved that the dependence in the theory developed here of the length dimensions of a body on the gravitational potential must be a general property of matter. He has shown that otherwise it would be possible to construct an apparatus with which one could pump energy out of the gravitational field. In Einstein's example one considers a non-deformable rod that can be tensioned moveably between two vertical rails. One could let the rod fall stressed, then relax it and raise it again. The rod has a greater weight when stressed than unstressed, and therefore it would provide greater work than would be consumed in raising the unstressed rod. However because of the lengthening of the rod in falling, the rails must diverge and the excess work in falling will be consumed again as the work of the tensioning forces on the ends of the rod.
>
> Let S be the total stress (stress times cross-sectional area) of the rod and l its length. Because of the stress, the gravitational mass of the rod is increased by[59]
>
> $$\frac{g(\Phi)}{c^2}Sl = \frac{1}{\Phi'}Sl\ .$$
>
> In falling [an infinitesimal distance in which the potential changes by $d\Phi'$ and the length of the rod by dl], this gravitational mass provides the extra work
>
> $$-\frac{1}{\Phi'}Sld\Phi'$$
>
> However at the same time at the ends of the rod the work is lost [to forces stressing the rod]. Setting equal these two expressions provides
>
> $$-\frac{1}{\Phi'}d\Phi' = \frac{1}{l}dl$$
>
> which yields on integration
>
> $$l\Phi' = \text{const.} \qquad [(45)]$$
>
> This, however, corresponds with [Nordström's] equation (25a) [the potential dependence of the radius of the electron].[60]

This result (45) was just one of a series of dependencies of basic physical quantities on gravitational potential. In preparation in Section 2 for his analysis of the electron,

---

58 He gives no further analysis. It is clear, however, that as the mirrored box is lowered into the field and expands, work would be done by the radiation pressure on its walls. It will be clear from the ensuing analysis that this work would reduce the total energy of the radiation by exactly the work released by the lowering of the gravitational mass of the system. Thus when the radiation is elevated in the mirrored shaft the radiation energy recovered would be diminished by exactly the amount needed to preserve conservation of energy in the entire cycle.

59 [JDN] To see this, align the x axis of the rest frame with the rod. The only non-vanishing component of $T_{\mu\nu}$ is T_{xx} which is the stress (per unit area) in the rod. Therefore, from (38) and (41), the gravitational mass

$$M_g = \int g(\Phi)\nu dv = -\frac{g(\Phi)}{c^2}Sl\ .$$

Nordström had already demonstrated that the inertial mass m of a complete stationary system varied in proportion to the external gravitational potential Φ':

$$\frac{m}{\Phi'} = \text{const.} \tag{46}$$

whereas the gravitational mass of the system M_g was independent of Φ'. The proof considered the special case of a complete stationary system for which the external field Φ'_a would be uniform over some sphere at sufficient distance from the system and directed perpendicular to the sphere. We might note that a complete stationary system of finite size within a constant potential field would have this property. He then imagined that the external field Φ'_a is altered by a slow displacement of yet more distant masses. He could read directly from the expression for the stress energy tensor of the gravitational field what was the resulting energy flow through the sphere enclosing the system and from this infer the alteration in total energy and therefore mass of the system. The result (46) followed immediately and from it the constancy of M_g through (44).[61] Calling on (45), (46) and other specific results in his analysis of the electron, Nordström was able to infer dependencies on gravitational potential for the stress S in the electron's surface, the gravitational source density v and the stress tensor p_{ab}:

$$\frac{S}{\Phi'^3} = \text{const.}\ \frac{\nu}{\Phi'^4} = \text{const.}\ \frac{P_{ab}}{\Phi'^4} = \text{const.}$$

Finally, after these results had been established, the entire Section 5 was given to establishing that the time of a process T would depend on the potential Φ' according to

$$T\Phi' = \text{const.} \tag{47}$$

In particular it followed from this result that the wave lengths of spectral lines depends on the gravitational potential. Nordström reported that the wavelength of light from the Sun's surface would be increased by one part in two million. He continued to report that "The same—possibly even observable—shift is given by several other new theories of gravitation." (p. 549) While he gave no citation, modern readers need hardly be told that all of Einstein's theories of gravitation from this period give this effect, including his 1907–1912 scalar theory of static fields, his *Entwurf* form of

60 A footnote here considers a complication that need not concern us. It reads:
"If the rod is deformable, in stressing it, some work will be expended and the rest energy of the rod will be correspondingly increased. Thereby the weight also experiences an increase, which provides the added work dA in falling. However, since in falling the rest energy diminishes, the work recovered in relaxing the rod is smaller than that consumed at the stressing and the difference amounts to exactly dA."

61 On pp. 545–46, he showed that, for a complete stationary system, the external potential Φ'_a varied in direct proportion to the potential Φ' at some arbitrary point within the system, so that Φ' and Φ'_a could be used interchangeably in expressing the proportionalities of the form of (45), (46) and (47).

general relativity and the final 1915 generally covariant version of the theory. The very numerical value that Nordström reported—one part in two million—was first reported in Einstein's earliest publication on gravitation, (Einstein 1907a, 459).

Nordström devoted some effort to the proof of (47). He noted that it followed immediately from (45) and the constancy of the velocity of light for the time taken by a light signal traversing a a rod. Anxious to show that it held for other systems, he considered a small mass orbiting another larger mass M_1 in a circular orbit of radius r within an external potential field Φ'_a. The analysis proved very simple since the speed of the small mass, its inertial mass and the potential along its trajectory were all constant with time. He showed that its orbital period T satisfied

$$\frac{M_1 c^2}{4\pi r \Phi'_a} = \frac{4\pi^2 r^2}{T^2}. \tag{48}$$

As the potential Φ'_a varies, $r\Phi'_a$ remains constant according to (45). Therefore the equality requires that T vary in direct proportion to r from which it follows that T satisfies (47). Again he showed the same effect for the period of a simple harmonic oscillator of small amplitude.

12.3 Applications of Nordström's Second Theory: The Spherical Electron and Free Fall

So far we have seen the content of Nordström's second theory and how he established its coherence. The paper also contained two interesting applications of the theory. The first was an analysis of a spherical electron given in his Section 3. It turned out to yield an especially pretty illustration of the result that a gravitational mass is associated with a stress. For the *entire* gravitational mass of Nordström's electron proves to be due to an internal stress. The electron was modelled as a massless shell carrying a charge distributed on its surface. (See Appendix for details.) The shell must be stressed to prevent mechanical disintegration of the electron due to repulsive forces between parts of the charge distribution. The electric field does not contribute to the electron's gravitational mass since the trace of its stress-energy tensor vanishes. Since the shell itself is massless it also does not contribute to the gravitational mass when unstressed. However, when stressed in reaction to the repulsive electric forces, it acquires a gravitational mass which comprises the entire gravitational mass of the electron.

Nordström's model of the electron was not self contained in the sense that it required only known theories of electricity and gravitation. Like other theories of the electron at this time, it had to posit that the stability of the electron depended on the presence of a stress bearing shell whose properties were largely unknown. While one might hope that the attractive forces of gravitation would replace this stabilizing shell, that was not the case in Nordström's electron. Rather he was superimposing the effects of gravitation on a standard model of the electron.[62]

The second illustration was an analysis in his final Section 7 of the motion in free fall of a complete stationary system. In particular, Nordström was concerned to determine just how close this motion was to the corresponding motion of a point mass. The results were not entirely satisfactory. He was able to show that complete stationary systems fell like point masses only for the case of a homogeneous external gravitational field, that is, one whose potential was a linear function of all four coordinates (x, y, z, u). He showed that a complete stationary system of mass m, falling with four velocity $\mathfrak{B}_\mu$ in a homogeneous external field Φ_a, obeys equations of motion

$$-g(\Phi'_a)m\frac{\partial\Phi_a}{\partial x_\mu} = \frac{d}{d\tau}m\mathfrak{B}_\mu \tag{49}$$

which corresponded to the equations (13) for a point mass in his first theory, excepting the added factor $g(\Phi'_a)$. Allowing that the mass m varies inversely with Φ'_a through equation (46), it follows that explicit mention of the mass m can be eliminated from these equations of motion which become[63]

$$-c^2\frac{\partial}{\partial x_\mu}\ln\Phi'_a = \frac{d}{d\tau}\mathfrak{B}_\mu - \mathfrak{B}_\mu\frac{d}{d\tau}\ln\Phi_a .$$

Nordström's concern was clearly still Einstein's original objection to this first theory recounted above in Section 7. A body rotating about its axis of symmetry could form a complete stationary system. He could now conclude that such a body would fall exactly as if it had no rotation, contrary, as he noted, to the result of his earlier theory. Also, he concluded without further discussion that molecular motions would have no influence on free fall. However, the vertical acceleration of free fall would continue to be slowed by its initial velocity according to (19) of his first theory.

We might observe that stresses would play a key role in the cases of the rotating body and the kinetic gas. The rotating body would be stressed to balance centrifugal forces and the walls containing a kinetic gas of molecules would be stressed by the forces of the gas pressure. These stresses would add to the gravitational mass of the spinning body and the contained gas allowing them to fall independently of their internal motions. No such compensating stresses would be present in the case of a point mass or a complete stationary system projected horizontally, so they would fall slower due to their horizontal velocity.

Thus, while Nordström's theory finally satisfied the requirement of equality of inertial and gravitational mass in Einstein's sense, it still did not satisfy the requirement that all bodies fall alike in a gravitational field. This Einstein (1911, §1) called "Galileo's principle," elsewhere (1913, 1251) citing it as the fact of experience supporting the equality of inertial and gravitational mass. Galileo's principle held only under rather restricted conditions: the system must be in vertical fall and in a homo-

---

62 We can see the Nordström had no real hope of eliminating this shell with gravitational attraction. For the electric field by itself generates no gravitational field in his theory. Another element must be present in the structure of the electron if gravitational forces are to arise.

63 I have corrected Nordström's incorrect "+" to "–" on the right-hand side.

geneous field.[64] At least, however, he could report that Einstein had extended the result to systems that were not complete stationary systems. He had shown that the *average* acceleration of an elastically oscillating system accorded with (19). Since this last result is nowhere reported in Einstein's publications, we must assume that he had it directly from Einstein.

Finally, in the course of his exposition, Nordström could note that the mass dependence on Φ' of relation (46) now replaces the corresponding condition (15) of his first theory. The new variable factor of $g(\Phi') = c^2/\Phi'$ in the in the equation of motion (49) causes (14) to be replaced by $\frac{mc^2}{\Phi'}\frac{d\Phi'}{d\tau} = c^2\frac{dm}{d\tau}$, which integrates to yield (46).

13. EINSTEIN FINALLY APPROVES: THE VIENNA LECTURE OF SEPTEMBER 1913

In September 1913, Einstein attended the 85th Congress of the German Natural Scientists and Physicians. There he spoke on the subject of the current state of the problem of gravitation, giving a presentation of his new *Entwurf* theory and engaging in fairly sharp dispute in discussion. A text for this lecture with ensuing discussion was published in the December issue of *Physikalische Zeitschrift* (Einstein 1913). Einstein made clear (p. 1250) his preference for Nordström's theory over other gravitation theories, including Abraham's and Mie's. Nordström's latest version of his gravitation theory was the only competitor to Einstein's own new *Entwurf* theory satisfying four requirements that could be asked of such gravitation theories:

1. "Satisfaction of the conservation law of momentum and energy;"
2. "Equality of inertial and gravitational masses of closed systems;"
3. Reduction to special relativity as a limiting case;
4. Independence of observable natural laws from the absolute value of the gravitational potential.

What Einstein did not say was that the satisfaction of 1. and 2. by Nordström's theory was due in significant measure to Einstein's pressure on Nordström and Einstein's own suggestions.

Einstein devoted a sizeable part of his lecture to Nordström's theory, giving a self-contained exposition of it in his Section 3. That exposition was a beautiful illustration of Einstein's ability to reduce the complex to its barest essentials and beyond. He simplified Nordström's development in many ways, most notably:

64 One might think that this would give Einstein grounds for rejecting the competing Nordström theory in favor of his own *Entwurf* theory. At the appropriate place, however, Einstein (1913, 1254) did *not* attack Nordström on these grounds. Perhaps that was for the better since it eventually turned out that general relativity fared no better. In general relativity, for example, a rotating body falls differently, in general, from a non-rotating body. See (Papapetrou 1951; Corinaldesi and Papapetrou 1951).

- Einstein selected the natural gauge (43) for the potential Φ', writing the resulting potential without the prime as φ.
- Einstein eradicated the implicit potential dependence of the mass m in (46), using a new mass m which did not vary with potential. This meant that Einstein's m coincided with the gravitational mass, not the inertial mass of a body.

To begin, Einstein used as the starting point the "Hamiltonian" equation of motion (36) which he had first recommended in Section 7 of his *Entwurf* paper. Using coordinates $(x_1, x_2, x_3, x_4) = (x, y, z, ict)$, he wrote this equation of motion of a mass m as

$$\delta\left\{\int H dt\right\} = 0 \tag{50}$$

where

$$H = -m\varphi\frac{d\tau}{dt} = m\varphi\sqrt{c^2 - q^2} .$$

Here q is the coordinate three-speed and τ is the Minkowski interval given by

$$d\tau^2 = -dx^2 - dy^2 - dz^2 + c^2dt^2 . \tag{51}$$

Since Einstein varied the three spatial coordinates of the particle trajectory x_i, the resulting equation of motion governed the three-velocity $\dot{x}_i = \frac{dx_i}{dt}$

$$\frac{d}{dt}\left\{m\varphi\frac{\dot{x}_i}{\sqrt{c^2 - q^2}}\right\} - m\frac{\partial\varphi}{\partial x_i}\sqrt{c^2 - q^2} = 0 .$$

It also followed that the momentum (increased by a multiplicative factor c) I_i and the conserved energy E were given by

$$I_i = m\varphi\frac{\dot{x}_i}{\sqrt{c^2 - q^2}} , \quad E = m\varphi\frac{c^2}{\sqrt{c^2 - q^2}} .$$

In particular, one could read directly from these formulae that the inertial mass of a body of mass m at rest is given by $m\varphi/c^2$ and that its energy is $m\varphi$.

Einstein then introduced the notion that had rescued Nordström's theory from his own recent attack: directly measured lengths and times might not coincide with those given by the Minkowski line element (51). He called the former quantities "natural" and indicated them with a subscript 0. He called the latter "coordinate" quantities. The magnitude of the effect was represented by a factor ω which would be a function of φ and was defined by

$$d\tau_0 = \omega d\tau . \tag{52}$$

Allowing for the dependence of energy on φ and the effects of the factor ω, Einstein developed an expression for the stress-energy tensor $T_{\mu\nu}$ of "flowing, incoherent matter"—we would now say "pressureless dust"—in terms of its natural mass density ρ_0 and the corresponding gravitational force density k_μ:

$$T_{\mu\nu} = \rho_0 c\varphi\omega^3 \frac{dx_\mu}{dt}\frac{dx_\nu}{dt}, \quad k_\mu = -\rho_0 c\varphi\omega^3 \frac{\partial\varphi}{\partial x_\mu}.$$

The two quantities were related by the familiar conservation law

$$\frac{\partial T_{\mu\nu}}{\partial x_\nu} = k_\mu.$$

The next task was to re-express this conservation law in terms of the trace $T_{\sigma\sigma}$ of the stress-energy tensor. Mentioning Laue's work, Einstein remarked that this quantity was the only choice for the quantity measuring the gravitational source density. For the special case of incoherent matter, $T_{\sigma\sigma} = -\rho_0 c\varphi\omega^3$, so that the conservation law took on a form independent of the special quantities involved in the case of incoherent matter flow

$$\frac{\partial T_{\mu\nu}}{\partial x_\nu} = T_{\sigma\sigma}\frac{1}{\varphi}\frac{\partial\varphi}{\partial x_\mu}, \tag{53}$$

Einstein announced what was really an assumption: this form of the law governed arbitrary types of matter as well.

This general form of the conservation law allowed Einstein to display the satisfaction by the theory of the second requirement he had listed. That was the equality of inertial and gravitational masses of closed systems. His purpose in including the additional words "closed systems" now became clear. In effect he meant by them Laue's complete static systems. His demonstration of the satisfaction of this result was admirably brief but damnably imprecise, compared to the careful attention Nordström had lavished on the same point. Einstein simply assumed that he had a system over whose spatial extension there was little variation in the φ term $\frac{1}{\varphi}\frac{\partial\varphi}{\partial x_\mu} = \frac{\partial \log\varphi}{\partial x_\mu}$ on the right-hand side of (53). An integration of (53) over the spatial volume v of such a system revealed that the four-force acting on the body is

$$\frac{\partial\log\varphi}{\partial x_\mu}\int T_{\sigma\sigma}dv = \frac{\partial\log\varphi}{\partial x_\mu}\int T_{44}dv,$$

where the terms in T_{11}, T_{22}, and T_{33} were eliminated by Laue's basic result (29). Since $\int T_{44}dv$ is the negative of the total energy of the system, Einstein felt justified to conclude: "Thereby is proven that the weight of a closed system is determined by its total measure [of energy]." Einstein's readers might well doubt this conclusion and suspect that the case of constant φ considered was a special case that may not be representative of the general case. Fortunately such readers could consult (Nordström 1913b) for a more precise treatment.

In his lecture, Einstein was seeking to give an exposition of both Nordström's and his new theory of gravitation and reasons for deciding between them. Thus we might anticipate that he had to cut corners somewhere. And that place turned out to be the singular novelty of Nordström's theory in 1913, the potential dependence of lengths and times. His introduction of this effect and concomitant retraction of his *Entwurf* objection was so brief that only someone who had followed the story closely and read the report of Einstein's argument in (Nordström 1913b) could follow it. Virtually all he had to say lay in a short paragraph (p. 1253):

> Further, equation [(53)] allows us to determine the function [ω] of φ left undetermined from the physical assumption that no work can be gained from a static gravitational field through a cyclical process. In §7 of my jointly published work on gravitation with Herr Grossmann I generated a contradiction between the scalar theory and the fundamental law mentioned. But I was there proceeding from the tacit assumption that ω = const[ant]. The contradiction is resolved, however, as is easy to show, if one sets[65]
>
> $$l = \frac{l_0}{\omega} = \frac{\text{const.}}{\varphi}$$
>
> or
>
> $$\omega = \text{const.} \cdot \varphi\,. \qquad [(54)]$$
>
> We will give yet a second substantiation for this stipulation later.

That second substantiation followed shortly, immediately after Einstein had given the field equation of Nordström's theory. He considered two clocks. The first was a "light clock," a rod of length l_0 with mirrors at either end and a light signal propagating in a vacuum and reflected between them. The second was a "gravitation clock," two gravitationally bound masses orbiting about one another at constant distance l_0. He gave no explicit analysis of these clocks. His only remark on their behavior was that their relative speed is independent of the absolute value of the gravitational potential, in accord with the fourth of the requirements he had laid out earlier for gravitation theories. This, he concluded, "is an indirect confirmation of the expression for ω given in equation [(54)]."

Einstein's readers would have had to fill in quite a few details here. Clearly the dependence of l_0 on the potential would cause the period of the light clock also to vary according to (47). But readers would also need to know of the analysis of the gravitation clock given by (Nordström 1913b) which led to (48) above and the same dependence on potential for the clock's period. Thus the dependence of both periods is the same so that the relative rate of the two clocks remains the same as the external potential changes. Had this result been otherwise, the fourth requirement would have been violated. That it was not presumably displays the coherence of the theory and thereby provides the "*indirect* confirmation." Curiously Einstein seems not to be

65 l is defined earlier as the length of a body. This retraction is also mentioned more briefly (p. 261) in the addendum to the later printing of (Einstein and Grossmann 1913) in the *Zeitschrift*.

making the obvious point that his equations (52) and (54) together yield the same potential dependence for periodic processes as follows from the behavior of these two clocks—or perhaps he deemed that point too obvious to mention.

The final component of the theory was its field equation. Recalling that "Laue's scalar" $T_{\sigma\sigma}$ must enter into this equation, Einstein simply announced it to be

$$-\kappa T_{\sigma\sigma} = \varphi \frac{\partial^2}{\partial x_\tau^2} \varphi \ . \tag{55}$$

It became apparent that the additional factor of φ on the right-hand side was included to ensure compatibility with the conservation of energy and momentum.[66] To display this compatibility he noted that stress-energy tensor $t_{\mu\nu}$ of the gravitational field is

$$t_{\mu\nu} = \frac{1}{\kappa}\left\{ \frac{\partial\varphi}{\partial x_\mu}\frac{\partial\varphi}{\partial x_\nu} - \frac{1}{2}\delta_{\mu\nu}\left(\frac{\partial\varphi}{\partial x_\tau}\right)^2 \right\} .$$

This tensor satisfies the equalities

$$T_{\sigma\sigma}\frac{1}{\varphi}\frac{\partial\varphi}{\partial x_\mu} = -\frac{1}{\kappa}\frac{\partial\varphi}{\partial x_\mu}\frac{\partial^2\varphi}{\partial x_\tau^2} = -\frac{\partial t_{\mu\nu}}{\partial x_\nu}$$

The first depends on substitution of $T_{\sigma\sigma}$ by the field equation and the second holds identically. Substituting into the conservation law (53) yields an expression for the joint conservation of gravitational and non-gravitational energy momentum,[67]

$$\frac{\partial}{\partial x_\nu}(T_{\mu\nu} + t_{\mu\nu}) = 0 \ .$$

All that remained for Einstein was to give his reasons for not accepting Nordström's theory. In our time, of course, the theory is deemed an empirical failure because it does not predict any deflection of a light ray by a gravitational field and does not explain the anomalous motion of Mercury. However in late 1913, there had been no celebrated eclipse expeditions and Einstein's own *Entwurf* theory also did not explain the anomalous motion of Mercury. Thus Einstein's sole objection to the theory was not decisive, although we should not underestimate its importance to Einstein.

---

66 Although Einstein does not make this point, it is helpful to divide both sides by φ and look upon $T_{\sigma\sigma}/\varphi$ as the gravitational source density. The trace $T_{\sigma\sigma}$ represents the mass-energy density and division by φ cancels out this density's φ dependence to return the gravitational mass density.

67 As Michel Janssen has repeatedly emphasized to me, Einstein's analysis is a minor variant of the method he described and used to generate the field equations of his *Entwurf* theory. Had Einstein begun with the identity mentioned, the expression for $t_{\mu\nu}$ and the conservation law (53), a reversal of the steps of Einstein's argument would generate the field equation. For further discussion of Einstein's method, see (Norton 1995).

According to Nordström's theory, the inertia of a body with mass m was $m\varphi/c^2$. Therefore, as the gravitational field in the neighborhood of the body was intensified by, for example, bringing other masses closer, the inertia of the body would actually decrease. This was incompatible with Einstein's idea of the "relativity of inertia" according to which the inertia of a body was caused by the remaining bodies of the universe, the precursor of what he later called "Mach's Principle." This deficiency enabled Einstein to ask after the possibility of extending the principle of relativity to accelerated motion, to see the real significance of the equality of inertial and gravitational mass in his principle of equivalence (which was not satisfied by Nordström's theory) and to develop his *Entwurf* theory.

14. EINSTEIN AND FOKKER: GRAVITATION IN NORDSTRÖM'S THEORY AS SPACETIME CURVATURE

It was clear by the time of Einstein's Vienna lecture that Nordström's most conservative of approaches to gravitation had led to a something more than a conservative Lorentz covariant theory of gravitation, for it had become a theory with kinematical effects very similar to those of Einstein's general theory of relativity. Gravitational fields would slow clocks and alter the lengths of rods. All that remained was the task of showing just how close Nordström's theory had come to Einstein's theory. This task was carried out by Einstein in collaboration with a student of Lorentz', Adriaan D. Fokker, who visited Einstein in Zurich in the winter semester of 1913–1914 (Pais 1982, 487). Their joint (Einstein and Fokker 1914), submitted on February 19, 1914, was devoted to establishing essentially one result, namely, in modern language, Nordström's theory was actually the theory of a spacetime that was only conformal to a Minkowski spacetime with the gravitational potential the conformal factor, so that the presence of a gravitational field coincided with deviations of the spacetime from flatness. That, of course, was not how Einstein and Fokker described the result. Their purpose, as they explained in the title and introduction of the paper, was to apply the new mathematical methods of Einstein's *Entwurf* theory to Nordström's theory. These methods were the "absolute differential calculus" of Ricci and Levi-Civita (1901). They enabled a dramatic simplification of Nordström's theory. It will be convenient here to summarize the content of the theory from this new perspective as residing in three basic assumptions:

I. Spacetime admits preferred coordinate systems $(x_1, x_2, x_3, x_4) = (x, y, z, ct)$ in which the spacetime interval is given by

$$ds^2 = \Phi^2(dx^2 + dy^2 + dz^2 - c^2dt^2) \tag{56}$$

and in which the trajectory of point masses in free fall is given by

$$\delta\int ds = 0 .$$

That such a characterization of the spacetime of Nordström's theory is possible is implicit in Einstein's Vienna lecture. In fact, once one knows the proportionality of ω and φ, the characterization can be read without calculation from Einstein's expression (52) for the natural proper time and the equation of motion (50). Einstein and Fokker emphasized that the preferred coordinate systems are ones in which the postulate of the constancy of the velocity of light obtains. For, along a light beam $ds^2 = 0$, so that

$$dx^2 + dy^2 + dz^2 - c^2 dt^2 = 0 \ .$$

We see here in simplest form the failure of the theory to yield a deflection of a light beam in a gravitational field. This failure is already evident, of course, from the fact that a light beam has no gravitational mass since the trace of its stress-energy tensor vanishes.

II. The conservation of gravitational and non-gravitational energy momentum is given by the requirement of the vanishing of the covariant divergence of the stress-energy tensor $T_{\mu\nu}$ for non-gravitational matter. At this time, Einstein preferred to write this condition as[68]

$$\sum_{\nu} \frac{\partial \mathfrak{T}_{\sigma\nu}}{\partial x_{\nu}} = \frac{1}{2} \sum_{\mu\nu\tau} \frac{\partial g_{\mu\nu}}{\partial x_{\sigma}} \gamma_{\mu\tau} \mathfrak{T}_{\tau\nu} \ ,$$

since they could interpret the term on the right-hand side as representing the gravitational force density.

Noting, as Einstein and Fokker did on pp. 322–23, that the $T_{\mu\nu}$ of the Vienna lecture corresponds to the tensor density $\mathfrak{T}_{\mu\nu}$ of the new development, they evaluated this conservation law in the preferred coordinate systems of I. It yielded the form of the conservation law (53) of the Vienna lecture.

Finally Einstein and Fokker turned to the field equation which was to have the form

$$\Gamma = k\mathfrak{T}$$

where κ is a constant. The quantity $\mathfrak{T}$ had to be a scalar representing material processes. In the light of the earlier discussion, we know there was only one viable choice, the trace of the stress-energy tensor $T = \frac{1}{\sqrt{-g}} \sum_{\tau} \mathfrak{T}_{\tau\tau}$. For the quantity Γ, which must be constructed from the metric tensor and its derivatives, they reported that the researches of mathematicians allowed only one quantity to be considered, the

68 Here Einstein had not yet begun to use modern notational conventions. Summation over repeated indices is not implied. All indices are written as subscript so that $\gamma_{\mu\nu}$ is the fully contravariant form of the metric, which we would now write as $g^{\mu\nu}$. $\mathfrak{T}_{\mu\nu}$ is the mixed tensor density which we could now write as $\sqrt{-g}T^{\nu}_{\mu}$.

full contraction of the Riemann-Christoffel tensor (ik, lm) of the fourth rank, where they allowed i, k, l and m to vary over 1, 2, 3 and 4. This assumed that the second derivative of $g_{\mu\nu}$ enters linearly into the equation. Therefore we have:

III. The gravitational field satisfies the field equation which asserts the proportionality of the fully contracted Riemann-Christoffel tensor and the trace of the stress energy tensor

$$\sum_{iklm} \gamma_{ik}\gamma_{lm}(ik, lm) = \kappa \frac{1}{\sqrt{-g}} \sum_{\tau} \mathfrak{T}_{\tau\tau} .$$

Evaluation of this field equation in the preferred coordinate systems of I. yields the field equation (55) of the Vienna lecture.

Einstein and Fokker were clearly and justifiably very pleased at the ease with which the methods of the *Entwurf* theory had allowed generation of Nordström's theory. In the paper's introduction they had promised to show that (p. 321)

> ... one arrives at Nordström's theory instead of the Einstein-Grossmann theory if one makes the single assumption that it is possible to choose preferred reference systems in such a way that the principle of the constancy of the velocity of light obtains.

Their concluding remarks shine with the glow of their success when they boast that (p. 328)

> ... one can arrive at Nordström's theory from the foundation of the principle of the constancy of the velocity of light through purely formal considerations, i.e. without assistance of further physical hypotheses. Therefore it seems to us that this theory earns preference over all other gravitation theories that retain this principle. From the physical stand point, this is all the more the case, as this theory achieves strict satisfaction of the equality of inertial and gravitational mass.

Of course Einstein retained his objection that Nordström's theory violates the requirement of the relativity of inertia.[69] The new formulation gives us vivid demonstration of this failure: the disposition of the preferred coordinate systems of I. will be entirely unaffected by the distribution of matter in spacetime. Einstein must then surely have been unaware that it would prove possible to give a generally covariant formulation of Nordström's theory on the basis of Weyl's work (Weyl 1918). The requirement that the preferred coordinate systems of I. exist could be replaced by the generally covariant requirement of the vanishing of the conformal curvature tensor. This formal trick, however, does not alter the theory's violation of the relativity of inertia and the presence of preferred coordinate systems in it.

---

69 As we know from lecture notes taken by a student, Walter Dallenbach, (EA 4 008, 41-42), Einstein in his teaching at the ETH in Zurich at this time included the claim that one arrives at the Nordström theory merely by assuming there are specialized coordinate system in which the speed of light is constant.There he remarks that this theory violates the relativity of inertia.

There remained a great irony in Einstein and Fokker's paper, which their readers would discover within two short years. While the existence of preferred coordinate systems was held against the Nordström theory, Einstein's own *Entwurf* theory was not itself generally covariant and would not be until November 1915, when Einstein would disclose the modern field equations to the Prussian Academy. Einstein and Grossmann (1913) had settled upon gravitational field equations which were not generally covariant. We now know that the generally covariant field equations of the completed general theory of relativity can be derived by means of the Riemann-Christoffel tensor through an argument very similar to the one used to arrive at the generally covariant form of the field equation of the Nordström theory. Einstein and Grossmann had considered and rejected this possibility in §4.2 of Grossmann's part of their joint paper. The obvious ease with which consideration of the Riemann-Christoffel tensor led to the field equation of Nordström's theory clearly gave Einstein an occasion to rethink that rejection. For Einstein and Fokker's paper concluded with the tantalizing remark that the reasons given in Grossmann's §4 of their joint paper against such a connection did not withstand further examination. Whatever doubt this raised in Einstein's mind seem to have subsided by March 1914, at which time he reported in a letter to this confidant Michele Besso that the "general theory of invariants functioned only as a hindrance" in construction of his system (Speziali 1972, 53).

Thus the conservative path struck by Nordström and Einstein led not just to the connection between gravitation and spacetime curvature but to the first successful field equation which set an expression in the Riemann-Christoffel curvature tensor proportional to one in the stress-energy tensor of matter.

15. WHAT EINSTEIN KNEW IN 1912

Einstein and Fokker's characterization in 1914 of the Nordström theory gives us a convenient vantage point from which to view Einstein's theory of 1912 for static gravitational fields. In particular we can see clearly that this theory already contained many of the components that would be assembled to form Nordström's theory. Indeed we shall see that Einstein's theory came very close to Nordström's theory. However we shall also see that a vital component was missing—the use of the stress-energy tensor and Laue's work on complete static systems. This component enables a scalar Lorentz covariant theory of gravitation to satisfy some version of the requirement of the equality of inertial and gravitational mass. We must already suspect that Einstein was unaware of this possibility prior to his August 1912 move to Zurich for his July 1912 response to Abraham (Einstein 1912d), quoted in Section 4 above, purports to show that no Lorentz covariant theory of gravitation could satisfy this requirement.

Einstein (1912a, 1912b) was the fullest development of a relativistic theory of static gravitational fields based on the principle of equivalence and in which the gravitational potential was the speed of light c. By Einstein's own account the following year (Einstein and Grossmann 1913, I, §1, §2), the theory was actually a theory of a spacetime with the line element

$$ds^2 = -dx^2 - dy^2 - dz^2 + c^2dt^2 \tag{57}$$

where c is now a function of x, y and z and behaves as a gravitational potential. Einstein (1912a, 360) offered the field equation

$$\nabla^2 c = kc\sigma \tag{58}$$

where k is a constant and σ the rest density of matter.[70] What Einstein did not mention in his *Entwurf* reformulation of the 1912 theory was that this field equation corresponded to the generally covariant field equation

$$R = \frac{k}{2}T$$

where R is the fully contracted Riemann-Christoffel tensor and T the trace of the stress–energy tensor, in the case of an unstressed, static matter distribution. This is exactly the field equation of Nordström's theory!

This field equation (58) had an extremely short life, for in (Einstein 1912b, §4), a paper submitted to *Annalen der Physik* on March 23, 1912, just a month after February 26, when he had submitted (Einstein 1912a), he revealed the disaster that had befallen his theory and would lead him to retract this field equation. Within the theory the force density $\mathfrak{F}$ on a matter distribution σ at rest is

$$\mathfrak{F} = -\sigma \operatorname{grad} c\ .$$

Einstein conjoined this innocuous result with the field equation (58) and applied it to a system of masses at rest held together in a rigid massless frame within a space in which c approached a constant value at spatial infinity. He concluded that the total gravitational force on the frame

$$\int \mathfrak{F} d\tau = -\int \sigma \operatorname{grad} c\, d\tau = -\frac{1}{k}\int \frac{\nabla^2 c}{c} \operatorname{grad} c\, d\tau$$

in general does not vanish. That is, the resultant of the gravitational forces exerted by the bodies on one another does not vanish. Therefore the system will set itself into motion, a violation of the equality of action and reaction, as Einstein pointed out. In effect the difficulty lay in the theory's failure to admit a gravitational field stress tensor, for the gravitational force density $\mathfrak{F}$ is equal to the divergence of this tensor. Were the tensor to be definable in Einstein's theory, that fact alone, through a standard application of Gauss' theorem, would make the net resultant force on the system vanish.[71]

70 The factor of c on the right-hand side of this otherwise entirely classical equation is introduced in order to leave c undetermined by a multiplicative gauge factor rather than an additive one.

Einstein then proceeded to consider a number of escapes from this disaster. The second and third escapes involved modifications to the force law and the field equation. The former failed but the latter proved workable. Einstein augmented the source density σ of (58) with a term in c:

$$\nabla^2 c = k\left\{c\sigma + \frac{1}{2k}\frac{\mathrm{grad}^2 c}{c}\right\}.$$

The extra term was constructed to allow the formation of a gravitational field stress tensor and the conclusion that there would be no net force on the system of masses. Einstein was especially pleased to find that this extra term proved to represent the gravitational field energy density so that the source term of the field equation was now the total energy density of the system, gravitational and non-gravitational.[72]

For our purposes what is most interesting is the first escape that Einstein considered and rejected. Mentioning vaguely "results of the old theory of relativity," he considered the possibility that the stressed frame of the system might have a gravitational mass. That possibility was dismissed however with an argument that is surprising to those familiar with his work of the following year: that possibility would violate the equality of inertial and gravitational mass! Einstein considered a box with mirrored walls containing radiation of energy E. He concluded from his theory that, if the box were sufficiently small, the radiation would exert a net force on the walls of the box of $-E\,\mathrm{grad}c$. He continued (Einstein 1912b, 453):

> This sum of forces must be equal to the resultant of forces which the gravitational field exerts on the whole system (box together with radiation), if the box is massless and if the circumstance that the box walls are subject to stresses as a result of the radiation pressure does not have the consequence that the gravitational field acts on the box walls. Were the

---

71 Writing $t_{im} = \frac{1}{kc}\left(\partial_i c\,\partial_m c - \frac{1}{2}\delta_{im}(\partial_n c\,\partial_n c)\right)$ for the quantity that comes closest to the stress tensor, we have the following in place of the standard derivation of the stress tensor (analogous to the derivation of (39)). Substituting field equation (58) into the expression for $\mathfrak{F}$, we recover:

$$\mathfrak{F}_i = -\sigma\partial_i c = -\frac{1}{kc}\partial_m c\,\partial_m c\,\partial_i c = -\partial_m t_{im} - \frac{1}{2k}\frac{\partial_m c\,\partial_m c}{c^2}\partial_i c\ .$$

The first term of the final sum is a divergence which would vanish by Gauss' theorem when integrated over the space containing the masses of the frame, leaving no net force. The problem comes from the second term, which is present only because of the factor of c on the source side of the field equation (58). In this integration it will not vanish in general, leaving the residual force on the masses. The need to eliminate this second term also dictates the precise form of the modification to the field equation that Einstein ultimately adopted. When the field equation source σ was augmented to become $\sigma + \frac{1}{2k}\frac{\partial_m c\,\partial_m c}{c^2}$, this second term no longer arose in the above expression for $\mathfrak{F}_i$.

72 However Einstein was disturbed to find that the new field equation only allowed his principle of equivalence to apply to infinitesimally small parts of space. See (Norton 1985, §4.2, §4.3).

> latter the case, then the resultant of the forces exerted by the gravitational field on the box (together with its contents) would be different from the value $-E\,\mathrm{grad}\,c$, *i.e.* the gravitational mass of the system would be different from E.

Einstein could not have written this were he aware of the relevant properties of "Laue's scalar" T. As Einstein himself showed the following year, the use of T as the gravitational source density in exactly this example of radiation enclosed in a mirrored cavity allowed one to infer *both* that the walls of the cavity acquired a gravitational mass because of their stressed state and that the gravitational mass of the entire system was given by its total energy. We must then take Einstein at his word and conclude that he learned of these properties of T from Laue. Presumably this means after his move to Zurich in August 1912 where Laue also was, and after completion of his work on his scalar theory of static gravitational fields in 1912.

Had Einstein been aware of these results earlier in 1912, they would probably not have pleased him in the long run. To begin, he did believe at the time of writing the *Entwurf* paper that the selection of T as the gravitational source density in a scalar theory of gravitation led to a contradiction with the conservation of energy. Had he seen past this to its resolution in the gravitational potential dependence of lengths he would have arrived at a most remarkable outcome: his theory of 1912 would have become exactly Nordström's final theory! As we saw above, his first field equation of 1912 was already equivalent to Nordström's final field equation in covariant terms. His equation of motion for a mass point was already the geodesic equation for a spacetime with the line element (57). This line element already entailed a dependence of times on the gravitational potential. The consistent use of Laue's scalar T as a source density would finally have led to a similar dependence for spatial length so that the line element (57) would be replaced by Nordström's (56). Since the expressed purpose of Einstein's 1912 theory was to extend the principle of relativity, this out come would not have been a happy one for Einstein. For his path would have led him to a theory which entailed the existence of coordinate systems in which the speed of light was globally constant. That is, the theory had resurrected the special coordinate systems of special relativity.

16. THE FALL OF NORDSTRÖM'S THEORY OF GRAVITATION

Revealing as Einstein and Fokker's formulation of the theory had been, Nordström himself clearly did not see it as figuring in the future development of his theory. Rather, Nordström embedded his 1913 formulation of his gravitation theory in his rather short lived attempts to generate a unified theory of electricity and gravitation within a five dimensional spacetime (Nordström 1914c, 1914d, 1915). Other work on the theory in this period was devoted to developing a clearer picture of the behavior of bodies in free fall and planetary motion according to the theory. Behacker (1913) had computed this behavior for Nordström's first theory and (Nordström 1914a) performed the same service for his second theory. In both cases the behavior demanded by the theories was judged to be in complete agreement with experience.

Nordström also had to defend his theory from an attack by Gustav Mie. Mie had made painfully clear in the discussion following Einstein's Vienna lecture of 1913 (published in *Physikalische Zeitschrift*, *14*, 1262–66) that he was outraged over Einstein's failure even to discuss Mie's own theory of gravitation in the lecture. Einstein explained that this omission derived from the failure of Mie's theory to satisfy the requirement of the equality of inertial and gravitational mass. Mie counterattacked with a two part assault (Mie 1914) on Einstein's theory. In an appendix (§10) Mie turned his fire upon Nordström's theory, claiming that it violated the principle of energy conservation. Nordström's (1914b) response was that Mie had erroneously inferred the contradiction within Nordström's theory by improperly importing a result from Mie's own theory into the derivation. Laue (1917, 310–13) pointed to errors on both sides of this dispute.

However it was not Mie's theory that led to the demise of Nordström's theory. Rather it was the rising fortunes of Einstein's general theory of relativity. Einstein completed the theory in a series of papers submitted to the Prussian Academy in November 1915. Within a few years, with the success of Eddington's eclipse expedition, Einstein had become a celebrity and his theory of gravitation eclipsed all others. One of the papers from that November 1915 (Einstein 1915) reported the bewitching success of the new theory in explaining the anomalous motion of Mercury. This success set new standards of empirical adequacy for gravitation theories. Prior to this paper, the pronouncements of a gravitation theory on the minutia of planetary orbits were not deemed the ultimate test of a new theory of gravitation. Einstein's own *Entwurf* theory failed to account for the anomalous motion of Mercury. Yet this failure is not mentioned in Einstein's publications from this period and one cannot even tell from these publications whether he was then aware of it. Thus the treatment in (Nordström 1914a) of the empirical adequacy of his theory to observed planetary motions was entirely appropriate by the standards of 1914. He showed that his theory predicted a very slow retardation of the major axis of a planet's elliptical orbit. Computing this effect for the Earth's motion he found it to be 0.0065 seconds of arc per year, which could be dismissed as "very small in relation to the astronomical perturbations [due to other planets]" (p.1109) Thus he could proceed to the overall conclusion (p. 1109) that

> ... the laws derived for [free] fall and planetary motion are in the *best* agreement with experience [my emphasis]

Standards had changed so much by the time of Laue's (1917) review article on the Nordström theory that even motions much smaller than the planetary perturbations were decisive in the evaluation of a gravitation theory. Einstein's celebrated 43 seconds of arc per century advance of Mercury's perihelion is less than a tenth of the perihelion motion due to perturbations from the other planets. Laue (p. 305) derived a formula for the predicted retardation—not advance—of a planet's perihelion. Without even bothering to substitute values into the formula he lamented

> Therefore the perihelion moves opposite to the sense of rotation of the orbit. In the case of Mercury, the impossibility of explaining its perihelion motion with this calculation lies already in this difference of sign concerning the perihelion motion.

Through this period, Nordström's theory had its sympathizers and the most notable of these was Laue himself.[73] He clearly retained this sympathy when he wrote the lengthy review article, (Laue 1917). Einstein's theory had become so influential by this time that Laue introduced the review with over four pages of discussion of Einstein's theory (pp. 266–70). That discussion conceded that Einstein's theory had attracted the most adherents of any relativistic gravitation theory. It also contained almost two pages of continuous and direct quotation from Einstein himself, as well as discussion of the epistemological and empirical foundations of Einstein' s theory. His discussion was not the most up-to-date, for he reported Einstein's *Entwurf* 0.84 seconds of arc deflection for a ray of starlight grazing the Sun, rather than the figure of 1.7 of the final theory of 1915. All this drove to the conclusion that there were no decisive grounds for accepting Einstein's theory and provided Laue with the opportunity to review a gravitation theory based on special relativity, Nordström's theory, which he felt had received less attention than it deserved.

The fall of Nordström's theory was complete by 1921. By this time even Laue had defected. In that year he published a second volume on general relativity to accompany his text on special relativity (Laue 1921). On p.17, he gave a kind appraisal of the virtues and vices of his old love, Nordström's theory. However he was firm in his concluding the superiority of Einstein's theory because of the failure of Nordström's theory to yield any gravitational light deflection—a defect, he urged, that must trouble any Lorentz covariant gravitation theory. Laue never lost his affection for the theory and years later took the occasion of Einstein's 70th birthday to recall the virtues of Nordström's theory (Laue 1949). The theory's obituary appeared in Pauli's encyclopedic distillation of all that was worth knowing in relativity theory (Pauli 1921, 144). He pronounced authoritatively

> The theory solves in a logically quite unexceptionable way the problem sketched out above, of how to bring the Poisson equation and the equation of motion of a particle into a Lorentz-covariant form. Also, the energy-momentum law and the theorem of the equality of inertial and gravitational mass are satisfied. If, in spite of this, Nordström's theory is not acceptable, this is due, in the first place, to the fact that it does not satisfy the principle of *general* relativity (or at least not in a simple and natural way ...). Secondly, it is in contradiction with experiment: it does not predict the bending of light rays and gives the displacement of the perihelion of Mercury with the wrong sign. (It is in agreement with Einstein's theory with regard to the red shift.)

He thereby rehearsed generations of physicists to come in the received view of Nordström's theory and relieved them of the need to investigate its content any further.

73 In a letter of October 10, 1915, to Wien, Mie had identified Laue as an adherent of Nordström's theory, explaining it through Laue's supposed failure to read anything else! I am grateful to John Stachel for this information.

17. CONCLUSION

The advent of the general theory of relativity was so entirely the work of just one person— Albert Einstein—that we cannot but wonder how long it would have taken without him for the connection between gravitation and spacetime curvature to be discovered. What would have happened if there were no Einstein? Few doubt that a theory much like special relativity would have emerged one way or another from the researches of Lorentz, Poincaré and others. But where would the problem of relativizing gravitation have led? The saga told here shows how even the most conservative approach to relativizing gravitation theory still did lead out of Minkowski spacetime to connect gravitation to a curved spacetime. Unfortunately we still cannot know if this conclusion would have been drawn rapidly without Einstein's contribution. For what led Nordström to the gravitational field dependence of lengths and times was a very Einsteinian insistence on just the right version of the equality of inertial and gravitational mass. Unceasingly in Nordström's ear was the persistent and uncompromising voice of Einstein himself demanding that Nordström see the most distant consequences of his own theory.

APPENDIX: NORDSTRÖM'S MODEL OF THE ELECTRON

Nordström's (1913b) development of his second theory contains (§3) a model of the electron which accounts for the effect of gravitation. The electron is modelled as a massless spherical shell of radius a carrying charge e distributed uniformly over its surface.[74] Three types of matter are present: an electric charge and its field; the shell stressed to balance the repulsive electric forces between different parts of the charge distribution; and the gravitational field generated by all three types of matter. See Fig. 4. Taking each in turn, we have

74 "Rational" units of charge are used, which means, in effect, that the electrostatic field equation is $\Delta\Psi = -\rho$, for charge density ρ.

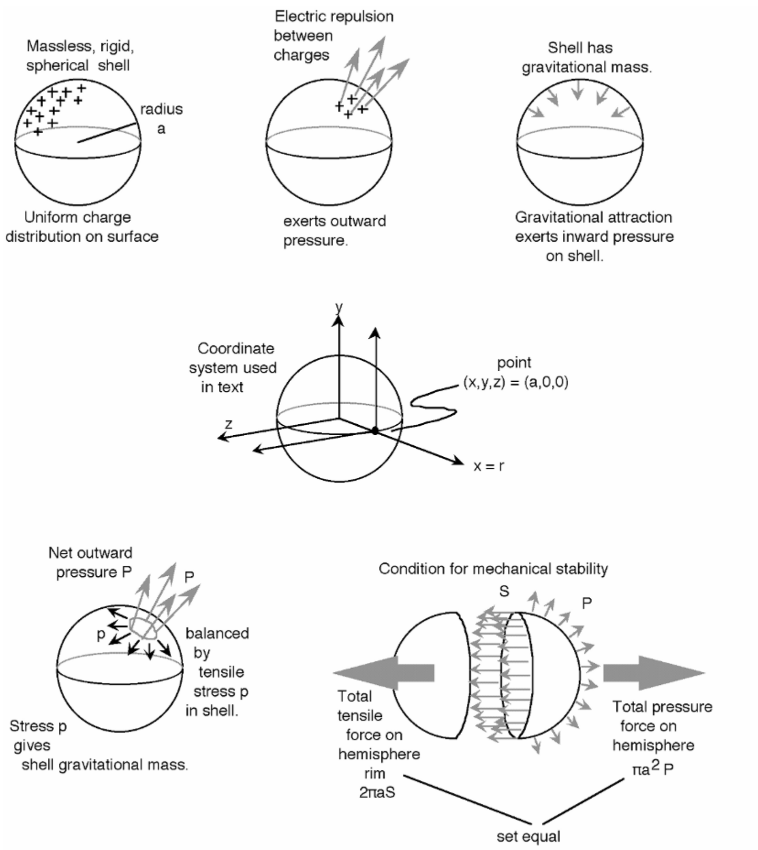


Figure 4: Nordström's Model of the Electron

Electric Charge and its Field

Using familiar results of electrostatics in the rational system of units, the electric charge e generates an electric potential Ψ at radius r from the center of the shell, for the case of $r \geq a$, which satisfies

$$\Psi = \frac{e}{4\pi r}, \quad \frac{\partial \Psi}{\partial r} = -\frac{e}{4\pi r^2}.$$

The latter value is all that is required to compute the Maxwell stress tensor at an arbitrary point on the shell which is representative of all its points due the rotational symmetry of the shell. We choose convenient coordinates $(x_1, x_2, x_3) = (x, y, z)$ for this point. We set the origin at the center of the shell, align the x axis with a radial arm and consider a point on the surface of the shell at which $(x, y, z) = (a, 0, 0)$. Writing ∂_i for $(\partial/\partial x_i)$, we have that the Maxwell stress tensor is[75]

$$L_{ik} = -\left(\partial_i \Psi \partial_k \Psi - \frac{1}{2}\delta_{ik}(\delta_m \Psi \partial_m \Psi)\right)$$

$$= -\begin{bmatrix} \frac{e^2}{32\pi^2 r^4} & 0 & 0 \\ 0 & -\frac{e^2}{32\pi^2 r^4} & 0 \\ 0 & 0 & -\frac{e^2}{32\pi^2 r^4} \end{bmatrix}.$$

We read directly from the coefficients of this tensor that the charges of the shell (at position $r = a$) are subject to an outwardly directed pressure of magnitude $e^2/32\pi a^4$ which seeks to cause the shell to explode radially outwards. That is, these charges are subject to a net electric force density given by the negative divergence of this stress tensor, $\partial_k L_{ik}$. With $r = a$, this force density is of magnitude $4e^2/32\pi^2 a^5$ directed radially outward.

Gravitational Field

The stresses in the shell will generate a gravitational field. For the moment, we shall write the total gravitational mass as M_g and note that it must be distributed uniformly over the shell. Since the source gravitational mass is all located in the shell over which the gravitational potential is constant, the field equation and stress tensor of the gravitational field reduce to the analogous equations of electrostatics, excepting a sign change. Thus the gravitational potential for $r \geq a$ satisfies

$$\Phi = -\frac{M_g}{4\pi r} + \Phi_a, \quad \frac{\partial \Phi}{\partial r} = \frac{M_g}{4\pi r^2},$$

where Φ_a is the external gravitational potential. Choosing the same point and coordinate system as in the analysis of the electric field, we find that the gravitational field stress tensor, as given by the spatial parts of the gravitational stress-energy tensor (39) is

75 The nonstandard minus sign follows the convention Nordström used in paper of requiring that (force density) = — (divergence of stress tensor). See (Nordström 1913b, 535, eq. 7).

$$G_{ik} = \begin{bmatrix} \frac{M_g^2}{32\pi^2 r^4} & 0 & 0 \\ 0 & -\frac{M_g^2}{32\pi^2 r^4} & 0 \\ 0 & 0 & -\frac{M_g^2}{32\pi^2 r^4} \end{bmatrix}.$$

This reveals an inwardly directed pressure $M_g^2/(32\pi^2 a^4)$ which seeks to implode the shell. That is, the shell is subject to a net gravitational force density given by the negative divergence of this stress tensor, $\partial_k T_{ik}$. With $r = a$, this force density is of magnitude $4M_g^2/32\pi^2 a^5$ directed radially inwards.

Stressed Shell

The combined effect of both electric and gravitational forces is a net outward pressure on the shell of magnitude

$$P = \frac{e^2 - M_g^2}{32\pi^2 a^4}. \tag{59}$$

Mechanical stability is maintained by a tensile stress p in the shell. At the point considered above in the same coordinate systems, this stress will correspond to a stress tensor T_{ik} given by

$$T_{ik} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & p & 0 \\ 0 & 0 & p \end{bmatrix},$$

where p will have a negative value. If this tensile stress is integrated across the thickness of the shell, we recover the tensile force S per unit length active in the shell

$$S = -\int p ds .$$

The condition for mechanical stability is[76]

$$2S = aP . \tag{60}$$

76 This standard result from the theory of statics can be derived most easily, as Nordström points out, by considering the pressure forces due to P acting on a hemisphere of the shell. A simple integration shows this force is $\pi a^2 P$. This force must be balanced exactly by the tensile force S along the rim of the hemisphere. That rim is of length $2\pi a$, so the total force is $2\pi a S$. Setting $2\pi a S = \pi a^2 P$ entails the result claimed.

Computation of gravitational mass M_g and inertial mass m of the electron

The as yet undetermined gravitational mass M_g of the electron is now recovered by combining the results for the three forms of matter. The source density v is determined by the stress-energy tensor $T_{\mu\nu}$ through equation (38). By assumption, there is no energy associated with the tensile stress in the shell in its rest frame. Thus in a rest frame $T_{uu} = 0$. The spatial components of $T_{\mu\nu}$ are given by the stress tensor T_{ik} above. Therefore

$$\nu = -\frac{1}{c^2}(T_{xx} + T_{yy} + T_{zz} + T_{uu}) = -\frac{1}{c^2}2p \ .$$

We can now recover the gravitational mass M_g from (41) by integrating over the shell

$$M_g = \int g(\Phi)\nu dv = \frac{g(\Phi)}{c^2}8\pi a^2 S \ .$$

We now substitute S in this expression with the condition (60) for mechanical stability and thence for P with the condition (59). By means of (42), we can also express $g(\Phi)$ in terms of $g(\Phi_a) = g_a$ using

$$g(\Phi) = \frac{g_a}{1 + \frac{g_a}{c^2}(\Phi - \Phi_a)} = \frac{g_a}{1 - \frac{g_a}{c^2}\frac{M_g}{4\pi a}} \ .$$

After some algebraic manipulation, we recover an implicit expression for M_g

$$M_g = \frac{g_a}{c^2}\frac{e^2 + M_g^2}{8\pi a} \ .$$

Since this gravitational mass M_g of this complete stationary system resides in an external potential Φ_a, the total mass of the system satisfies $M_g = g_a m$, so that we have for the rest mass m and rest energy E_0 of the electron

$$m = \frac{E_0}{c^2} = \frac{e^2 + M_g^2}{8\pi c^2 a} \ . \tag{61}$$

As Nordström points out of this final result of §3 of his paper is an extremely satisfactory one. The total energy of his electron is made up solely of the sum of an electric component $e^2/8\pi a$ and a gravitational component $M_g^2/8\pi a$. These two components agree exactly with the corresponding classical values. This agreement is not a foregone conclusion since the gravitational mass of the electron arises in an entirely non-classical way: it derives from the fact that the electron shell is stressed. Presumably this agreement justifies Nordström's closing remark in his §3, "Thus the expression found for m contains a verification of the theory."

In his §4, Nordström proceeded to use his expression (61) for the mass m of an electron to introduce the dependence of length on gravitational potential. In accordance with (46), derived in his §2, the mass m must vary in proportion to the external field Φ'_a in the appropriate gauge. However it was not clear how one could recover this same variability from the quantities in the expression (61) for m. He had found in §2 that M_g is independent of the gravitational potential and he asserted that the same held for e according to the basic equations of electrodynamics. Thus he concluded that the radius a of the electron must vary with gravitational potential according to (45). He then turned to Einstein' s more general argument for (45).

ACKNOWLEDGEMENTS

This paper was first published in *Archive for History of Exact Sciences* (Norton 1992b). I am grateful to the Albert Einstein Archives, Hebrew University of Jerusalem, the copyright holder, for their kind permission to quote from Einstein's unpublished writings. I am grateful also to Michel Janssen and Jürgen Renn for helpful discussion.

REFERENCES

Abraham, Max. 1909. "Zur elektromagnetischen Mechanik." *Physikalische Zeitschrift*, 10: 737–41.

———. 1910. "Die Bewegungsgleichungen eines Massenteilchens in der Relativtheorie." *Physikalische Zeitschrift*, 11: 527–31.

———. 1912a. "Zur Theorie der Gravitation." *Physikalische Zeitschrift*, 13: 1–4. (English translation in this volume.)

———. 1912b. "Das Elementargesetz der Gravitation." *Physikalische Zeitschrift*, 13: 4–5.

———. 1912c. "Relativität und Gravitation. Erwiderung auf einer Bemerkung des Hrn. A. Einstein." *Annalen der Physik*, 38: 1056–58.

———. 1912d. "Nochmals Relativität und Gravitation. Bemerkung zu A. Einsteins Erwiderung." *Annalen der Physik*, 39: 444–48.

———. 1912e. "Eine neue Gravitationstheorie." *Archiv der Mathematik und Physik* (3), XX: 193–209. (English translation in this volume.)

Behacker, M. 1913. "Der freie Fall and die Planetenbewegung in Nordströms Gravitationstheorie." *Physikalische Zeitschrift*, 14: 989–992.

Corinaldesi, E. and A. Papapetrou, A. 1951. "Spinning Test Particles in General Relativity II." *Proceedings of the Royal Society, London*, A209: 259–68.

CPAE 2. 1989. John Stachel, David C. Cassidy, Jürgen Renn, and Robert Schulmann (eds.), *The Collected Papers of Albert Einstein*. Vol. 2. *The Swiss Years: Writings, 1900–1909*. Princeton: Princeton University Press.

CPAE 3. 1993. Martin J. Klein, A. J. Kox, Jürgen Renn, and Robert Schulmann (eds.), *The Collected Papers of Albert Einstein*. Vol. 3. *The Swiss Years: Writings, 1909–1911*. Princeton: Princeton University Press.

CPAE 4. 1995. Martin J. Klein, A. J. Kox, Jürgen Renn, and Robert Schulmann (eds.), *The Collected Papers of Albert Einstein*. Vol. 4. *The Swiss Years: Writings, 1912–1914*. Princeton: Princeton University Press.

CPAE 6. 1996. A. J. Kox, Martin J. Klein, and Robert Schulmann (eds.), *The Collected Papers of Albert Einstein*. Vol. 6. *The Berlin Years: Writings, 1914–1917*. Princeton: Princeton University Press.

Ehrenfest, Paul. 1907. "Die Translation deformierbarer Elektronen und der Flächensatz." *Annalen der Physik*, 23: 204–205.

Einstein, Albert. 1907a. "Über das Relativitätsprinzip und die aus demselben gezogenen Folgerungen." *Jahrbuch der Radioaktivität und Elektronik*, (1907) 4: 411–462; (1908) 5: 98–99, (CPAE 2, Doc. 47).

———. 1907b. "Bemerkungen zu der Notiz von Hrn Paul Ehrenfest: 'Die Translation deformierbarer Elektronen und der Flächensatz'." *Annalen der Physik*, 23: 206–208, (CPAE 2, Doc. 44).

———. 1907c. “Über die vom Relativitätsprinzip gefordete Trägheit der Energie.” *Annalen der Physik*, 23: 371–384.

———. 1911. “Über den Einfluss der Schwerkraft auf die Ausbreitung des Lichtes.” *Annalen der Physik*, 35: 898–908, (CPAE 3, Doc. 23).

———. 1912a. “Lichtgeschwindigkeit und Statik des Gravitationsfeldes.” *Annalen der Physik*, 38: 355–69, (CPAE 4, Doc. 3).

———. 1912b. “Zur Theorie des Statischen Gravitationsfeldes.” *Annalen der Physik*, 38: 443–58, (CPAE 4, Doc. 4).

———. 1912c. “Gibt es eine Gravitationswirkung die der elektrodynamischen Induktionswirkung analog ist?” *Vierteljahrsschrift für gerichtliche Medizin und öffentliches Sanitätswesen*, 44: 37–40, (CPAE 4, Doc. 7).

———. 1912d. “Relativität und Gravitation. Erwiderung auf eine Bemerkung von M. Abraham.” *Annalen der Physik*, 38: 1059–64, (CPAE 4, Doc. 8).

———. 1912e. “Bemerkung zu Abrahams vorangehender Auseinandersetzung ‘Nochmals Relativität und Gravitation’.” *Annalen der Physik*, 39: 704, (CPAE 4, Doc. 9).

———. 1913. “Zum gegenwärtigen Stande des Gravitationsproblems.” *Physikalische Zeitschrift*, 14: 1249–1262, (CPAE 4, Doc. 17). (English translation in this volume.)

———. 1915. “Erklärung der Perihelbewegung des Merkur aus der allgemeinen Relativitätstheorie.” *Königlich Preussische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte*. 1915: 831–39, (CPAE 6, Doc. 24).

———. 1933. “Notes on the Origin of the General Theory of Relativity.” In *Ideas and Opinions*, 285–290. Translated by Sonja Bargmann. New York: Crown, 1954.

———. 1949. *Autobiographical Notes*. Open Court, 1979.

Einstein, Albert and Adriaan D. Fokker. 1914. “Die Nordströmsche Gravitationstheorie vom Standpunkt des absoluten Differentialkalküls.” *Annalen der Physik*, 44: 321–28, (CPAE 4, Doc. 28).

Einstein, Albert and Marcel Grossmann. 1913. *Entwurf einer verallgemeinerten Relativitätstheorie und einer Theorie der Gravitation*. Leipzig: B.G.Teubner (separatum); with addendum by Einstein in *Zeitschrift für Mathematik und Physik*, 63(1913): 225–61, (CPAE 4, Doc. 13).

Holton, Gerald. 1975. “Finding Favor with the Angel of the Lord. Notes towards the Psychobiographical Study of Scientific Genius.” In Yehuda Elkana (ed.), *The Interaction between Science and Philosophy.* Humanities Press.

Isaksson, Eva. 1985. “Der finnische Physiker Gunnar Nordström und sein Beitrag zur Entstehung der allgemeinen Relativitätstheorie Albert Einsteins.” *NTM-Schriftennr. Gesch. Naturwiss., Technik, Med.* Leipzig, 22, 1: 29–52.

Janssen, Michel. 1995. “A Comparison between Lorentz’s Ether Theory and Einstein’s Special Theory of Relativity in the Light of the Experiments of Trouton and Noble.” PhD dissertation, University of Pittsburgh.

Laue, Max. 1911a. “Zur Dynamik der Relativitätstheorie.” *Annalen der Physik*, 35: 524–542.

———. 1911b. *Das Relativitätsprinzip*. Braunschweig: Friedrich Vieweg und Sohn.

———. 1911c. “Ein Beispiel zur Dynamik der Relativitätstheorie.” *Verhandlungen der Deutschen Physikalischen Gesellschaft*, 1911: 513–518.

———. 1912. “Zur Theorie des Versuches von Trouton und Noble.” *Annalen der Physik*, 38: 370–84.

———. 1917. “Die Nordströmsche Gravitationstheorie.” *Jahrbuch der Radioaktivität und Elektronik*, 14: 263–313.

———. 1921. *Die Relativitätstheorie*. Vol. 2: *Die allgemeine Relativitätstheorie und Einsteins Lehre von der Schwerkraft*. Braunschweig: Friedrich Vieweg und Sohn.

———. 1949. “Zu Albert Einsteins 70-tem Geburtstag.” *Rev. Mod. Phys.*, 21: 348–49.

Lewis, Gilbert N. and Richard C. Tolman. 1909. “The Principle of Relativity, and Non-Newtonian Mechanics.” *Philosophical Magazine*, 18: 510–523.

Liu, Chuang. 1991. *Relativistic Thermodynamics: Its History and Foundation*. Ph.D. Dissertation, University of Pittsburgh.

Maxwell, James C. 1864. “A dynamical theory of the electromagnetic field.” In W. D. Niven (ed.), *The Scientific Papers of James Clerk Maxwell*. Cambridge University Press, 1890. Reprinted 1965, New York: Dover, 526–597.

Mie, Gustav. 1914. “Bemerkungen zu der Einsteinschen Gravitationstheorie. I and II.” *Physikalische Zeitschrift*, 14: 115–122; 169–176.

Minkowski, Hermann. 1908. “Die Grundgleichungen für die elektromagnetischen Vorgänge in bewegten Körpern.” *Königlichen Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse, Nachrichten* 1908: 53–111; In *Gesammelte Abhandlung von Hermann Minkowski* Vol.2, Leipzig, 1911, 352–404. Reprinted New York: Chelsea. Page citations from this edition. (English translation of the appendix “Mechanics and the Relativity Postulate” in this volume.)

———. 1909. “Raum und Zeit.” *Physikalische Zeitschrift*, 10: 104–111. In *Gesammelte Abhandlung von Hermann Minkowski* Vol.2, Leipzig, 1911, 431–444. Reprinted New York: Chelsea (page citations from this edition). Translated as “*Space and Time*,” in H.A.Lorentz et al., *Principle of Relativity*, 1923, 75–91. Reprinted in 1952, New York: Dover.

Misner, Charles W., Kip S. Thorne, and John A. Wheeler. 1973. *Gravitation*. San Francisco: Freeman.

Nordström, Gunnar. 1909. “Zur Elektrodynamik Minkowskis.” *Physikalische Zeitschrift*, 10: 681–87.

———. 1910. “Zur elektromagnetischen Mechanik.” *Physikalische Zeitschrift*, 11: 440–45.

———. 1911. “Zur Relativitätsmechanik deformierbar Körper.” *Physikalische Zeitschrift*, 12: 854–57.

———. 1912. “Relativitätsprinzip und Gravitation.” *Physikalische Zeitschrift*, 13: 1126–29. (English translation in this volume.)

———. 1913a. “Träge und schwere Masse in der Relativitätsmechanik.” *Annalen der Physik*, 40: 856–78. (English translation in this volume.)

———. 1913b. “Zur Theorie der Gravitation vom Standpunkt des Relativitätsprinzips.” *Annalen der Physik*, 42: 533–54. (English translation in this volume.)

———. 1914a. “Die Fallgesetze und Planetenbewegungen in der Relativitätstheorie.” *Annalen der Physik*, 43: 1101–10.

———. 1914b. “Über den Energiesatz in der Gravitationstheorie.” *Physikalische Zeitschrift*, 14: 375–80.

———. 1914c. “Über die Möglichkeit, das elektromagnetische Feld und das Gravitationsfeld zu vereinigen.” *Physikalische Zeitschrift*, 15: 504–506.

———. 1914d. “Zur Elektrizitäts- und Gravitationstheorie.” *Ofversigt af Finska Vetenskaps-Societetens Förhandlingar*, 57, (1914–1915), Afd. A, N:o.4: 1–15.

———. 1915. “Über eine Mögliche Grundlage einer Theorie der Materie,” *Ofversigt af Finska Vetenskaps-Societetens Förhandlingar*, 57, (1914–1915), Afd. A, N:o.28: 1–21.

Norton, John D. 1984. “How Einstein found his Field Equations: 1912–1915.” *Historical Studies in the Physical Sciences*, 14: 253–316. Reprinted 1989 in Don Howard and John Stachel (eds.), *Einstein and the History of General Relativity* (*Einstein Studies* vol. 1). Boston/Basel/Berlin: Birkhäuser, 101–159.

———. 1985. “What was Einstein’ s Principle of Equivalence?” *Studies in History and Philosophy of Science*, 16, 203–246. Reprinted 1989 in Don Howard and John Stachel (eds.), *Einstein and the History of General Relativity.* (*Einstein Studies* vol. 1.) Boston/Basel/Berlin: Birkhäuser, 3–47. (Page citations from the former.)

———. 1992a. “The Physical Content of General Covariance.” In J. Eisenstaedt and A. Kox (eds.), *Studies in the History of General Relativity.* (*Einstein Studies*, vol. 3.) Boston/Basel/Berlin: Birkhäuser.

———. 1992b. “Einstein, Nordström and the Early Demise of Scalar, Lorentz Covariant Theories of Gravitation.” *Archive for History of Exact Sciences* 45: 17–94.

———. 1993. “Einstein and Nordström: Some Lesser-Known Thought Experiments in Gravitation.” In J. Earman, M. Janssen and J. D. Norton (eds.), *The Attraction of Gravitation: New Studies in the History of General Relativity.* (*Einstein Studies,* vol. 5.) Boston/Basel/Berlin: Birkhäuser, 3–29.

———. 1995. “Eliminative Induction as a Method of Discovery: How Einstein Discovered General Relativity.” In J. Leplin (ed.), *The Creation of Ideas in Physics: Studies for a Methodology of Theory Construction*. Dordrecht: Kluwer, 29–69.

Pais, Abraham. 1982. *Subtle is the Lord ...: The Science and the Life of Albert Einstein*. Oxford: Clarendon.

Papapetrou, Achilles. 1951. “Spinning Test Particle in General Relativity I.” *Proceedings of the Royal Society, London*, A209, 248–58.

Pauli, Wolfgang. 1921. “Relativitätstheorie.” In *Encyklopädie der mathematischen Wissenschaften, mit Einschluss an ihrer Anwendung*. Vol. 5, *Physik*, Part 2. Arnold Sommerfeld (ed.). Leipzig: B.G. Teubner, 1904–1922, 539–775. [Issued November 15, 1921.] English translation *Theory of Relativity.* With supplementary notes by the author. G. Field. Translated in 1958, London: Pergamon. (Citations from the Pergamon edition.)

Planck, Max. 1908. “Zur Dynamik bewegter Körper.” *Annalen der Physik*, 26: 1–34.

Poincaré, Henri. 1905. “Sur la Dynamique de l’ Électron.” *Comptes Rendus des Séances de l’Academie des Sciences*, 140: 1504–1508.

———. 1906. “Sur la Dynamique de l’ Électron,” *Rendiconti del Circolo Matematico di Palermo*, 21: 129–75.

Ricci, Gregorio and Tullio Levi-Civita. 1901. “Méthodes de Calcul Différentiel Absolu et leurs Application.” *Math. Ann.*, 54: 125–201. Reprinted 1954 in T. Levi-Civita, *Opere Matematiche*, vol. 1, Bologna, 479–559.

Rohrlich, Fritz. 1960. “Self-Energy and Stability of the Classical Electron.” *American Journal of Physics*, 28: 639–43.

Sommerfeld, Arnold. 1910. “Zur Relativitätstheorie I. Vierdimensionale Vektoralgebra.” *Annalen der Physik*, 32: 749–776; “Zur Relativitätstheorie II. Vierdimensionale Vektoranalysis.” *Annalen der Physik*, 33: 649–89.

Speziali, Pierre (ed.). 1972. *Albert Einstein-Michele Besso: Correspondance 1903–1955*. Paris: Hermann.

Tolman, Richard C. 1934. *Relativity, Thermodynamics and Cosmology*. Oxford: Oxford University Press; Dover reprint, 1987.

Trouton, Frederick T. and Henry R. Noble. 1903. "The Mechanical Forces Acting on a Charged Condensor moving through Space." *Philosophical Transactions of the Royal Society of London*, 202: 165–181.

Weyl, Hermann. 1918. "Reine Infinitesimal Geometrie." *Mathematische Zeitschrift*, 2: 384–411.

Wheeler, John A. 1979. "Einstein's Last Lecture." In G. E. Tauber (ed.), *Albert Einstein's Theory of General Relativity*. New York: Crown, 187–190.

GUNNAR NORDSTRÖM

THE PRINCIPLE OF RELATIVITY AND GRAVITATION

Originally published as "Relativitätsprinzip und Gravitation" in Physikalische Zeitschrift 13, 1912, 23, pp. 1126–1129. Received October 23, 1912. Author's date: Helsingfors, October 20, 1912.

Einstein's hypothesis that the speed of light c depends upon the gravitational potential[1] leads to considerable difficulties for the principle of relativity, as the discussion between Einstein and Abraham shows us.[2] Hence, one is led to ask if it would not be possible to replace Einstein's hypothesis with a different one, which leaves c constant and still adapts the theory of gravitation to the principle of relativity in such a way that gravitational and inertial mass are equal.[3] I believe that I have found such a hypothesis, and I will present it in the following.

Let x, y, z, u be the four coordinates, with

$$u = ict.$$

Like Abraham,[4] I set

$$\frac{\partial^2\Phi}{\partial x^2} + \frac{\partial^2\Phi}{\partial y^2} + \frac{\partial^2\Phi}{\partial z^2} + \frac{\partial^2\Phi}{\partial u^2} = 4\pi f\gamma, \tag{1}$$

designating the rest density of matter by γ and the gravitational potential by Φ. Φ as well as γ are four-dimensional quantities; f is the gravitational constant. In a gravitational field we have a four-vector

1 A. Einstein, *Ann. d. Phys.* **35**, 898, 1911

2 See *Ann. d. Phys.* **38**, 355, 1056, 1059; **39**, 444, 1912

3 By the equality of inertial and gravitational mass, I do not mean, however, that every inertial phenomenon is caused by the inertial and gravitational mass. For elastically stressed bodies, according to Laue (see below), one obtains a momentum that cannot at all be traced back to a mass. I will return to these questions in a future communication.

4 M. Abraham, this journal, **13**, 1, 1912.

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 3
Gravitation in the Twilight of Classical Physics: Between Mechanics, Field Theory, and Astronomy.

$$\left.\begin{array}{ll} \mathfrak{F}_x = -\dfrac{\partial\Phi}{\partial x}, & \mathfrak{F}_y = -\dfrac{\partial\Phi}{\partial y}, \\ \mathfrak{F}_z = -\dfrac{\partial\Phi}{\partial z}, & \mathfrak{F}_u = -\dfrac{\partial\Phi}{\partial u}, \end{array}\right\} \quad (2)$$

which must be the cause of the acceleration of a mass point located in the field. However, if one considers the four-vector $\mathfrak{F}$ as the *accelerating force* [*bewegende Kraft*] acting on an unchanging unit mass, then the constancy of the speed of light cannot be maintained. In this case, namely, $\mathfrak{F}$ would be equal to the four-dimensional acceleration vector of a mass point and could not remain perpendicular to the velocity vector $\mathfrak{a}$ for arbitrary directions of motion, as demanded by the constancy of the speed of light.[5]

Keeping the speed of light constant, one can nevertheless still eliminate the difficulty in two ways. Either one takes not $\mathfrak{F}$ itself but only its component perpendicular to the velocity vector as the *accelerating force*,[6] or one takes the mass of a mass point to be not constant but dependent on the gravitational potential. On each of these two assumptions, the four-vectors $\mathfrak{F}$ and $\mathfrak{a}$ do not remain parallel: in the first case due to an extra force [*Zusatzkraft*] added to $\mathfrak{F}$, in the second case due to the variability of the mass. As we shall see, the two methods lead to the same laws for the motion of a mass point, but they correspond to two different interpretations of the concept of force.

In accordance with the position of most researchers in the domain of relativity theory, I will first use the second method. Thus, we treat

$$m\mathfrak{F}_x = -m\frac{\partial\Phi}{\partial x}\ etc.$$

as the components of an *accelerating* force acting on a mass point, but view the rest mass of that point as variable. If the components of the velocity vector are $\mathfrak{a}_x, \mathfrak{a}_y, \mathfrak{a}_z, \mathfrak{a}_u$, and τ is the proper time, the equations of motion of the mass point are

$$\left.\begin{array}{l} -m\dfrac{\partial\Phi}{\partial x} = \dfrac{d}{d\tau}(m\mathfrak{a}_x) = m\dfrac{d\mathfrak{a}_x}{d\tau} + \mathfrak{a}_x\dfrac{dm}{d\tau}, \\ -m\dfrac{\partial\Phi}{\partial y} = \dfrac{d}{d\tau}(m\mathfrak{a}_y) = m\dfrac{d\mathfrak{a}_y}{d\tau} + \mathfrak{a}_y\dfrac{dm}{d\tau}, \\ -m\dfrac{\partial\Phi}{\partial z} = \dfrac{d}{d\tau}(m\mathfrak{a}_z) = m\dfrac{d\mathfrak{a}_z}{d\tau} + \mathfrak{a}_z\dfrac{dm}{d\tau}, \\ -m\dfrac{\partial\Phi}{\partial u} = \dfrac{d}{d\tau}(m\mathfrak{a}_u) = m\dfrac{d\mathfrak{a}_u}{d\tau} + \mathfrak{a}_u\dfrac{dm}{d\tau}. \end{array}\right\} \quad (3)$$

5 M. Abraham, loc. cit., eq. (5).

6 Minkowski treats the electrodynamic force in a similar way. Compare *Gött. Nachr.*, 1908, p. 98, eq. (98).

| We multiply the equations in turn by $\mathfrak{a}_x, \mathfrak{a}_y, \mathfrak{a}_z, \mathfrak{a}_u$ and add them. Since [1127]

$$\frac{\partial\Phi}{\partial x}\mathfrak{a}_x + \frac{\partial\Phi}{\partial y}\mathfrak{a}_y + \frac{\partial\Phi}{\partial z}\mathfrak{a}_z + \frac{\partial\Phi}{\partial u}\mathfrak{a}_u = \frac{\partial\Phi}{\partial\tau},$$

$$\mathfrak{a}_x\frac{d\mathfrak{a}_x}{d\tau} + \mathfrak{a}_y\frac{d\mathfrak{a}_y}{d\tau} + \mathfrak{a}_z\frac{d\mathfrak{a}_z}{d\tau} + \mathfrak{a}_u\frac{d\mathfrak{a}_u}{d\tau} = 0,$$

$$\mathfrak{a}_x{}^2 + \mathfrak{a}_y{}^2 + \mathfrak{a}_z{}^2 + \mathfrak{a}_u{}^2 = -c^2,$$

we obtain

$$\left.\begin{aligned} -m\frac{d\Phi}{d\tau} &= -c^2\frac{dm}{d\tau}, \\ \frac{1}{m}\frac{dm}{d\tau} &= \frac{1}{c^2}\frac{d\Phi}{d\tau}. \end{aligned}\right. \tag{4}$$

Integration yields

$$\log m = \frac{1}{c^2}\Phi + \text{const},$$

or

$$m = m_0 e^{\frac{\Phi}{c^2}}. \tag{5}$$

This equation shows that the mass m depends on the gravitational potential according to a simple law.

Using (4) the equations of motion (3) can also be written in the following form:

$$\left.\begin{aligned} -\frac{\partial\Phi}{\partial x} &= \frac{d\mathfrak{a}_x}{d\tau} + \frac{\mathfrak{a}_x}{c^2}\frac{d\Phi}{d\tau}, \\ -\frac{\partial\Phi}{\partial y} &= \frac{d\mathfrak{a}_y}{d\tau} + \frac{\mathfrak{a}_y}{c^2}\frac{d\Phi}{d\tau}, \\ -\frac{\partial\Phi}{\partial z} &= \frac{d\mathfrak{a}_z}{d\tau} + \frac{\mathfrak{a}_z}{c^2}\frac{d\Phi}{d\tau}, \\ -\frac{\partial\Phi}{\partial u} &= \frac{d\mathfrak{a}_u}{d\tau} + \frac{\mathfrak{a}_u}{c^2}\frac{d\Phi}{d\tau}. \end{aligned}\right\} \tag{6}$$

As one can see, the mass m drops out of these equations. The laws according to which a mass point moves in a gravitational field are thus completely independent of the mass of the point.

The considerations so far are based on the assumption that $m\mathfrak{F}$ is the accelerating force. Now, for the moment, we wish to assume that the component of $m\mathfrak{F}$ perpendicular to the velocity vector $\mathfrak{a}$, rather than $m\mathfrak{F}$ itself, is the accelerating force. This part of $m\mathfrak{F}$ is a four-vector having an x-component[7]

$$m\mathfrak{F}_x + m\frac{\mathfrak{a}_x}{c^2}\{\mathfrak{F}_x\mathfrak{a}_x + \mathfrak{F}_y\mathfrak{a}_y + \mathfrak{F}_z\mathfrak{a}_z + \mathfrak{F}_u\mathfrak{a}_u\}.$$

The second term is the x-component of an extra force added to $m\mathfrak{F}$. According to (2) the expression can be changed to

$$-m\left\{\frac{\partial\Phi}{\partial x} + \frac{\mathfrak{a}_x}{c^2}\frac{\partial\Phi}{d\tau}\right\}.$$

Since we now view m as constant, the first of the equations of motion of a mass point is

$$-m\left\{\frac{\partial\Phi}{\partial x} + \frac{\mathfrak{a}_x}{c^2}\frac{\partial\Phi}{d\tau}\right\} = m\frac{d\mathfrak{a}_x}{d\tau}.$$

But this is just the first of the equations of motion (6).

From the two alternative assumptions we obtain precisely the same laws describing the motion of a mass point in a gravitational field, only the force and the mass are conceptualized differently in the two cases. The latter way of thinking corresponds to Minkowski's original, the way treated first corresponds to that held by Laue and Abraham.[8]

So far we have considered an isolated point mass. Now we would like to investigate the motion of arbitrary bodies in a gravitational field and develop the law of conservation of energy for this process. We assume only that the mass of each particle of the bodies actually is something real, so that we can speak of the rest density γ of the spacetime points. This is certainly the case when no tangential stresses are present in the body.[9] The rest density γ is of course a function of the four coordinates

$$\gamma = \gamma(x, y, z, u).$$

We again view mass as variable and accept the concept of force equation (3) is based on. Then the components of the force exerted by gravitation on a *unit volume* of matter are[10]

$$\left.\begin{aligned} \mathfrak{K}_x &= -\gamma\frac{\partial\Phi}{\partial x}, & \mathfrak{K}_y &= -\gamma\frac{\partial\Phi}{\partial y}, \\ \mathfrak{K}_z &= -\gamma\frac{\partial\Phi}{\partial z}, & \mathfrak{K}_u &= -\gamma\frac{\partial\Phi}{\partial u}. \end{aligned}\right\} \tag{7}$$

7 Cf. H. Minkowski, loc. cit.

8 Cf. the discussion between Abraham and the author, this journal **10**, 681, 737, 1909; **11**, 440, 527, 1910. I now take the position then taken up by Abraham.

9 Cf. M. Laue, *Das Relativitätsprinzip*, Braunschweig 1911, p. 151 f.; G. Nordström, this journal **12**, 854, 1911.

10 If the four-vector $\mathfrak{K}$ is taken to be the *accelerating* force, then it should be designated as the "*accelerating force per unit rest volume*."

For the sake of generality, we assume that besides gravitation an "external" force $\mathfrak{K}'$ with components

$$\mathfrak{K}_x{}',\ \mathfrak{K}_y{}',\ \mathfrak{K}_z{}',\ \mathfrak{K}_u{}'$$

acts on the unit volume of matter. We can then write the equations of motion of matter in the following general form[11]

$$\left.\begin{aligned}
-\gamma\frac{\partial\Phi}{\partial x}+\mathfrak{K}_x{}' &= \frac{\partial}{\partial x}\gamma\mathfrak{a}_x{}^2+\frac{\partial}{\partial y}\gamma\mathfrak{a}_x\mathfrak{a}_y+\frac{\partial}{\partial z}\gamma\mathfrak{a}_x\mathfrak{a}_z+\frac{\partial}{\partial u}\gamma\mathfrak{a}_x\mathfrak{a}_u,\\
&-\ -\ -\ -\ -\ -\ -\ -\ -\ -\ -\ -\ -\ -\ -\\
&-\ -\ -\ -\ -\ -\ -\ -\ -\ -\ -\ -\ -\ -\ -\\
-\gamma\frac{\partial\Phi}{\partial u}+\mathfrak{K}_u{}' &= \frac{\partial}{\partial x}\gamma\mathfrak{a}_u\mathfrak{a}_x+\frac{\partial}{\partial y}\gamma\mathfrak{a}_u\mathfrak{a}_y+\frac{\partial}{\partial z}\gamma\mathfrak{a}_u\mathfrak{a}_z+\frac{\partial}{\partial u}\gamma\mathfrak{a}_u{}^2.
\end{aligned}\right\}\qquad(8)$$ [1128]

If we wish to introduce the ordinary three-dimensional velocity v and the ordinary mass density ρ, we have to set

$$\mathfrak{a}_x = \frac{\mathfrak{v}_x}{\sqrt{1-\mathfrak{q}^2}},\ \cdot\ \cdot\ \cdot\ \cdot\ ,\ \mathfrak{a}_u = \frac{ic}{\sqrt{1-\mathfrak{q}^2}},$$
$$\gamma = \rho\sqrt{1-\mathfrak{q}^2},$$

where $\mathfrak{q} = \mathfrak{v}/c$ has been substituted for reasons of simplicity. We multiply the last of the equations (8) by $-ic$ and insert the expressions above into its right-hand side. Continuing to use the notation of three-dimensional vector analysis, the equation becomes

$$\gamma\frac{\partial\Phi}{\partial t}-ic\mathfrak{K}_u{}' = c^2\mathrm{div}\frac{\rho\mathfrak{v}}{\sqrt{1-\mathfrak{q}^2}}+c^2\frac{\partial}{\partial t}\frac{\rho}{\sqrt{1-\mathfrak{q}^2}}.\qquad(9)$$

We wish to transform the first term. Equation (1) yields

$$4\pi f\gamma = \mathrm{div}\nabla\Phi-\frac{1}{c^2}\frac{\partial^2\Phi}{\partial t^2},$$

and thus

11 G. Nordström, this journal **11**, 441, eq. (4′), 1910.

$$\gamma\frac{\partial\Phi}{\partial t} = \frac{1}{4\pi f}\left\{\frac{\partial\Phi}{\partial t}\operatorname{div}\nabla\Phi - \frac{1}{c^2}\frac{\partial\Phi}{\partial t}\frac{\partial}{\partial t}\left(\frac{\partial\Phi}{\partial t}\right)\right\}$$

$$= \frac{1}{4\pi f}\left\{\operatorname{div}\left(\frac{\partial\Phi}{\partial t}\nabla\Phi\right) - \nabla\Phi\cdot\frac{\partial}{\partial t}\nabla\Phi - \frac{1}{2c^2}\frac{\partial}{\partial t}\left(\frac{\partial\Phi}{\partial t}\right)^2\right\},$$

$$\gamma\frac{\partial\Phi}{\partial t} = \frac{1}{4\pi f}\operatorname{div}\left(\frac{\partial\Phi}{\partial t}\nabla\Phi\right) - \frac{1}{8\pi f}\frac{\partial}{\partial t}\left\{(\nabla\Phi)^2 + \frac{1}{c^2}\left(\frac{\partial\Phi}{\partial t}\right)^2\right\}.$$

We insert this expression into equation (9) and obtain the following equation, which expresses the law of conservation of energy:

$$\left.\begin{aligned} -ic\mathfrak{K}_u{}' = -\frac{1}{4\pi f}\operatorname{div}\left(\frac{\partial\Phi}{\partial t}\nabla\Phi\right) + \frac{1}{8\pi f}\frac{\partial}{\partial t}\left\{(\nabla\Phi)^2 + \frac{1}{c^2}\left(\frac{\partial\Phi}{\partial t}\right)^2\right\} \\ + c^2\operatorname{div}\frac{\rho\mathfrak{v}}{\sqrt{1-\mathfrak{q}^2}} + c^2\frac{\partial}{\partial t}\frac{\rho}{\sqrt{1-\mathfrak{q}^2}}. \end{aligned}\right\} \tag{10}$$

The quantity $-ic\mathfrak{K}_u{}'$ represents the energy influx caused by the external force $\mathfrak{K}'$ per unit volume and per unit time. Of the terms on the right-hand side, the first two terms relate to the gravitational field, the last two to the matter of the bodies. We set

$$\mathfrak{S}^g = -\frac{1}{4\pi f}\frac{\partial\Phi}{\partial t}\nabla\Phi. \tag{11}$$

$$\psi^g = \frac{1}{8\pi f}\left\{(\nabla\Phi)^2 + \frac{1}{c^2}\left(\frac{\partial\Phi}{\partial t}\right)^2\right\}. \tag{12}$$

$$\mathfrak{S}^m = \frac{c^2\rho\mathfrak{v}}{\sqrt{1-\mathfrak{q}^2}}, \tag{13}$$

$$\psi^m = \frac{c^2\rho}{\sqrt{1-\mathfrak{q}^2}}. \tag{14}$$

ψ^g is the *energy density* of the gravitational field, $\mathfrak{S}^g$ is the *energy flux* of this field, ψ^m and $\mathfrak{S}^m$ are the energy density and the convective energy flux of matter. For these quantities, we have found the expressions (13) and (14) already known earlier.[12]

We note that according to (12) the energy density of the field is always positive.

Finally, the energy equation is written as

12 G. Nordström, 1oc. cit., eqs. (11) and (12); M. Laue, loc. cit., § 24.

$$-ic\mathfrak{K}_u{}' = \operatorname{div}(\mathfrak{S}^g + \mathfrak{S}^m) + \frac{\partial}{\partial t}(\psi^g + \psi^m). \tag{15}$$

We see that the law of conservation of energy is satisfied.

The quantities $\mathfrak{S}^g$ and ψ^g depend on a four-dimensional tensor, which also yields fictitious stresses for the gravitational force $\mathfrak{K}$. This tensor is precisely the same as that which Abraham obtained using different assumptions.[13] The ten components of the gravitation tensor are

$$\left.\begin{aligned}
X_x &= \frac{1}{4\pi f}\left\{-\left(\frac{\partial\Phi}{\partial x}\right)^2 + \Psi\right\}, \\
&- - - - - - - - \\
&- - - - - - - - \\
U_u &= \frac{1}{4\pi f}\left\{-\left(\frac{\partial\Phi}{\partial u}\right)^2 + \Psi\right\}, \\
X_y &= Y_x = -\frac{1}{4\pi f}\frac{\partial\Phi}{\partial x}\frac{\partial\Phi}{\partial y}, \\
&- - - - - - - - \\
&- - - - - - - - \\
&- - - - - - - - \\
&- - - - - - - - \\
Z_u &= U_z = -\frac{1}{4\pi f}\frac{\partial\Phi}{\partial z}\frac{\partial\Phi}{\partial u},
\end{aligned}\right\} \tag{16}$$

where Ψ is the following four-dimensional scalar

$$\Psi = \frac{1}{2}\left\{\left(\frac{\partial\Phi}{\partial x}\right)^2 + \left(\frac{\partial\Phi}{\partial y}\right)^2 + \left(\frac{\partial\Phi}{\partial z}\right)^2 + \left(\frac{\partial\Phi}{\partial u}\right)^2\right\}. \tag{16a}$$ [1129]

It can be easily shown that in fact

$$-\gamma\frac{\partial\Phi}{\partial x} = \frac{\partial X_x}{\partial x} + \frac{\partial X_y}{\partial y} + \frac{\partial X_z}{\partial z} + \frac{\partial X_u}{\partial u}, \qquad \text{etc.}$$

$$\mathfrak{S}_x{}^g = icU_x, \qquad \mathfrak{S}_y{}^g = icU_y, \qquad \mathfrak{S}_z{}^g = icU_z, \qquad \psi^g = U_u.$$

13 M. Abraham, loc. cit., p. 3.

Because the gravitation tensor is symmetric, the momentum density is equal to the energy flux divided by c^2.

Equation (4), which expresses the variability of the mass of a mass point, can be easily generalized to extended masses. For this purpose, we have to treat the system of equations (8) in the same way as we treated the system of equations (3) earlier. We multiply the equations (8) in turn by $\mathfrak{a}_x, \mathfrak{a}_y, \mathfrak{a}_z, \mathfrak{a}_u$ and add them. If no causes other than the gravitational field lead to a variability of mass, the external force $\mathfrak{K}'$ is perpendicular to $\mathfrak{a}$, and after some rearranging one obtains

$$\frac{\partial}{\partial x}\gamma\mathfrak{a}_x + \frac{\partial}{\partial y}\gamma\mathfrak{a}_y + \frac{\partial}{\partial z}\gamma\mathfrak{a}_z + \frac{\partial}{\partial u}\gamma\mathfrak{a}_u = \frac{\gamma}{c^2}\frac{d\Phi}{d\tau}, \tag{17}$$

or

$$\mathrm{div}\rho\mathfrak{v} + \frac{\partial\rho}{\partial t} = \frac{\rho}{c^2}\left\{\mathfrak{v}\nabla\Phi + \frac{\partial\Phi}{\partial t}\right\}, \tag{18}$$

or still

$$\frac{d}{dt}(\rho dv) = \frac{\rho dv}{c^2}\frac{d\Phi}{dt} \tag{18a}$$

(dv is the volume element). These three equivalent equations express in general the law of the dependence of mass on the gravitational field.

Equation (1) can be integrated in a well-known manner. One obtains the well-known expression for the retarded potential

$$\Phi(x_0, y_0, z_0, t) = -f\cdot\int\frac{dx\ dy\ dz}{r}\gamma_{t-\frac{r}{c}} + \text{const}, \tag{19}$$

$$\text{where } r = \sqrt{(x-x_0)^2+(y-y_0)^2+(z-z_0)^2}\,.$$

The integration is over three-dimensional space, and γ is evaluated at the time $t-\frac{r}{c}$.

From equations (5) and (19) it becomes apparent that point masses cannot really exist because within such a mass point $\Phi = -\infty$, and hence the mass would be zero. If a body contracts, its mass decreases and with vanishing volume its mass would also vanish. As far as I can see, these consequences of the theory do not lead to contradictions.

Obviously, the theory developed here has much in common with the one which Abraham presented in this journal **13**, 1, 1912, but later refuted.[14] The theory developed here, however, is free from all the maladies which are brought about by the variability of the speed of light in the theories of Einstein and Abraham.

Addendum to proofs: From a letter from Herr Prof. Dr. A. Einstein I learn that earlier he had already concerned himself with the possibility I used above for treating gravitational phenomena in a simple way. However, he became convinced that the

14 M. Abraham, this journal **13**, 793, 1912.

consequences of such a theory cannot correspond with reality. In a simple example he shows that, according to this theory, a rotating system in a gravitational field will acquire a smaller acceleration than a non-rotating system.

I do not find this result dubious in itself, for the difference is too small to yield a contradiction with experience. Of course, the result under discussion shows that my theory is not compatible with Einstein's hypothesis of equivalence, according to which an unaccelerated reference system in a homogeneous gravitational field is equivalent to an accelerated reference system in a gravitation free space.

In this circumstance, however, I do not see a sufficient reason to reject the theory. For, even though Einstein's hypothesis is extraordinarily ingenious, on the other hand it still provides great difficulties. Therefore other attempts at treating gravitation are also desirable and I want to provide a contribution to them with my communication.

GUNNAR NORDSTRÖM

INERTIAL AND GRAVITATIONAL MASS IN RELATIVISTIC MECHANICS

Originally published as "Träge und schwere Masse in der Relativitätsmechanik" in Annalen der Physik, 40, 1913, pp. 856–878. Received 21. January 1913. Author's date: Helsingfors, January 1913.

In several recent papers in the field of relativistic mechanics, the concept of the mass of bodies plays a very subordinate role. The reason is easy to understand. As Laue[1] and Herglotz[2] have shown, one can develop the entire mechanics of extended bodies without exploiting the concept of inertial mass in any way. Thus the concept of mass is not absolutely essential for mechanics, and on the other hand, if one considers bodies subject to arbitrary elastic stresses, this concept is also not sufficient to describe all inertial phenomena of matter.

But the question of the mass of matter is nevertheless of considerable importance for the theory of relativity, especially for the assessment of the way in which the theory of gravitation is to be integrated into the theory of relativity. In any case inertia and gravity [*Schwere*] of matter must stand in close relation to each other, and it would be easiest to account for this unity of essence [*Wesenseinheit*] via the mass underlying these two phenomena. One would attempt to retain such a concept of mass, even though it is known that according to relativity theory there exist inertial phenomena which cannot be traced back to mass in any way. In such cases, one must make use of a specially defined momentum, which depends, for example, upon the state of elastic stress of a body | rather than upon its mass. [857]

In the present paper, I will treat the relativistic mechanics of deformable bodies in such a way that the possibility of generally maintaining the concept of mass is clearly emphasized. On this occasion, I will also investigate the influence of the heat conduction on mechanical processes. Finally, I will consider gravitation by also ascribing gravity to the inertial mass.

1 M. Laue, *Das Relativitätsprinzip*, Braunschweig 1911, VII; *Ann. d. Phys.*, 35, p. 524, 1911.
2 G. Herglotz, *Ann. d. Phys.,* 36, p. 493, 1911.

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 3
Gravitation in the Twilight of Classical Physics: Between Mechanics, Field Theory, and Astronomy.

1. THE FOUNDATIONS OF THE RELATIVISTIC MECHANICS OF DEFORMABLE BODIES

We consider a body in an arbitrary state of motion and arbitrary state of stress. In addition to the elastic forces, a spatially distributed ponderomotive force $\mathfrak{K}$ of any kind may act on the bodies. $\mathfrak{K}$ is a four-vector which is to be designated the "external" ponderomotive force per unit volume, or the "external" accelerating force [*bewegende Kraft*] per unit of rest volume.[3]

According to Laue,[4] there is a symmetric four-dimensional tensor $\boldsymbol{T}$, whose components give the spatial stresses as well as the mechanical energy-momentum density. Accordingly, we can write the equations of motion of the body in the following form:

$$\begin{aligned}
\mathfrak{K}_x &= \frac{\partial \boldsymbol{T}_{xx}}{\partial x} + \frac{\partial \boldsymbol{T}_{xy}}{\partial y} + \frac{\partial \boldsymbol{T}_{xz}}{\partial z} + \frac{\partial \boldsymbol{T}_{xu}}{\partial u}, \\
\mathfrak{K}_y &= \frac{\partial \boldsymbol{T}_{yx}}{\partial x} + \frac{\partial \boldsymbol{T}_{yy}}{\partial y} + \frac{\partial \boldsymbol{T}_{yz}}{\partial z} + \frac{\partial \boldsymbol{T}_{yu}}{\partial u}, \\
\mathfrak{K}_z &= \frac{\partial \boldsymbol{T}_{zx}}{\partial x} + \frac{\partial \boldsymbol{T}_{zy}}{\partial y} + \frac{\partial \boldsymbol{T}_{zz}}{\partial z} + \frac{\partial \boldsymbol{T}_{zu}}{\partial u}, \\
\mathfrak{K}_u &= \frac{\partial \boldsymbol{T}_{ux}}{\partial x} + \frac{\partial \boldsymbol{T}_{uy}}{\partial y} + \frac{\partial \boldsymbol{T}_{uz}}{\partial z} + \frac{\partial \boldsymbol{T}_{uu}}{\partial u},
\end{aligned} \tag{1}$$

where $x, y, z, u = ict$ are the four coordinates; the speed of light c is supposed to be a universal constant. |

[858] We want to assign to each spacetime point of matter a certain *rest-mass density* ν. This quantity is to be a four-dimensional scalar, but otherwise for the time being we leave it completely undetermined, so that we still have the freedom to further specify the concept of mass. From the rest density, the usual mass density μ is determined by the equation

$$\nu = \mu\sqrt{1 - \frac{\mathfrak{v}^2}{c^2}}\,, \tag{2}$$

where $\mathfrak{v}$ represents the (three-dimensional) velocity of the point in question. For simplicity we set

$$\mathfrak{q} = \frac{\mathfrak{v}}{c},$$

and therefore have

$$\nu = \mu\sqrt{1 - \mathfrak{q}^2}.$$

---

3 H. Minkowski, *Gött. Nachr.*, 1908, p. 107 and 108 [excerpts from this article are contained in this volume]; compare also eqs. (6) and (9) below.

4 M. Laue, *Das Relativiätsprinzip*, p. 149.

Now we take the four-dimensional tensor $\boldsymbol{T}$ to be the sum of two such tensors by setting

$$\left.\begin{aligned}
\boldsymbol{T}_{xx} &= \boldsymbol{p}_{xx} + \nu\mathfrak{B}_x^{\ 2},\\
&\cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\\
&\cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\\
\boldsymbol{T}_{uu} &= \boldsymbol{p}_{uu} + \nu\mathfrak{B}_u^{\ 2},\\
\boldsymbol{T}_{xy} &= \boldsymbol{p}_{xy} + \nu\mathfrak{B}_x\mathfrak{B}_y,\\
&\cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\\
&\cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\\
&\cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\\
&\cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\\
\boldsymbol{T}_{zu} &= \boldsymbol{p}_{zu} + \nu\mathfrak{B}_z\mathfrak{B}_u,
\end{aligned}\right\} \tag{3}$$

where $\mathfrak{B}$ represents the four-dimensional velocity vector, which, as is well known, is related to the velocity $\mathfrak{v}$ by the equations

$$\mathfrak{B}_x = \frac{\mathfrak{v}_x}{\sqrt{1-\mathfrak{q}^2}}, \ \ldots\ \mathfrak{B}_u = \frac{ic}{\sqrt{1-\mathfrak{q}^2}}. \tag{4}$$

We call the four-dimensional tensor $\boldsymbol{p}$ introduced in eq. (3)[1] the *elastic stress tensor*. Like $\boldsymbol{T}$, it is symmetric, since $\boldsymbol{p}_{xy} = \boldsymbol{p}_{yx}$ etc. The second part of the tensor $\boldsymbol{T}$ can be called the *material* tensor. |

We set [859]

$$\left.\begin{aligned}
\mathfrak{K}_x^{\ e} &= -\frac{\partial\boldsymbol{p}_{xx}}{\partial x} - \frac{\partial\boldsymbol{p}_{xy}}{\partial y} - \frac{\partial\boldsymbol{p}_{xz}}{\partial z} - \frac{\partial\boldsymbol{p}_{xu}}{\partial u},\\
&\cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\\
&\cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\\
\mathfrak{K}_u^{\ e} &= -\frac{\partial\boldsymbol{p}_{ux}}{\partial x} - \frac{\partial\boldsymbol{p}_{uy}}{\partial y} - \frac{\partial\boldsymbol{p}_{uz}}{\partial z} - \frac{\partial\boldsymbol{p}_{uu}}{\partial u},
\end{aligned}\right\} \tag{5}$$

and call the vector $\mathfrak{K}^e$ the *elastic* ponderomotive force. Our equations of motion (1) are now

$$\left.\begin{aligned}
\mathfrak{K}_x + \mathfrak{K}_x^{\;e} &= \frac{\partial}{\partial x}\nu\mathfrak{B}_x^{\;2} + \frac{\partial}{\partial y}\nu\mathfrak{B}_x\mathfrak{B}_y + \frac{\partial}{\partial z}\nu\mathfrak{B}_x\mathfrak{B}_z + \frac{\partial}{\partial u}\nu\mathfrak{B}_x\mathfrak{B}_u,\\
\mathfrak{K}_y + \mathfrak{K}_y^{\;e} &= \frac{\partial}{\partial x}\nu\mathfrak{B}_y\mathfrak{B}_x + \frac{\partial}{\partial y}\nu\mathfrak{B}_y^{\;2} + \frac{\partial}{\partial z}\nu\mathfrak{B}_y\mathfrak{B}_z + \frac{\partial}{\partial u}\nu\mathfrak{B}_y\mathfrak{B}_u,\\
\mathfrak{K}_z + \mathfrak{K}_z^{\;e} &= \frac{\partial}{\partial x}\nu\mathfrak{B}_z\mathfrak{B}_x + \frac{\partial}{\partial y}\nu\mathfrak{B}_z\mathfrak{B}_y + \frac{\partial}{\partial z}\nu\mathfrak{B}_z^{\;2} + \frac{\partial}{\partial u}\nu\mathfrak{B}_z\mathfrak{B}_u,\\
\mathfrak{K}_u + \mathfrak{K}_u^{\;e} &= \frac{\partial}{\partial x}\nu\mathfrak{B}_u\mathfrak{B}_x + \frac{\partial}{\partial y}\nu\mathfrak{B}_u\mathfrak{B}_y + \frac{\partial}{\partial z}\nu\mathfrak{B}_u\mathfrak{B}_z + \frac{\partial}{\partial u}\nu\mathfrak{B}_u^{\;2}.
\end{aligned}\right\} \tag{6}$$

In order to understand the meaning of the right-hand sides, we transform them. We denote the volume of a material particle of the body by dv, and the rest volume of the same by dv_0, where

$$dv_0 = \frac{dv}{\sqrt{1-\mathfrak{q}^2}}. \tag{7}$$

Furthermore, if τ denotes the proper time

$$d\tau = dt\sqrt{1-\mathfrak{q}^2}, \tag{8}$$

then by introducing $\mathfrak{v}$ and using a well known formula[5] one obtains |

[860]
$$\left.\begin{aligned}
&\frac{\partial}{\partial x}\nu\mathfrak{B}_x^{\;2} + \frac{\partial}{\partial y}\nu\mathfrak{B}_x\mathfrak{B}_y + \frac{\partial}{\partial z}\nu\mathfrak{B}_x\mathfrak{B}_z + \frac{\partial}{\partial u}\nu\mathfrak{B}_x\mathfrak{B}_u\\
&\quad = \frac{1}{dv_0}\frac{d}{d\tau}(\nu\mathfrak{B}_x dv_0) = \frac{1}{dv}\frac{d}{dt}\left\{\frac{\mu\mathfrak{v}_x}{\sqrt{1-\mathfrak{q}^2}}dv\right\}.
\end{aligned}\right\} \tag{9}$$

Changing the index x to y, z, u respectively, one obtains the corresponding equations. Inserting these expressions into (6), one obtains the equations of motion in a similar form, appropriate for a material point.

As usual, the first three of the equations of motion are supposed to express the law of conservation of momentum, the fourth that of conservation of energy. In order to study the first law more closely, we set

$$\mathfrak{g}_x^{\;e} = -\frac{i}{c}\boldsymbol{p}_{xu}, \qquad \mathfrak{g}_y^{\;e} = -\frac{i}{c}\boldsymbol{p}_{yu}, \qquad \mathfrak{g}_z^{\;e} = -\frac{i}{c}\boldsymbol{p}_{zu}, \tag{10}$$

---

5 $$\frac{d}{dt}\int\varphi dv = \int\left\{\operatorname{div}\ \varphi\mathfrak{v} + \frac{\partial\varphi}{\partial t}\right\}dv,$$
where φ is an arbitrary function of x, y, z, t and the integration on the left extends over any particular part of the matter. The vector symbols in this essay are those explained in Abraham's *Theorie der Elektrizität*, Vol. I; they always refer to *three-dimensional* vectors.

and call the three-dimensional vector $\mathfrak{g}^e$ the *elastic momentum density*. The vector $\mathfrak{g}^m$, constructed in a similar way from the material tensor, with the components

$$\left(\mathfrak{g}_x{}^m = -\frac{i}{c}\nu\mathfrak{B}_x\mathfrak{B}_u\right) \text{ etc.,}$$

shall be called the *material momentum density*. One finds

$$\mathfrak{g}^m = \frac{\mu\mathfrak{v}}{\sqrt{1-\mathfrak{q}^2}}. \tag{11}$$

We furthermore introduce the *relative stresses* $\boldsymbol{t}$ [6] through the following equation

$$\left.\begin{aligned} \boldsymbol{t}_{xx} &= \boldsymbol{p}_{xx} + \frac{i}{c}\boldsymbol{p}_{xu}\mathfrak{v}_x, \\ \boldsymbol{t}_{xy} &= \boldsymbol{p}_{xy} + \frac{i}{c}\boldsymbol{p}_{xu}\mathfrak{v}_y, \\ &\text{etc.,} \end{aligned}\right\} \tag{12}$$

or by (10),

$$\left.\begin{aligned} \boldsymbol{t}_{xx} &= \boldsymbol{p}_{xx} - \mathfrak{g}_x{}^e\mathfrak{v}_x, \\ \boldsymbol{t}_{xy} &= \boldsymbol{p}_{xy} - \mathfrak{g}_x{}^e\mathfrak{v}_y, \\ &\text{etc.,} \end{aligned}\right\} \tag{12a}$$

| The relative stresses form a three-dimensional asymmetric tensor. Clearly, the calculation of these stresses is similar to that of the pressure on a moving surface in electrodynamics. Furthermore, it should be noted that writing $\boldsymbol{T}$ instead of $\boldsymbol{p}$ in (12) yields the same relative stresses, because the second (material) tensor into which we partitioned $\boldsymbol{T}$ contributes zero relative stress. For this reason, the relative stresses defined by eq. (12) are identical with those introduced by Laue (loc. cit.). [861]

We can now transform the expressions for the spatial components of $\mathfrak{K}^e$. We obtain from (10) and (12a)

$$\begin{aligned} \mathfrak{K}_x{}^e = &-\left\{\frac{\partial\boldsymbol{t}_{xx}}{\partial x} + \frac{\partial\boldsymbol{t}_{xy}}{\partial y} + \frac{\partial\boldsymbol{t}_{xz}}{\partial z}\right\} \\ &-\left\{\frac{\partial}{\partial x}\mathfrak{g}_x{}^e\mathfrak{v}_x + \frac{\partial}{\partial y}\mathfrak{g}_x{}^e\mathfrak{v}_y + \frac{\partial}{\partial z}\mathfrak{g}_x{}^e\mathfrak{v}_z + \frac{\partial}{\partial t}\mathfrak{g}_x{}^e\right\}, \end{aligned} \tag{13}$$

and the corresponding expressions for $\mathfrak{K}_y{}^e$ and $\mathfrak{K}_z{}^e$.

---

6 M. Abraham, “Zur Elektrodynamik bewegter Körper,” *Rend. Circ. Matem. Palermo*, eq. (10) 1909; M. Laue, loc. cit. p. 151.

We multiply eq. (13) by dv and integrate over a (three-dimensional) space v filled with mass. The integral of the expression in the first bracket can be converted into a surface integral by Gauss's theorem. Also, applying the formula in footnote 5 [p. 859 in the original] to the final bracketed expression, we obtain

$$\int \mathfrak{K}_x^e dv = -\int\{\boldsymbol{t}_{xx}d\mathfrak{f}_x + \boldsymbol{t}_{xy}d\mathfrak{f}_y + \boldsymbol{t}_{xz}d\mathfrak{f}_z\} - \frac{d}{dt}\int \mathfrak{g}_x^e dv. \tag{14}$$

Here, $d\mathfrak{f}_x$, $d\mathfrak{f}_y$, $d\mathfrak{f}_z$ are the components of an area element of the surface bounding the region v under consideration, with $d\mathfrak{f}$ taken as a vector in the direction of the external normal. The symbol d/dt denotes the temporal change in a bounded region of the matter.

Corresponding expressions hold for the remaining spatial axis directions, and it is clear that the elastic force is partially determined by the relative elastic stresses acting as area forces [*Flächenkräfte*], and partially by the change of the elastic momentum.

According to (6), (9) and (14), we can now write the first of the equations of motion in the following integral form: |

[862]

$$\begin{aligned} &\int \mathfrak{K}_x dv - \int\{\boldsymbol{t}_{xx}d\mathfrak{f}_x + \boldsymbol{t}_{xy}d\mathfrak{f}_y + \boldsymbol{t}_{xz}d\mathfrak{f}_z\} - \frac{d}{dt}\int \mathfrak{g}_x{}^e dv \\ &\qquad = \frac{d}{dt}\int \frac{\mu\mathfrak{v}_x}{\sqrt{1-\mathfrak{q}^2}}dv = \frac{d}{dt}\int \mathfrak{g}_x{}^m dv. \end{aligned} \tag{15}$$

These equations and the two analogous ones for the remaining axis directions express the law of conservation of momentum for a bounded region of the matter.

The asymmetry of the relative stress tensor implies that the elastic forces in general apply a torque[7] to each part of the body. According to the theory of elasticity, the torque acting on the unit volume about an axis parallel to the x-axis is

$$\boldsymbol{t}_{yz} - \boldsymbol{t}_{zy} = \mathfrak{v}_y\mathfrak{g}_x{}^e - \mathfrak{v}_z\mathfrak{g}_y{}^e.$$

Hence, expressed vectorially one has for the torque $\mathfrak{n}$ per unit volume

$$\mathfrak{n} = [\mathfrak{v}\mathfrak{g}^e]. \tag{16}$$

Thus this torque must always appear when the momentum density has a component perpendicular to the velocity. The torque is thus necessary also to retain the uniform translatory motion of the elastically stressed body. As is well known, this is a significant difference between classical and relativistic mechanics, the reason for which becomes clear when establishing the area law.[8] However, we do not want to discuss this here.

7 M. Laue, loc. cit., p. 168.
8 M.Laue, *Ann. d. Phys.*, 35, p. 536, 1911.

Whereas the first three equations of motion express the law of conservation of momentum, the last equation expresses the law of conservation of energy. We set

$$\mathfrak{S}^e = c^2\mathfrak{g}^e, \tag{17}$$

$$\mathfrak{S}^m = c^2\mathfrak{g}^m = \frac{c^2\mu\mathfrak{v}}{\sqrt{1-\mathfrak{q}^2}}, \tag{18}$$

hence

$$\mathfrak{S}_x{}^e = -ic\boldsymbol{p}_{xu} \qquad \text{etc.,} \tag{17a}$$

$$\mathfrak{S}_x{}^m = -ic\nu\mathfrak{V}_x\mathfrak{V}_u \qquad \text{etc.,} \tag{18a}$$

| and furthermore [863]

$$\psi^e = -\boldsymbol{p}_{uu}, \tag{19}$$

$$\psi^m = -\nu\mathfrak{V}_u{}^2 = \frac{c^2\mu}{\sqrt{1-\mathfrak{q}^2}}. \tag{20}$$

Equation (5) yields an expression for $ic\mathfrak{K}_u{}^e$, which is, written in vector form,

$$ic\mathfrak{K}_u{}^e = \operatorname{div}\mathfrak{S}^e + \frac{\partial\psi^e}{\partial t}. \tag{21}$$

We can now write the last of the equations of motion (6), multiplied by $-ic$, as

$$-ic\mathfrak{K}_u = \operatorname{div}(\mathfrak{S}^e + \mathfrak{S}^m) + \frac{\partial}{\partial t}(\psi^e + \psi^m). \tag{22}$$

This is the equation expressing the conservation of energy. We recognize that the vectors $\mathfrak{S}^e$ and $\mathfrak{S}^m$ express the *elastic* and the *material energy flux* respectively and that the quantities ψ^e and ψ^m are the corresponding *energy densities*. The quantity $-ic\mathfrak{K}_u$ gives the energy influx per unit volume and time produced by the action of the external force $\mathfrak{K}$. The meaning of the right-hand side is obvious. By integrating over an arbitrary volume and applying Gauss's theorem, one obtains the law of conservation of energy, expressed for a spatial region, fixed in the (x, y, z) reference frame employed.

2. DETAILED INVESTIGATION OF THE ELASTIC STATE VARIABLES

In order to gain a clear conception of the elastic variables, we transform the stress tensor $\boldsymbol{p}$ to rest at the spacetime point under consideration. Then the components of the tensor take the form

$$\left.\begin{array}{cccc} \boldsymbol{p}^0_{xx} & \boldsymbol{p}^0_{xy} & \boldsymbol{p}^0_{xz} & 0, \\ \boldsymbol{p}^0_{yx} & \boldsymbol{p}^0_{yy} & \boldsymbol{p}^0_{yz} & 0, \\ \boldsymbol{p}^0_{zx} & \boldsymbol{p}^0_{zy} & \boldsymbol{p}^0_{zz} & 0, \\ 0 & 0 & 0 & \boldsymbol{p}^0_{uu}; \end{array}\right\} \tag{23}$$

because in the case of rest the elastic state of stress can produce no energy flux and thus also no momentum.[9] |

[864] We note incidentally that for the case of rest the usual laws of the theory of elasticity apply. We can thus relate the six spatial stress components $\boldsymbol{p}^0_{xx}, \boldsymbol{p}^0_{xy}, \ldots$ to the deformation quantities at rest.[10] However, we do not want to discuss this in more detail.

One easily sees that $\boldsymbol{p}^0_{uu}$ must be a four-dimensional scalar. It is

$$\begin{aligned} -c^2\boldsymbol{p}^0_{uu} = {} & \boldsymbol{p}_{xx}\mathfrak{B}_x{}^2 + \boldsymbol{p}_{yy}\mathfrak{B}_y{}^2 + \boldsymbol{p}_{zz}\mathfrak{B}_z{}^2 + \boldsymbol{p}_{uu}\mathfrak{B}_u{}^2 \\ & + 2\boldsymbol{p}_{xy}\mathfrak{B}_x\mathfrak{B}_y + 2\boldsymbol{p}_{xz}\mathfrak{B}_x\mathfrak{B}_z + \ldots, \end{aligned}$$

since the right-hand side is invariant under Lorentz transformations, and upon transformation to rest, nine of the ten terms disappear, whereby one obtains the identity $-c^2\boldsymbol{p}^0_{uu} = -c^2\boldsymbol{p}^0_{uu}$.

Furthermore, by (19) and (20)

$$\Psi = -\boldsymbol{p}^0_{uu} + c^2\nu \tag{24}$$

is the *rest-energy density* of the matter, which is also a four-dimensional scalar. Since ν is still undetermined, one could define this quantity in such a way that $\boldsymbol{p}^0_{uu} = 0$. However, for the time being we do not want to impose such a condition.

By means of this transformation to rest, one also recognizes the validity of the following system of equations:

$$\left.\begin{aligned} \boldsymbol{p}_{xx}\mathfrak{B}_x + \boldsymbol{p}_{xy}\mathfrak{B}_y + \boldsymbol{p}_{xz}\mathfrak{B}_z + \boldsymbol{p}_{xu}\mathfrak{B}_u &= \boldsymbol{p}^0_{uu}\mathfrak{B}_x, \\ \boldsymbol{p}_{yx}\mathfrak{B}_x + \boldsymbol{p}_{yy}\mathfrak{B}_y + \boldsymbol{p}_{yz}\mathfrak{B}_z + \boldsymbol{p}_{yu}\mathfrak{B}_u &= \boldsymbol{p}^0_{uu}\mathfrak{B}_y, \\ \boldsymbol{p}_{zx}\mathfrak{B}_x + \boldsymbol{p}_{zy}\mathfrak{B}_y + \boldsymbol{p}_{zz}\mathfrak{B}_z + \boldsymbol{p}_{zu}\mathfrak{B}_u &= \boldsymbol{p}^0_{uu}\mathfrak{B}_z, \\ \boldsymbol{p}_{ux}\mathfrak{B}_x + \boldsymbol{p}_{uy}\mathfrak{B}_y + \boldsymbol{p}_{uz}\mathfrak{B}_z + \boldsymbol{p}_{uu}\mathfrak{B}_u &= \boldsymbol{p}^0_{uu}\mathfrak{B}_u. \end{aligned}\right\} \tag{25}$$

9 If heat flow occurs, its influence is to be included in the action of the external force $\mathfrak{K}$. Compare section 5 below.

10 Cf. Herglotz, loc. cit.

From the first three of these equations we obtain the expressions for the components of the elastic energy flux and the momentum density. Namely, using eqs. (4) we find that

$$-ic\boldsymbol{p}_{ux} = -\overset{0}{\boldsymbol{p}}_{uu}\mathfrak{v}_x + \boldsymbol{p}_{xx}\mathfrak{v}_x + \boldsymbol{p}_{xy}\mathfrak{v}_y + \boldsymbol{p}_{xz}\mathfrak{v}_z,$$

hence

$$\left.\begin{aligned}
\mathfrak{S}_x^{\;e} &= c^2\mathfrak{g}_x^{\;e} = -\overset{0}{\boldsymbol{p}}_{uu}\mathfrak{v}_x + \boldsymbol{p}_{xx}\mathfrak{v}_x + \boldsymbol{p}_{xy}\mathfrak{v}_y + \boldsymbol{p}_{xz}\mathfrak{v}_z,\\
\mathfrak{S}_y^{\;e} &= c^2\mathfrak{g}_y^{\;e} = -\overset{0}{\boldsymbol{p}}_{uu}\mathfrak{v}_y + \boldsymbol{p}_{yx}\mathfrak{v}_x + \boldsymbol{p}_{yy}\mathfrak{v}_y + \boldsymbol{p}_{yz}\mathfrak{v}_z,\\
\mathfrak{S}_z^{\;e} &= c^2\mathfrak{g}_z^{\;e} = -\overset{0}{\boldsymbol{p}}_{uu}\mathfrak{v}_z + \boldsymbol{p}_{zx}\mathfrak{v}_x + \boldsymbol{p}_{zy}\mathfrak{v}_y + \boldsymbol{p}_{zz}\mathfrak{v}_z.
\end{aligned}\right\} \tag{26}$$

| We can also express these vector components in terms of the relative stresses, by [865]
using (12a) to eliminate $\boldsymbol{p}$. We obtain

$$\mathfrak{S}_x^{\;e}(1-\mathfrak{q}^2) = -\overset{0}{\boldsymbol{p}}_{uu}\mathfrak{v}_x + \boldsymbol{t}_{xx}\mathfrak{v}_x + \boldsymbol{t}_{xy}\mathfrak{v}_y + \boldsymbol{t}_{xz}\mathfrak{v}_z, \tag{26a}$$

and the corresponding expressions for the two remaining components.

The following expression for the elastic energy density arises from the last of eqs. (25):

$$\psi^e = -\overset{0}{\boldsymbol{p}}_{uu} + \mathfrak{g}^e\mathfrak{v}. \tag{27}$$

3. THE CHANGES OF THE MASS AND OF THE REST-ENERGY

In order to obtain the law describing the variability of mass, we multiply the eqs. (6) in turn by $\mathfrak{B}_x$, $\mathfrak{B}_y$, $\mathfrak{B}_z$, $\mathfrak{B}_u$ and add them, where we take into account that

$$\begin{aligned}
&\mathfrak{B}_x\left\{\frac{\partial}{\partial x}\nu\mathfrak{B}_x^{\;2} + \frac{\partial}{\partial y}\nu\mathfrak{B}_x\mathfrak{B}_y + \frac{\partial}{\partial z}\nu\mathfrak{B}_x\mathfrak{B}_z + \frac{\partial}{\partial u}\nu\mathfrak{B}_x\mathfrak{B}_u\right\}\\
&= \mathfrak{B}_x^{\;2}\left\{\frac{\partial}{\partial x}\nu\mathfrak{B}_x + \frac{\partial}{\partial y}\nu\mathfrak{B}_y + \frac{\partial}{\partial z}\nu\mathfrak{B}_z + \frac{\partial}{\partial u}\nu\mathfrak{B}_u\right\}\\
&\quad + \frac{1}{2}\nu\left\{\mathfrak{B}_x\frac{\partial\mathfrak{B}_x^{\;2}}{\partial x} + \mathfrak{B}_y\frac{\partial\mathfrak{B}_x^{\;2}}{\partial y} + \mathfrak{B}_z\frac{\partial\mathfrak{B}_x^{\;2}}{\partial z} + \mathfrak{B}_u\frac{\partial\mathfrak{B}_x^{\;2}}{\partial u}\right\} \quad \text{etc.}
\end{aligned}$$

Since furthermore, according to the principles of the theory of relativity,

$$\mathfrak{B}_x^{\;2} + \mathfrak{B}_y^{\;2} + \mathfrak{B}_z^{\;2} + \mathfrak{B}_u^{\;2} = -c^2, \tag{28}$$

we obtain

$$\left.\begin{aligned}&\mathfrak{B}_x(\mathfrak{K}_x+\mathfrak{K}_x{}^e)+\mathfrak{B}_y(\mathfrak{K}_y+\mathfrak{K}_y{}^e)+\mathfrak{B}_z(\mathfrak{K}_z+\mathfrak{K}_z{}^e)+\mathfrak{B}_u(\mathfrak{K}_u+\mathfrak{K}_u{}^e)\\&\qquad\qquad = -c^2\left\{\frac{\partial}{\partial x}\nu\mathfrak{B}_x+\frac{\partial}{\partial y}\nu\mathfrak{B}_y+\frac{\partial}{\partial z}\nu\mathfrak{B}_z+\frac{\partial}{\partial u}\nu\mathfrak{B}_u\right\}.\end{aligned}\right\} \tag{29}$$

This equation gives the variability of mass with respect to time, because if dv_0 designates the rest volume of a material particle, then (compare eq. (9))

$$\frac{\partial}{\partial x}\nu\mathfrak{B}_x+\frac{\partial}{\partial y}\nu\mathfrak{B}_y+\frac{\partial}{\partial z}\nu\mathfrak{B}_z+\frac{\partial}{\partial u}\nu\mathfrak{B}_u = \frac{1}{d\tau_0}\frac{d}{d\tau}(\nu dv_0), \tag{30}$$

where $\nu dv_0 = \mu dv$ is the mass of the particle.

Therefore, if the sum of the external and elastic forces are perpendicular to the velocity vector $\mathfrak{B}$, the mass of the matter is constant in time, but otherwise it is not. |

[866] Further, we want to develop a formula for the elastic force, and for this purpose differentiate the equations (25) with respect to x, y, z, u and add. Simply converting terms, taking (5) into consideration, we thus obtain

$$\left.\begin{aligned}&\mathfrak{B}_x\mathfrak{K}_x{}^e+\mathfrak{B}_y\mathfrak{K}_y{}^e+\mathfrak{B}_z\mathfrak{K}_z{}^e+\mathfrak{B}_u\mathfrak{K}_u{}^e\\&= \boldsymbol{p}_{xx}\frac{\partial\mathfrak{B}_x}{\partial x}+\boldsymbol{p}_{yy}\frac{\partial\mathfrak{B}_y}{\partial y}+\boldsymbol{p}_{zz}\frac{\partial\mathfrak{B}_z}{\partial z}+\boldsymbol{p}_{uu}\frac{\partial\mathfrak{B}_u}{\partial u}\\&\quad+\boldsymbol{p}_{xy}\left\{\frac{\partial\mathfrak{B}_x}{\partial y}+\frac{\partial\mathfrak{B}_y}{\partial x}\right\}+\boldsymbol{p}_{xz}\left\{\frac{\partial\mathfrak{B}_x}{\partial z}+\frac{\partial\mathfrak{B}_z}{\partial x}\right\}\\&\quad+\boldsymbol{p}_{xu}\{..\}+\boldsymbol{p}_{yz}\{..\}+\boldsymbol{p}_{yu}\{..\}+\boldsymbol{p}_{zu}\{..\}\\&\quad-\left\{\frac{\partial}{\partial x}\boldsymbol{p}^0_{uu}\mathfrak{B}_x+\frac{\partial}{\partial y}\boldsymbol{p}^0_{uu}\mathfrak{B}_y+\frac{\partial}{\partial z}\boldsymbol{p}^0_{uu}\mathfrak{B}_z+\frac{\partial}{\partial u}\boldsymbol{p}^0_{uu}\mathfrak{B}_u\right\}.\end{aligned}\right\} \tag{31}$$

Subtracting (29) from (31), taking (24) into account, one obtains

$$\left.\begin{aligned}
&\frac{\partial}{\partial x}\Psi\mathfrak{B}_x+\frac{\partial}{\partial y}\Psi\mathfrak{B}_y+\frac{\partial}{\partial z}\Psi\mathfrak{B}_z+\frac{\partial}{\partial u}\Psi\mathfrak{B}_u\\
&=-\{\mathfrak{B}_x\mathfrak{K}_x+\mathfrak{B}_y\mathfrak{K}_y+\mathfrak{B}_z\mathfrak{K}_z+\mathfrak{B}_u\mathfrak{K}_u\}\\
&\quad-\boldsymbol{p}_{xx}\frac{\partial\mathfrak{B}_x}{\partial x}-\boldsymbol{p}_{yy}\frac{\partial\mathfrak{B}_y}{\partial y}-\boldsymbol{p}_{zz}\frac{\partial\mathfrak{B}_z}{\partial z}-\boldsymbol{p}_{uu}\frac{\partial\mathfrak{B}_u}{\partial u}\\
&\quad-\boldsymbol{p}_{xy}\left\{\frac{\partial\mathfrak{B}_x}{\partial y}+\frac{\partial\mathfrak{B}_y}{\partial x}\right\}-\boldsymbol{p}_{xz}\left\{\frac{\partial\mathfrak{B}_x}{\partial z}+\frac{\partial\mathfrak{B}_z}{\partial x}\right\}\\
&\quad-\boldsymbol{p}_{xu}\{\ldots\}-\boldsymbol{p}_{yz}\{\ldots\}-\boldsymbol{p}_{yu}\{\ldots\}-\boldsymbol{p}_{zu}\{\ldots\}.
\end{aligned}\right\}\tag{32}$$

This equation expresses the law of conservation of energy for a rest volume carried along by the matter, in contrast to eq. (22), which refers to a unit volume fixed in the spatial coordinate system used. Equation (32) is completely symmetric with respect to x, y, z, u; that it really expresses the law of conservation of energy is easily seen by transforming to rest.

4. THE DEFINITION OF INERTIAL MASS

Until now, we have taken the rest-mass density to be a completely arbitrary function of the four coordinates of the spacetime points of matter. Now we want to end this indeterminacy, | and for that reason we will focus on various possibilities for the [867]
moment.

In eq. (24),

$$\Psi = -\boldsymbol{p}^0_{uu}+c^2\nu$$

the rest-energy density Ψ is a defined quantity, one of the quantities $\boldsymbol{p}^0_{uu}$ and ν, however, can be freely specified. We demand that $\boldsymbol{p}^0_{uu}$ becomes zero if no elastic stresses exist in the body under consideration, because the tensor $\boldsymbol{p}$ should represent the elastic state of stress and only that state. Therefore, if all spatial components of $\boldsymbol{p}$ are zero when transformed to the state of rest (schema (23)), then $\boldsymbol{p}^0_{uu}$ should also be zero. But this can be achieved in various ways.

If one restricts attention to bodies in which there is a normal pressure from all sides, one can easily define the rest density ν in such a way that (if no heat conduction takes place) the total inertia of the body will be determined by its mass. To do so, one only has to set

$$\begin{aligned}
\boldsymbol{p}^0_{xx} &= \boldsymbol{p}^0_{yy} = \boldsymbol{p}^0_{zz} = \boldsymbol{p}^0_{uu},\\
0 &= \boldsymbol{p}^0_{xy} = \boldsymbol{p}^0_{xz} = \boldsymbol{p}^0_{xu} = \ldots,
\end{aligned}$$

in the schema of (23),[11] from which ν is determined according to (24).

Then the stress tensor $\boldsymbol{p}$ has degenerated into a scalar and the elastic momentum density $\mathfrak{g}^e$ becomes equal to zero, independent of the state of motion.

However, as far as I am aware, this conception of mass cannot be naturally extended to the general case of bodies in which (relative) tangential stresses also exist. For the general case, it appears to me that the simplest and most expedient definition lies in the stipulation

$$c^2\nu = \Psi. \tag{33}$$

The rest-mass density is thus set proportional to the rest-energy density. Then, according to (24)

$$\boldsymbol{p}^0_{uu} = 0, \tag{34}$$

[868] | and several of our previous equations become simpler as a result.

Of course, the factual content of relativistic mechanics remains completely unaffected by the way in which we define inertial mass. Our definition gains more than a merely formal meaning only later, in section §6, when we also attribute weight to the inertial mass.

Having now fixed the concept of mass through the definition (33), we have to note that each moving and elastically stressed body possesses a momentum $\mathfrak{g}^e$ that is determined not by the mass but by the state of elastic stress of the body. According to (26) and (26a) we obtain

$$\left.\begin{aligned} \mathfrak{g}_x{}^e &= \frac{1}{c^2}(\boldsymbol{p}_{xx}\mathfrak{v}_x + \boldsymbol{p}_{xy}\mathfrak{v}_y + \boldsymbol{p}_{xz}\mathfrak{v}_z) \\ &= \frac{1}{c^2(1-\mathfrak{q}^2)}(\boldsymbol{t}_{xx}\mathfrak{v}_x + \boldsymbol{t}_{xy}\mathfrak{v}_y + \boldsymbol{t}_{xz}\mathfrak{v}_z) \quad \text{etc.} \end{aligned}\right\} \tag{35}$$

This momentum also appears when there is a normal pressure from all sides in the body under consideration. In this case, $\mathfrak{g}^e$ can be derived from a *virtual* inertial mass which is added to that defined by eq. (33).

5. THE INFLUENCE OF HEAT CONDUCTION

All the equations developed above are also valid if heat conduction takes place in the bodies considered, because the effects of the heat conduction can be attributed to a ponderomotive force $\mathfrak{K}^w$ appearing in the heat conduction field, and this force can be included in the external force $\mathfrak{K}$. Of the heat conduction force $\mathfrak{K}^w$ the energy component $\mathfrak{K}_u{}^w$ plays the essential role. According to our fundamental assumptions, like all ponderomotive forces, $\mathfrak{K}^w$ should also be derivable from a symmetric four-dimensional tensor. We denote the heat conductivity tensor by $\boldsymbol{w}$, and thus have

11 G. Nordström, *Physik. Zeitschr.*, 12. p. 854, 1911; M. Laue, *Das Relativitätsprinzip,* p. 151.

$$\left.\begin{aligned}\mathfrak{K}_x{}^w &= -\frac{\partial \boldsymbol{w}_{xx}}{\partial x}-\frac{\partial \boldsymbol{w}_{xy}}{\partial y}-\frac{\partial \boldsymbol{w}_{xz}}{\partial z}-\frac{\partial \boldsymbol{w}_{xu}}{\partial u},\\ &\cdots\cdots\cdots\cdots\\ &\cdots\cdots\cdots\cdots\\ \mathfrak{K}_u{}^w &= -\frac{\partial \boldsymbol{w}_{ux}}{\partial x}-\frac{\partial \boldsymbol{w}_{uy}}{\partial y}-\frac{\partial \boldsymbol{w}_{uz}}{\partial z}-\frac{\partial \boldsymbol{w}_{uu}}{\partial u},\end{aligned}\right\} \tag{36}$$

where, e. g., [869]

$$\boldsymbol{w}_{xu} = \boldsymbol{w}_{ux} \quad \text{etc.}$$

For the case of rest, the tensor $\boldsymbol{w}$ takes the following schema:

$$\left.\begin{matrix} 0 & 0 & 0 & \boldsymbol{w}^0_{xu}, \\ 0 & 0 & 0 & \boldsymbol{w}^0_{yu}, \\ 0 & 0 & 0 & \boldsymbol{w}^0_{zu}, \\ \boldsymbol{w}^0_{ux} & \boldsymbol{w}^0_{uy} & \boldsymbol{w}^0_{uz} & 0, \end{matrix}\right\} \tag{37}$$

because in the state of rest, all the spatial stresses are given by the tensor $\boldsymbol{p}$ (schema 23) and the total energy density of the matter is given by $\Psi = c^2\nu$. Therefore, the real components of $\boldsymbol{w}$ must be zero for the state of rest.

If we take heat conduction into account, then we have three four-dimensional tensors pertaining to matter: the thermal conductivity tensor, the elastic tensor, and the material tensor. For the case of rest, all three can be combined in the following common schema:

$$\left.\begin{array}{ccc|c} \boldsymbol{p}^0_{xx} & \boldsymbol{p}^0_{xy} & \boldsymbol{p}^0_{xz} & \boldsymbol{w}^0_{xu} \\ \boldsymbol{p}^0_{yx} & \boldsymbol{p}^0_{yy} & \boldsymbol{p}^0_{yz} & \boldsymbol{w}^0_{yu} \\ \boldsymbol{p}^0_{zx} & \boldsymbol{p}^0_{zy} & \boldsymbol{p}^0_{zz} & \boldsymbol{w}^0_{zu} \\ \hline \boldsymbol{w}^0_{ux} & \boldsymbol{w}^0_{uy} & \boldsymbol{w}^0_{uz} & -c^2\nu \end{array}\right\}. \tag{38}$$

Because we made the stipulation (33), the decomposition of the total tensor into the three parts is unambiguous.

We can introduce a four-vector $\mathfrak{W}$ by the system of equations

$$\left.\begin{array}{l} \boldsymbol{W}_{xx}\mathfrak{V}_x + \boldsymbol{W}_{xy}\mathfrak{V}_y + \boldsymbol{W}_{xz}\mathfrak{V}_z + \boldsymbol{W}_{xu}\mathfrak{V}_u = -\mathfrak{W}_x, \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ \boldsymbol{W}_{ux}\mathfrak{V}_x + \boldsymbol{W}_{uy}\mathfrak{V}_y + \boldsymbol{W}_{uz}\mathfrak{V}_z + \boldsymbol{W}_{uu}\mathfrak{V}_u = -\mathfrak{W}_u. \end{array}\right\} \tag{39}$$

As we are going to show, this vector is to be designated the *rest heat flow*. The law of energy conservation for heat conduction (compare equation 21) is expressed by the equation

$$ic\mathfrak{K}_u^w = -ic\left\{\frac{\partial \boldsymbol{W}_{ux}}{\partial x} + \frac{\partial \boldsymbol{W}_{uy}}{\partial y} + \frac{\partial \boldsymbol{W}_{uz}}{\partial z} + \frac{\partial \boldsymbol{W}_{uu}}{\partial u}\right\},$$

[870] | and for the case of rest one obtains from (39)

$$\begin{aligned} ic\boldsymbol{W}_{xu}^0 &= -\mathfrak{W}_x^0, \\ ic\boldsymbol{W}_{yu}^0 &= -\mathfrak{W}_y^0, \\ ic\boldsymbol{W}_{zu}^0 &= -\mathfrak{W}_z^0, \\ 0 &= -\mathfrak{W}_u^0, \end{aligned}$$

from which the asserted meaning of the vector $\mathfrak{W}$ becomes clear.

From the last equations one also sees that the four-vector $\mathfrak{W}$ is orthogonal to the velocity vector $\mathfrak{V}$, so that

$$\mathfrak{W}_x\mathfrak{V}_x + \mathfrak{W}_y\mathfrak{V}_y + \mathfrak{W}_z\mathfrak{V}_z + \mathfrak{W}_u\mathfrak{V}_u = 0, \tag{40}$$

since the left hand side of this equation is invariant under Lorentz transformations, and it equals zero when transformed to rest.

The tensor $\boldsymbol{W}$ can also be expressed as the "tensor product"[12] of the two four-vectors $\mathfrak{W}$ and $\mathfrak{V}$. Transforming to the state of rest, one finds the following expressions for the components of $\boldsymbol{W}$

12 Cf. W. Voigt, *Gött. Nachr.*, p. 500, 1904.

$$\left.\begin{aligned}\boldsymbol{w}_{xx} &= \frac{2}{c^2}\mathfrak{W}_x\mathfrak{V}_x,\\ \boldsymbol{w}_{uu} &= \frac{2}{c^2}\mathfrak{W}_u\mathfrak{V}_u,\\ \boldsymbol{w}_{xy} &= \frac{1}{c^2}\{\mathfrak{W}_x\mathfrak{V}_y+\mathfrak{W}_y\mathfrak{V}_x\},\\ \boldsymbol{w}_{xu} &= \frac{1}{c^2}\{\mathfrak{W}_x\mathfrak{V}_u+\mathfrak{W}_u\mathfrak{V}_x\},\\ &\text{etc.}\end{aligned}\right\} \tag{41}$$

For the energy density ψ^w and for the energy flux $\mathfrak{S}^w$ of the heat conduction field one has of course

$$\psi^w = -\boldsymbol{w}_{uu}, \tag{42}$$

$$\mathfrak{S}_x^w = -ic\boldsymbol{w}_{xu}, \qquad \mathfrak{S}_y^w = -ic\boldsymbol{w}_{yu}, \qquad \mathfrak{S}_z^w = -ic\boldsymbol{w}_{zu}. \tag{43}$$

These quantities can also be expressed using the vector $\mathfrak{W}$. First, we find from (40)

$$-ic\mathfrak{W}_u = \mathfrak{W}\mathfrak{v}, \tag{44}$$

| where the right-hand side is the scalar product of two three-dimensional vectors. [871] Furthermore, from (41) we obtain

$$\psi^w = \frac{2}{c^2\sqrt{1-\mathfrak{q}^2}}\mathfrak{W}\mathfrak{v}, \tag{42a}$$

$$\left.\begin{aligned}\mathfrak{S}^w &= \frac{1}{\sqrt{1-\mathfrak{q}^2}}\left\{\mathfrak{W}+\frac{\mathfrak{v}}{c^2}(\mathfrak{W}\mathfrak{v})\right\}\\ &= \frac{\mathfrak{W}}{\sqrt{1-\mathfrak{q}^2}}+\mathfrak{v}\frac{\psi^w}{2}.\end{aligned}\right\} \tag{43a}$$

The energy flux $\mathfrak{S}^w$ corresponds, of course, to the momentum density

$$\mathfrak{g}^w = \frac{1}{c^2}\mathfrak{S}^w.$$

We can write the last of the equations (36), multiplied by ic, as

$$ic\mathfrak{K}_u^w = \operatorname{div}\mathfrak{S}^w+\frac{\partial\psi^w}{\partial t}, \tag{45}$$

which can be inserted into the energy eq. (22). If $\mathfrak{K}^w$ is the only "external" force acting, naturally one has to set $\mathfrak{K} = \mathfrak{K}^w$ in all the equations of the previous sections, and thus specifically in (22) $\mathfrak{K}_u = \mathfrak{K}_u{}^w$.

We further wish to develop a few formulas for $\mathfrak{K}^w$. Inserting the expressions (41) into the system of equations (36), we obtain after simple transformations[13]

$$\left.\begin{aligned} c^2\mathfrak{K}_x{}^w = &-\frac{d\mathfrak{W}_x}{d\tau} - \mathfrak{V}_x\left\{\frac{\partial\mathfrak{W}_x}{\partial x} + \frac{\partial\mathfrak{W}_y}{\partial y} + \frac{\partial\mathfrak{W}_z}{\partial z} + \frac{\partial\mathfrak{W}_u}{\partial u}\right\} \\ &-\left\{\mathfrak{W}_x\frac{\partial\mathfrak{V}_x}{\partial x} + \mathfrak{W}_y\frac{\partial\mathfrak{V}_x}{\partial y} + \mathfrak{W}_z\frac{\partial\mathfrak{V}_x}{\partial z} + \mathfrak{W}_u\frac{\partial\mathfrak{V}_x}{\partial u}\right\} \\ &-\mathfrak{W}_x\left\{\frac{\partial\mathfrak{V}_x}{\partial x} + \frac{\partial\mathfrak{V}_y}{\partial y} + \frac{\partial\mathfrak{V}_z}{\partial z} + \frac{\partial\mathfrak{V}_u}{\partial u}\right\}, \end{aligned}\right\} \quad (46)$$

and corresponding expressions for the remaining components of $\mathfrak{K}^w$. Multiplying
[872] these expressions by $\mathfrak{V}_x, \mathfrak{V}_y, \mathfrak{V}_z, \mathfrak{V}_u$ | and adding them, we obtain further, in light of (40) and (28),

$$\left.\begin{aligned} &\mathfrak{V}_x\mathfrak{K}_x{}^w + \mathfrak{V}_y\mathfrak{K}_y{}^w + \mathfrak{V}_z\mathfrak{K}_z{}^w + \mathfrak{V}_u\mathfrak{K}_u{}^w \\ &= \frac{\partial\mathfrak{W}_x}{\partial x} + \frac{\partial\mathfrak{W}_y}{\partial y} + \frac{\partial\mathfrak{W}_z}{\partial z} + \frac{\partial\mathfrak{W}_u}{\partial u} \\ &+ \frac{1}{c^2}\left\{\mathfrak{W}_x\frac{d\mathfrak{V}_x}{d\tau} + \mathfrak{W}_y\frac{d\mathfrak{V}_y}{d\tau} + \mathfrak{W}_z\frac{d\mathfrak{V}_z}{d\tau} + \mathfrak{W}_u\frac{d\mathfrak{V}_u}{d\tau}\right\}. \end{aligned}\right\} \quad (47)$$

This equation makes it possible to take heat conduction into account in the formulae (29) and (32).

13 If φ is an arbitrary function of the four coordinates, then

$$\frac{d\varphi}{d\tau} = \mathfrak{V}_x\frac{\partial\varphi}{\partial x} + \mathfrak{V}_y\frac{\partial\varphi}{\partial y} + \mathfrak{V}_z\frac{\partial\varphi}{\partial z} + \mathfrak{V}_u\frac{\partial\varphi}{\partial u}.$$

6. GRAVITATION

Various approaches to treat the phenomena of gravitation from the standpoint of relativity theory have been attempted. The theories of Einstein[14] and Abraham[15] are especially noteworthy. According to these two theories, however, the speed of light c would depend on the gravitational field rather than being constant, and this would require at least a complete revolution of the foundations of the present theory of relativity.

However, through a modification of the theory of Abraham one can, as I have shown elsewhere,[16] maintain the constancy of the speed of light, and develop a theory of gravitation which is compatible with the theory of relativity in its present form. Since I want to generalize this theory in one respect, its foundations are briefly recounted here.

I introduce the gravitational potential Φ and set, using rational units,

$$\frac{\partial^2\Phi}{\partial x^2}+\frac{\partial^2\Phi}{\partial y^2}+\frac{\partial^2\Phi}{\partial z^2}+\frac{\partial^2\Phi}{\partial u^2} = g\nu. \tag{48}$$

Here, ν is the rest density of matter as defined in eq. (33). The gravitational potential Φ and the quantity g are also four-dimensional scalars; we call g the *gravitational factor.* |

The gravitational field exerts forces on the bodies present within the field. For the ponderomotive gravitational force $\mathfrak{K}^g$ per unit volume, I set [873]

$$\mathfrak{K}_x{}^g = -g\nu\frac{\partial\Phi}{\partial x},\quad \mathfrak{K}_y{}^g = -g\nu\frac{\partial\Phi}{\partial y},\quad \mathfrak{K}_z{}^g = -g\nu\frac{\partial\Phi}{\partial z},\quad \mathfrak{K}_u{}^g = -g\nu\frac{\partial\Phi}{\partial u}. \tag{49}$$

The equations (48) and (49), along with the principle of constant c,

$$c = \text{a universal constant}, \tag{50}$$

constitute the complete basis of my theory of gravitation. These equations also determine the rational units of Φ and g. For the time being, we consider the gravitational factor g to be a universal constant, but here I would like to remark that since g occurs only as a factor of ν, nothing prevents us from assuming that g depends upon the internal state of matter.

The fundamental equations (48), (49) and (50) demand that the mass of a material particle depends on the gravitational potential at its location. In order to obtain the law of this dependence, accordingly we consider the motion of a mass point of mass m in an arbitrary gravitational field. We assume that no forces except gravitation act on the mass point. Then we can write the equations of motion of the mass point in the following way (compare eq. (6) and (9))

---

14 A. Einstein, *Ann. d. Phys.*, 35, p. 898, 1911.

15 M. Abraham, *Physik. Zeitschr.*, 13, p. 1, 1912 [in this volume].

16 G. Nordström, *Physik. Zeitschr.*, 13, p. 1126, 1912 [in this volume].

$$\left.\begin{aligned}
-gm\frac{\partial\Phi}{\partial x} &= m\frac{d\mathfrak{V}_x}{d\tau}+\mathfrak{V}_x\frac{dm}{d\tau},\\
-gm\frac{\partial\Phi}{\partial y} &= m\frac{d\mathfrak{V}_y}{d\tau}+\mathfrak{V}_y\frac{dm}{d\tau},\\
-gm\frac{\partial\Phi}{\partial z} &= m\frac{d\mathfrak{V}_z}{d\tau}+\mathfrak{V}_z\frac{dm}{d\tau},\\
-gm\frac{\partial\Phi}{\partial u} &= m\frac{d\mathfrak{V}_u}{d\tau}+\mathfrak{V}_u\frac{dm}{d\tau}.
\end{aligned}\right\}\tag{51}$$

We multiply the equations in turn by $\mathfrak{V}_x, \mathfrak{V}_y, \mathfrak{V}_z, \mathfrak{V}_u$ and add them. Taking (28) into account, due to

$$\frac{d\Phi}{d\tau} = \mathfrak{V}_x\frac{\partial\Phi}{\partial x}+\mathfrak{V}_y\frac{\partial\Phi}{\partial y}+\mathfrak{V}_z\frac{\partial\Phi}{\partial z}+\mathfrak{V}_u\frac{\partial\Phi}{\partial u},$$

we obtain

$$-gm\frac{d\Phi}{d\tau} = -c^2\frac{dm}{d\tau},$$

[874] | or

$$\frac{1}{m}\frac{dm}{d\tau} = \frac{g}{c^2}\frac{d\Phi}{d\tau}.\tag{52}$$

If g is assumed to be constant, integration yields

$$m = m_0 e^{\frac{g}{c^2}\Phi},\tag{53}$$

and this equation gives the dependence of the mass on the gravitational potential.

The equations of motion can also be written in the following form, from (52):

$$\left.\begin{aligned}
-g\frac{\partial\Phi}{\partial x} &= \frac{d\mathfrak{V}_x}{d\tau}+\frac{g}{c^2}\mathfrak{V}_x\frac{d\Phi}{d\tau},\\
-g\frac{\partial\Phi}{\partial y} &= \frac{d\mathfrak{V}_y}{d\tau}+\frac{g}{c^2}\mathfrak{V}_y\frac{d\Phi}{d\tau},\\
-g\frac{\partial\Phi}{\partial z} &= \frac{d\mathfrak{V}_z}{d\tau}+\frac{g}{c^2}\mathfrak{V}_z\frac{d\Phi}{d\tau},\\
-g\frac{\partial\Phi}{\partial u} &= \frac{d\mathfrak{V}_u}{d\tau}+\frac{g}{c^2}\mathfrak{V}_u\frac{d\Phi}{d\tau},
\end{aligned}\right\}\tag{54}$$

whereby the mass m cancels out of the equations of motion.

The reason for the variability of the mass m is that the gravitational force $\mathfrak{K}^g$ is not orthogonal to the velocity vector $\mathfrak{V}$ (compare p. 507 [p. 865 in original]). Multiplying the equations (49) by $\mathfrak{V}_x$, $\mathfrak{V}_y$, $\mathfrak{V}_z$, $\mathfrak{V}_u$ and adding them, we obtain

$$\mathfrak{V}_x\mathfrak{K}_x{}^g + \mathfrak{V}_y\mathfrak{K}_y{}^g + \mathfrak{V}_z\mathfrak{K}_z{}^g + \mathfrak{V}_u\mathfrak{K}_u{}^g = -g\nu\frac{d\Phi}{d\tau}. \tag{55}$$

We can insert this expression into the eq. (29) for the change of the mass; the gravitational force $\mathfrak{K}^g$ is, of course, part of the "external" force $\mathfrak{K}$.

The gravitational force $\mathfrak{K}^g$ is derived from a symmetric four-dimensional tensor $\boldsymbol{G}$, in that

$$\left.\begin{array}{l} \mathfrak{K}_x{}^g = -\dfrac{\partial \boldsymbol{G}_{xx}}{\partial x} - \dfrac{\partial \boldsymbol{G}_{xy}}{\partial y} - \dfrac{\partial \boldsymbol{G}_{xz}}{\partial z} - \dfrac{\partial \boldsymbol{G}_{xu}}{\partial u}, \\ \cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot \\ \cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot \\ \mathfrak{K}_u{}^g = -\dfrac{\partial \boldsymbol{G}_{ux}}{\partial x} - \dfrac{\partial \boldsymbol{G}_{uy}}{\partial y} - \dfrac{\partial \boldsymbol{G}_{uz}}{\partial z} - \dfrac{\partial \boldsymbol{G}_{uu}}{\partial u}. \end{array}\right\} \tag{56}$$

One obtains equations of this form by inserting the expression (48) for $g\nu$ into (49), and then performing a further | transformation. Then one also finds the following expressions for the tensor components:[17] [875]

$$\left.\begin{array}{l} \boldsymbol{G}_{xx} = \dfrac{1}{2}\left\{\left(\dfrac{\partial\Phi}{\partial x}\right)^2 - \left(\dfrac{\partial\Phi}{\partial y}\right)^2 - \left(\dfrac{\partial\Phi}{\partial z}\right)^2 - \left(\dfrac{\partial\Phi}{\partial u}\right)^2\right\}, \\ \cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot \\ \cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot \\ \boldsymbol{G}_{uu} = \dfrac{1}{2}\left\{-\left(\dfrac{\partial\Phi}{\partial x}\right)^2 - \left(\dfrac{\partial\Phi}{\partial y}\right)^2 - \left(\dfrac{\partial\Phi}{\partial z}\right)^2 + \left(\dfrac{\partial\Phi}{\partial u}\right)^2\right\}, \\ \boldsymbol{G}_{xy} = \dfrac{\partial\Phi}{\partial x}\dfrac{\partial\Phi}{\partial y}, \\ \cdot\;\cdot\;\cdot\;\cdot\;\cdot \\ \cdot\;\cdot\;\cdot\;\cdot\;\cdot \\ \cdot\;\cdot\;\cdot\;\cdot\;\cdot \\ \cdot\;\cdot\;\cdot\;\cdot\;\cdot \\ \boldsymbol{G}_{zu} = \dfrac{\partial\Phi}{\partial z}\dfrac{\partial\Phi}{\partial u}. \end{array}\right\} \tag{57}$$

17 Abraham obtains precisely the same expression in his theory mentioned above; M. Abraham, loc. cit. p. 3.

These quantities give the fictitious gravitational stresses (pressure taken as positive), as well as momentum density, energy flux and energy density in the gravitational field. For the energy flux $\mathfrak{S}^g$ and for the momentum density $\mathfrak{g}^g$, one has

$$\mathfrak{S}_x{}^g = c^2\mathfrak{g}_x{}^g = -ic\boldsymbol{G}_{xu} \qquad \text{etc.,}$$

and for the energy density ψ^g

$$\psi^g = -\boldsymbol{G}_{uu}.$$

Hence, according to (57), in the notation of vector analysis

$$\mathfrak{S}^g = c^2\mathfrak{g}^g = -\frac{\partial\Phi}{\partial t}\nabla\Phi, \tag{58}$$

$$\psi^g = \frac{1}{2}\left\{(\nabla\Phi)^2 + \frac{\mathrm{i}}{c^2}\left(\frac{\partial\Phi}{\partial t}\right)^2\right\}. \tag{59}$$

One sees that ψ^g is always positive.

The last of the equations (56), multiplied by ic, is now

$$ic\mathfrak{K}_u{}^g = \operatorname{div}\mathfrak{S}^g + \frac{\partial\psi^g}{\partial t}, \tag{60}$$

which is the equation expressing the law of conservation of energy for the gravitational field. For regions outside of the material bodies, we have, | of course,
[876] $\mathfrak{K}_u{}^g = 0$. For regions within the bodies, eq. (60) is to be combined with eq. (22).

Equation (48) can obviously be viewed as a four-dimensional Poisson equation, and its integration can be performed accordingly.[18] However, the form of the eq. (48) also shows that one can calculate Φ according to the well-known formula for the retarded potential. Taking into account the possibility that g might be variable, one has

$$\left.\begin{aligned}\Phi(x_0, y_0, z_0, t_0) &= -\frac{1}{4\pi}\int\frac{dxdydz}{r}(g\nu)_{x,\,y,\,z,\,t} + \text{const.}\\ &= -\frac{1}{4\pi}\int\frac{dxdydz}{r}(g\mu\sqrt{1-\mathfrak{q}^2})_{x,\,y,\,z,\,t} + \text{const.,}\end{aligned}\right\} \tag{61}$$

where

18 M. Abraham, *Physik. Zeitschr.* 13, p. 5, 1912; A. Sommerfeld, *Ann. d. Phys.*, 33, p. 665, 1910.

$$\left.\begin{aligned} r &= \sqrt{(x-x_0)^2+(y-y_0)^2+(z-z_0)^2}, \\ t &= t_0 - \frac{r}{c}. \end{aligned}\right\} \tag{61a}$$

The integration is to be carried out over three-dimensional space.

7. FREE-FALL MOTION

We first wish to establish an equation for the motion of a mass point in an arbitrary *static* gravitational field. On this occasion we should make two comments. First, our theory does not allow real point-like masses, because at such a point by (61) we would have $\Phi = -\infty$, and hence, by (53), the mass of the point would be zero. Thus a "mass point" must always have a certain extension. Second, it should be noted that in order to allow one to treat the field as static, the particle moving in the field must be constituted such that its own field is weak in comparison to the external field, even in its immediate vicinity. |

In the static field one has [877]

$$\frac{\partial \Phi}{\partial t} = 0.$$

We multiply the first three of the equations (54) by $\mathfrak{v}_x$, $\mathfrak{v}_y$, $\mathfrak{v}_z$ and add them. On the left hand side we obtain $-g\mathfrak{v}\nabla\Phi$. Furthermore, one has generally

$$\frac{d\mathfrak{B}_x}{d\tau} = \frac{1}{1-\mathfrak{q}^2}\frac{d\mathfrak{v}_x}{dt} + \frac{\mathfrak{v}_x}{c^2(1-\mathfrak{q}^2)^2}\mathfrak{v}\frac{d\mathfrak{v}}{dt} \qquad \text{etc.,} \tag{62}$$

and hence

$$\mathfrak{v}_x\frac{d\mathfrak{B}_x}{d\tau} + \mathfrak{v}_y\frac{d\mathfrak{B}_y}{d\tau} + \mathfrak{v}_z\frac{d\mathfrak{B}_z}{d\tau} = \frac{1}{(1-\mathfrak{q}^2)^2}\mathfrak{v}\frac{d\mathfrak{v}}{dt}.$$

Since furthermore in our case

$$\frac{d\Phi}{d\tau} = \frac{1}{\sqrt{1-\mathfrak{q}^2}}\mathfrak{v}\nabla\Phi,$$

we obtain

$$-g\mathfrak{v}\nabla\Phi = \frac{\mathrm{i}}{(1-\mathfrak{q}^2)^2}\mathfrak{v}\frac{dv}{dt} + g\frac{\mathfrak{q}^2}{1-\mathfrak{q}^2}\mathfrak{v}\nabla\Phi,$$

and finally

$$-g\mathfrak{v}\nabla\Phi = \frac{1}{1-\mathfrak{q}^2}\mathfrak{v}\frac{dv}{dt}. \tag{63}$$

Now we wish to assume more specifically that the gravitational field is homogeneous and parallel to the z-axis, and hence that

$$\frac{\partial\Phi}{\partial z} = \text{const.}, \quad \frac{\partial\Phi}{\partial x} = \frac{\partial\Phi}{\partial y} = \frac{\partial\Phi}{\partial u} = 0,$$

and investigate the motion of a mass point in this field. The third of the equations (54) gives

$$-g\frac{\partial\Phi}{\partial z} = \frac{1}{1-\mathfrak{q}^2}\frac{d\mathfrak{v}_z}{dt} + \frac{\mathfrak{v}_z}{c^2(1-\mathfrak{q}^2)^2}\mathfrak{v}\frac{d\mathfrak{v}}{dt} + g\frac{\mathfrak{v}_z{}^2}{c^2(1-\mathfrak{q}^2)}\frac{\partial\Phi}{\partial z};$$

taking (63) into account we find that the last two terms cancel one another and we obtain

$$\frac{d\mathfrak{v}_z}{dt} = -(1-\mathfrak{q}^2)g\frac{\partial\Phi}{\partial z}.$$

The first of the equations (54) yields in a similar manner

$$0 = \frac{1}{1-\mathfrak{q}^2}\frac{d\mathfrak{v}_x}{dt} + \frac{\mathfrak{v}_x}{c^2(1-\mathfrak{q}^2)^2}\mathfrak{v}\frac{d\mathfrak{v}}{dt} + g\frac{\mathfrak{v}_x\mathfrak{v}_z}{c^2(1-\mathfrak{q}^2)}\frac{\partial\Phi}{\partial z}.$$

Here too, the last two terms cancel one another, and the equation becomes
[878] $d\mathrm{v}_x/dt = 0$. Since the same must be true for $d\mathrm{v}_y/dt$, | for a mass-point in a homogeneous gravitational field we obtain the equations of motion

$$\left.\begin{aligned} \frac{d\mathfrak{v}_z}{dt} &= -\left(1-\frac{\mathfrak{v}^2}{c^2}\right)g\frac{\partial\Phi}{\partial z}, \\ \frac{d\mathfrak{v}_x}{dt} &= \frac{d\mathfrak{v}_y}{dt} = 0. \end{aligned}\right\} \tag{64}$$

These equations state the following: *The velocity component perpendicular to the field direction is uniform. Gravitational acceleration becomes smaller as the velocity increases, but this is independent of the direction of the velocity. A body projected horizontally falls slower than one without initial velocity falling vertically.* One also sees that a rotating body must fall slower than a non-rotating one. Of course, for attainable rotational speeds the difference is much too small to be amenable to observation.

These results raise the question of whether the molecular motions within a falling body also have an influence on the gravitational acceleration. At least one cannot deny the possibility that this is the case. The theory of gravitation is then simply to be modified by considering the gravitational factor g as dependent on the molecular motions within the body rather than as a constant. For this reason we have left this possibility open in the foregoing treatment. In this context, it should be pointed out that also the mass density of a body depends upon the molecular motions, since the rest-energy density, which determines ν according to eq. (33), is influenced by these motions.

However, those questions of the theory of gravitation which are related to the atomic structure of matter lie beyond the scope of this essay.

EDITORIAL NOTE

[1] In the original, Nordström mistakenly refers to eq. (4) rather than eq. (3).

GUNNAR NORDSTRÖM

ON THE THEORY OF GRAVITATION FROM THE STANDPOINT OF THE PRINCIPLE OF RELATIVITY

Originally published as "Zur Theorie der Gravitation vom Standpunkt des Relativitätsprinzips" in Annalen der Physik, 42, 1913, pp. 533–554. Received July 24, 1913. Author's date: Zurich, July 1913.

In the present communication, I wish to develop further several aspects of the theory of gravitation whose fundamentals I published in two previous essays and discuss it.[1] The theory presented in the last essay is not completely unambiguous. First—as emphasized on p. 509 [p. 867 in the original]—the rest density of matter was defined in a fairly arbitrary way; though a different definition of the concept of mass would not change the general laws of mechanics, it would modify the laws of gravitation. Second, in the theory of gravitation, the possibility has been left open that the gravitational factor g is not a constant, but could depend on various circumstances. One can think of this scalar quantity as being dependent on the internal state of the object as well as on the gravitational potential at the location in question. A dependence of the gravitational factor on the state of stress of the body is equivalent to a change in the definition of mass, but a dependence on the gravitational potential will have a deeper significance for the theory.

1. DEFINITE FORMULATION OF THE THEORY

All the aforementioned ambiguities of the theory can be removed by a very plausible stipulation which I owe to Mr. Laue and Mr. Einstein. Mr. Laue has shown that one can maintain Einstein's theorem of equivalence—though not in its full scope—by defining | the rest density of matter in an appropriate manner,[2] namely by means of the sum [p. 534]

$$\boldsymbol{T}_{xx} + \boldsymbol{T}_{yy} + \boldsymbol{T}_{zz} + \boldsymbol{T}_{uu} = -D \tag{1}$$

1 G. Nordström, *Physik. Zeitschr.,* 13, p. 1126, 1912; *Ann. d. Phys.,* 40, p. 856, 1913 [both in this volume]. The present communication is a continuation of the latter, and the symbol loc. cit. in the text refers to the same.

2 See A. Einstein, "Entwurf einer verallgemeinerten Relativitätstheorie und einer Theorie der Gravitation." *Zeitschr. f. Math. Phys.,* 62, p. 21, 1913.

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 3
Gravitation in the Twilight of Classical Physics: Between Mechanics, Field Theory, and Astronomy.

of the diagonal components of the tensor $\boldsymbol{T}$, which represents the state of the matter. $\boldsymbol{T}$ is the dynamical tensor introduced by Laue,[3] and it is equal to the sum of the two tensors which I earlier called the elastic stress tensor and the material tensor.[4] Following Einstein, we will call the invariant D defined by eq. (1) Laue's scalar, and we will find that when divided by c^2 it represents the rest density.

Furthermore, it will turn out that Einstein's theorem of equivalence demands a very particular dependence of the gravitational factor g on the gravitational potential Φ; we put

$$g = g(\Phi).$$

If we furthermore denote the rest density of matter by ν, the fundamental equations for the gravitational field are[5]

$$\frac{\partial^2\Phi}{\partial x^2} + \frac{\partial^2\Phi}{\partial y^2} + \frac{\partial^2\Phi}{\partial z^2} + \frac{\partial^2\Phi}{\partial u^2} = g(\Phi)\nu, \tag{2}$$

$$\left.\begin{aligned} \mathfrak{K}_x^g &= -g(\Phi)\nu\frac{\partial\Phi}{\partial x}, \qquad & \mathfrak{K}_y^g &= -g(\Phi)\nu\frac{\partial\Phi}{\partial y}, \\ \mathfrak{K}_z^g &= -g(\Phi)\nu\frac{\partial\Phi}{\partial z}, \qquad & \mathfrak{K}_u^g &= -g(\Phi)\nu\frac{\partial\Phi}{\partial u}. \end{aligned}\right\} \tag{3}$$

Equation (2) determines the gravitational field produced by a given distribution of masses. The system of eqs. (3) determines the ponderomotive force $\mathfrak{K}^g$, which the field exerts on matter.

The task now is to define the rest density ν and to determine the function $g(\Phi)$ in such a way that the theorem of equivalence is valid in the widest possible sense.

For this purpose, we consider a system of finite bodies such that an appropriate
[p. 535] reference system exists, | in which the gravitational field is static so that we have everywhere $\partial\Phi/\partial t = 0$. However, bodies rotating about their symmetry axis and stationary flows of fluids may occur. In any case, in the reference system under consideration, the total momentum is equal to zero,[6]

$$\mathfrak{G} = \int g dv = 0,$$

and the system as a whole is at rest. A system which satisfies these conditions is to be called a complete stationary system.

---

3 M. Laue, *Das Relativitätsprinzip*, 2nd. Ed., p. 182.

4 G. Nordström, *Ann. d. Phys.,* loc. cit., p. 858.

5 G. Nordström, loc. cit., eqs. (48) and (49).

6 We must exclude heat transport, since otherwise the total momentum is not zero. Besides, heat transport would change with time the energy distribution, and thus also the mass distribution, and therefore make the gravitational field time-dependent.

As the state does not change with time, eq. (2) yields, according to the usual potential theory

$$\Phi = -\frac{1}{4\pi}\int\frac{dv}{r}g(\Phi)\nu + \Phi_a, \tag{4}$$

where the integral extends over all of xyz-space. Φ_a is the value of Φ at infinity and has its origin in other systems of masses, which we assume to be far away. For large distances r, one has

$$|\nabla\Phi| = \frac{1}{4\pi r^2}\int g(\Phi)\nu dv, \tag{5}$$

and the direction of $\nabla\Phi$ is away from the system of masses.

In the most general case, we have in the system three different world-tensors, which give the spatial stresses, the energy flux, and the energy-momentum density: the elastic-material tensor $\boldsymbol{T}$, the gravitation tensor $\boldsymbol{G}$ and the electromagnetic tensor $\boldsymbol{L}$. For the components of the gravitation tensor, the eqs. (57) apply, loc. cit.

$$\left.\begin{aligned} \boldsymbol{G}_{xx} &= \frac{1}{2}\left\{\left(\frac{\partial\Phi}{\partial x}\right)^2 - \left(\frac{\partial\Phi}{\partial y}\right)^2 - \left(\frac{\partial\Phi}{\partial z}\right)^2 - \left(\frac{\partial\Phi}{\partial u}\right)^2\right\}, \\ \boldsymbol{G}_{xy} &= \frac{\partial\Phi}{\partial x}\frac{\partial\Phi}{\partial y} \qquad \text{etc.,} \end{aligned}\right\} \tag{6}$$

and we have

$$\mathfrak{K}_x^g = -\left\{\frac{\partial\boldsymbol{G}_{xx}}{\partial x} + \frac{\partial\boldsymbol{G}_{xy}}{\partial y} + \frac{\partial\boldsymbol{G}_{xz}}{\partial z} + \frac{\partial\boldsymbol{G}_{xu}}{\partial u}\right\} \qquad \text{etc.} \tag{7}$$

| We want to form the sum of the diagonal components for the *total tensor* $\boldsymbol{T} + \boldsymbol{G} + \boldsymbol{L}$ and to integrate over all of three-dimensional space in our system of reference. Since $\partial\Phi/\partial t = 0$ the trace of the gravitation tensor is equal to $-(\nabla\Phi)^2$; the trace of $\boldsymbol{T}$ has been denoted by $-D$, and the trace of the electromagnetic tensor is equal to zero. Hence, we form the integral [p. 536]

$$-\int\{D + (\nabla\Phi)^2\}dv$$

extended over all of space. But since, according to a theorem of Laue[7] we have

$$\int\{\boldsymbol{T}_{xx} + \boldsymbol{G}_{xx} + \boldsymbol{L}_{xx}\}dv = 0$$

and two corresponding equations for the yy- and zz- components hold, we obtain

7 M. Laue, loc. cit., p. 209.

$$-\int\{D+(\nabla\Phi)^2\}dv = \int\{\boldsymbol{T}_{uu}+\boldsymbol{G}_{uu}+\boldsymbol{L}_{uu}\}dv = -E_0,$$

where E_0 denotes the energy of the whole system in the frame of reference used (the rest energy).

Since $\partial^2\Phi/\partial u^2 = 0$, eq. (2) yields

$$\operatorname{div}\nabla\Phi = g(\Phi)\nu,$$
$$\operatorname{div}\Phi\nabla\Phi = \Phi g(\Phi)\nu+(\nabla\Phi)^2.$$

The integration of $(\nabla\Phi)^2$ over a ball of infinitely large radius, taking (5) into account, yields

$$\int(\nabla\Phi)^2 dv = -\int(\Phi-\Phi_a)g(\Phi)\nu dv.$$

Therefore, for E_0 one obtains

$$E_0 = \int Ddv-\int(\Phi-\Phi_0)g(\Phi)\nu dv. \tag{8}$$

Since the total momentum in the system of reference under consideration is zero, in a different system of reference, in which our system of masses is moving with the speed $\mathfrak{v}$, one has the following expressions for the energy E and for the momentum $\mathfrak{G}$,[8]

$$E = \frac{E_0}{\sqrt{1-\mathfrak{q}^2}}, \qquad \mathfrak{G} = \frac{E_0\cdot\mathfrak{v}}{c^2\sqrt{1-\mathfrak{q}^2}},$$

where we have set $q = \mathfrak{v}/c$. The inertial mass of the system is thus:

$$m = \frac{E_0}{c^2} = \frac{1}{c^2}\int\{D-(\Phi-\Phi_a)g(\Phi)\nu\}dv \tag{9}$$

From eqs. (4), (5) and (3), one sees that the quantity |

[p. 537] $$M_g = \int g(\Phi)\nu dv \tag{10}$$

determines the gravitational effects that the system exerts and experiences. M_g will be called the *gravitational mass* of the system. Einstein's equivalence theorem implies that for various systems M_g and m are proportional to each other. Then M_g/m can only be a function of Φ_a, and this function can be none but $g(\Phi_a)$, so that one has

8 M. Laue, 1oc. cit., p. 209.

$$g(\Phi_a)m = \int g(\Phi)\nu dv. \tag{11}$$

Now the task is to define the rest density ν and to determine the function $g(\Phi)$ by means of this stipulation. We equate the expressions obtained for m from (9) and (11) and obtain

$$\frac{1}{c^2}\int D dv = \frac{1}{c^2}\int g(\Phi)\nu\left\{\Phi - \Phi_a + \frac{c^2}{g(\Phi_a)}\right\}dv.$$

In order to satisfy this equation identically, we set

$$D = g(\Phi)\nu\left\{\Phi - \Phi_a + \frac{c^2}{g(\Phi_a)}\right\}. \tag{12}$$

Since furthermore, the potential Φ_a of the external field must drop out of the equation, we set

$$\frac{c^2}{g(\Phi)} - \Phi = A, \qquad g(\Phi) = \frac{c^2}{A + \Phi}, \tag{13}$$

where A signifies a universal constant.

From (12), one now obtains

$$\nu = \frac{1}{c^2}D = -\frac{1}{c^2}(\boldsymbol{T}_{xx} + \boldsymbol{T}_{yy} + \boldsymbol{T}_{zz} + \boldsymbol{T}_{uu}), \tag{14}$$

whereby the rest density of the matter is defined.[9]

It is to be noted that the value of the constant A is unknown, since the absolute value of the potential Φ cannot be calculated at any point. Denoting the gravitational potential at a point accessible to investigation by Φ, | and that part of Φ which arises [p. 538] from masses external to our solar system by Φ_0, then only the difference $\Phi - \Phi_0$ can be determined by means of any kind of observations concerning the quantities Φ and Φ_0. Let $g(\Phi_0)$ be denoted by g_0. If we eliminate the quantity A from the equations

$$g(\Phi) = \frac{c^2}{A + \Phi} \quad \text{and} \quad g_0 = \frac{c^2}{A + \Phi_0},$$

we obtain for the function $g(\Phi)$:

---

9 In addition it should be noted that one can also express ν by means of the relative stress t and the rest energy density Ψ of matter. One finds

$$\nu = \frac{1}{c^2}\{\Psi - t_{xx} - t_{yy} - t_{zz}\}.$$

$$g(\Phi) = \frac{g_0}{1 + \frac{g_0}{c^2}(\Phi - \Phi_0)}. \tag{15}$$

Only experimentally accessible quantities appear in this equation. Naturally, one could also fix the initial potential Φ_0 in a different way, since for two arbitrary values of Φ one has

$$g(\Phi_2) = \frac{g(\Phi_1)}{1 + \frac{g(\Phi_1)}{c^2}(\Phi_2 - \Phi_1)}. \tag{15a}$$

The eqs. (14) and (15) for ν and g uniquely determine the theory of gravitation. If we set

$$\Phi' = \Phi + A = \frac{c^2}{g_0} + \Phi - \Phi_0, \tag{16}$$

then the new gravitational potential Φ' is not afflicted with the ambiguity of Φ. We obtain from (13) and (11)

$$g(\Phi) = \frac{c^2}{\Phi'}, \tag{17}$$

$$M_g = \frac{mc^2}{\Phi_a'} = \frac{E_0}{\Phi_a'}, \tag{18}$$

and if one inserts Φ' into the fundamental eqs. (2), (3), they become

$$\Phi'\left\{\frac{\partial^2\Phi'}{\partial x^2} + \frac{\partial^2\Phi'}{\partial y^2} + \frac{\partial^2\Phi'}{\partial z^2} + \frac{\partial^2\Phi'}{\partial u^2}\right\} = c^2\nu, \tag{19}$$

$$\left.\begin{aligned} \mathfrak{K}_x^g &= -c^2\nu\frac{\partial}{\partial x}\ln\Phi', & \mathfrak{K}_y^g &= -c^2\nu\frac{\partial}{\partial y}\ln\Phi', \\ \mathfrak{K}_z^g &= -c^2\nu\frac{\partial}{\partial z}\ln\Phi', & \mathfrak{K}_u^g &= -c^2\nu\frac{\partial}{\partial u}\ln\Phi'. \end{aligned}\right\} \tag{20}$$

If they are written in this way, no universal constant corresponding to the gravita-
[p. 539] tional constant appears in the fundamental equations.[10] | In eq. (19) one can, to a certain approximation, take Φ' in front of the brackets as a constant equal to c^2/g_0. By integration one then obtains the usual formula for the retarded potential.

10 But in §6 it is to be shown that such a universal constant plays a role in the definition of the fundamental units.

2. DEPENDENCE OF A BODY'S MASS ON THE GRAVITATIONAL POTENTIAL

We wish to prove that the inertial mass of a system depends on the properties of the gravitational potential Φ_a of the external field outlined in §1. We examine the state of affairs from the system of reference in which the bodies produce a static gravitational field and construct about the bodies a spherical surface of very large radius r. At the points on this surface, $\nabla\Phi$ is directed vertically outwards and has, according to (5), the magnitude

$$|\nabla\Phi| = \frac{M_g}{4\pi r^2}.$$

We imagine that the gravitational potential Φ_a of the external field is produced by masses which lie very far away from our system of bodies and outside the spherical surface. For the time being, Φ_a is spatially and temporally constant inside this surface. Then we imagine Φ_a being changed by $d\Phi_a$ due to a slow displacement of the distant masses. This change engenders a certain flow of energy through the spherical surface, which we want to calculate. The energy flux $\mathfrak{S}^g$ in the gravitational field is according to eq. (58), loc. cit., as well as by (6)

$$\mathfrak{S}^g = -\frac{\partial\Phi}{\partial t}\nabla\Phi.$$

By integrating over the spherical surface, we find that the change $d\Phi_a$ of the external potential results in energy transport through the spherical surface to the interior, having the magnitude

$$4\pi r^2|\nabla\Phi|d\Phi_a = M_g d\Phi_a.$$

Hence, the amount by which the rest energy E_0 of our system has been increased is:

$$dE_0 = M_g d\Phi_a.$$

| If we insert here $E_0 = c^2 m$ from (9) and $M_g = g(\Phi_a)m$ from (11), we obtain the equation [p. 540]

$$\frac{1}{m}dm = \frac{g(\Phi_a)}{c^2}d\Phi_a, \tag{a}$$

which agrees with the eq. (52), loc. cit., found by a different method. According to (17), we have further

$$d\ln m = d\ln\Phi_a',$$

and obtain finally through integration

$$\frac{m}{\Phi_a'} = \text{const.} \tag{21}$$

The inertial mass of a body is thus directly proportional to the gravitational potential $\Phi_a{}'$ of the external field. According to (16), we can also write this dependence in the following form

$$m = m_0\left\{1 + \frac{g_0}{c_2}(\Phi_a - \Phi_0)\right\}, \tag{21a}$$

where m_0 is the inertial mass associated with the gravitational potential Φ_0.[11]

According to (21) and (18) we have

$$M_g = \int g(\Phi)\nu dv = \text{const.} \tag{22}$$

Thus, in contrast to the inertial mass, the gravitational mass is a characteristic constant for each body that does not depend upon the external gravitational potential.

3. INERTIAL AND GRAVITATIONAL MASS OF A SPHERICAL ELECTRON

As an example of the theory, we want to establish the formulas for the inertial and gravitational mass of a spherical electron with uniform surface charge. Let the electric charge of the electron expressed in rational units be e, and the radius a. In order to prevent the unlimited expansion of the electron as a result of the force of repulsion between equal electric charges, certain elastic stresses must act within the electron. Most conveniently, we assume that these stresses are concentrated on the surface of
[p. 541] the electron as well. | According to eq. (14), the elastic stresses give the electron a mass, which also produces gravitational effects. The gravitational field is superimposed on the electric field and both fields act back on the electron. The electron is at rest. We have to think of the surface of the electron as an infinitely thin shell, in which the elastic tensor $\boldsymbol{T}$ is different from zero. We assume that the component $\boldsymbol{T}_{uu}{}^0$ equals zero, so that, for the case of rest, the tensor $\boldsymbol{T}$ does not contribute to the energy of the electron. Hence, the tensor $\boldsymbol{T}$ reduces (for the case of rest) to a spatial stress tensor, which for reasons of symmetry must have one principal axis in the direction of the radius. Since we should have a tensile stress [*Zugspannung*] parallel to the shell only, the principal component of $\boldsymbol{T}$ in the direction of the radius equals zero; the two other principal components are equal to each other and will be called p. The trace of $\boldsymbol{T}$ is thus $2p$ and the rest density ν becomes

$$\nu = -\frac{1}{c^2}2p.$$

The tensile stress S in the surface is equal to the line integral

11 The eqs. (21) and (21a) take the place of (53), loc. cit., which presupposes a constant g and thus has now lost its validity.

$$S = -\int p ds$$

integrated across the shell.

We now form the integral $\int \nu dv$ extended over the whole shell and obtain

$$\int \nu dv = -\frac{2}{c^2} 4\pi a^2 \int p ds = \frac{1}{c^2} 8\pi a^2 S.$$

Since for reasons of symmetry, $g(\Phi)$ must have the same value at all points of the surface, we obtain for the gravitational mass M_g of the electron

$$M_g = \int g(\Phi) \nu dv = \frac{g(\Phi)}{c^2} 8\pi a^2 S.$$

On the surface of the electron, the gravitational potential has the value

$$-\frac{M_g}{4\pi a} + \Phi_a ,$$

where Φ_a is the potential of the external field (not produced by the electron). Hence, according to (15a), we can substitute the expression |

$$g(\Phi) = \frac{g_a}{1 - \frac{g_a}{c^2} \frac{M_g}{4\pi a}}, \tag{23}$$ [p. 542]

for $g(\Phi)$, and thus obtain the equation

$$M_g = \frac{g_a}{c^2} \frac{8\pi a^2 S}{1 - \frac{g_a}{c^2} \frac{M_g}{4\pi a}}. \tag{a}$$

The task is now to calculate S from the forces the electric field and the gravitational field exert on the surface of the electron. On each element of the electron's surface, the electric field exerts a force perpendicular to and directed outwards from the surface, whose magnitude per unit of surface area is given by the Maxwellian stresses. One finds for this outward-directed force per unit area the expression

$$\frac{e^2}{32\pi^2 a^4}.$$

In a similar manner, the gravitational field exerts a force on each surface element of the electron's surface, which is perpendicular but directed inwards. Outside of the electron one has

$$\Phi = -\frac{M_g}{4\pi r} + \Phi_a, \qquad \frac{d\Phi}{dr} = \frac{M_g}{4\pi r^2}.$$

The fictitious gravitational stress on the surface is perpendicular to the surface, and by (6) it has the magnitude

$$\frac{1}{2}\left(\frac{d\Phi}{dr}\right)^2 = \frac{M_g{}^2}{32\pi^2 a^4}.$$

This force is exerted by the gravitational field on the electron's surface per unit of surface area. The combined force which the two fields exert on a unit of surface area is thus

$$P = \frac{e^2 - M_g{}^2}{32\pi^2 a^4}, \tag{b}$$

where positive is outward-directed. The force P and the elastic stress S in the electron's surface should now maintain equilibrium. It is easy to find that the condition for equilibrium is

$$2S = aP. \tag{c}$$

To derive this relation, one can for example think of the spherical electron surface as
[p. 543] divided into two equal halves. | The normal force P attempts to drive the two halves apart with a total force of $\pi a^2 P$, whereas the stress S holds the two halves together with a total force of $2\pi a S$. By equating the two expressions for the force, one obtains the relation (c).

The eqs. (a), (b), (c) yield

$$M_g = \frac{g_a}{c^2}\frac{e^2 - M_g{}^2}{8\pi a} \cdot \frac{1}{1 - \frac{g_a}{c^2}\frac{M_g}{4\pi a}}.$$

We thus have a quadratic equation for M_g:

$$M_g - \frac{g_a}{c^2}\frac{M_g{}^2}{8\pi a} = \frac{g_a}{c^2}\frac{c^2}{8\pi a}.$$

From this we obtain

$$M_g = \frac{g_a}{c^2} \cdot \frac{e^2 + M_g{}^2}{8\pi a}, \tag{23}$$

and since according to (11) $M_g = g_a \cdot m$, we obtain for the inertial mass m of the electron the expression

$$m = \frac{e^2 + M_g{}^2}{8\pi c^2 a} = \frac{E_0}{c^2}. \tag{24}$$

The rest energy E_0 is composed of two parts: the energy of the electric field and that of the gravitational field. These two parts of E_0 are

$$\frac{e^2}{8\pi a} \quad \text{and} \quad \frac{M_g{}^2}{8\pi a}.$$

Thus the expression found for m contains a confirmation of the theory.

4. DEPENDENCE OF THE DIMENSIONS OF LENGTH ON THE GRAVITATIONAL POTENTIAL

An important conclusion can furthermore be drawn from the expression (24) for the inertial mass of an electron. If the gravitational potential Φ_a of the external field (not stemming from the electron) is changed, then m changes according to eq. (21). However, by (22), M_g remains constant, and the same applies for e, according to the fundamental equations of electrodynamics. Therefore, the radius a must vary inversely with m, and one obtains from (21)

$$a\Phi_a{}' = \text{const.} \tag{25}$$

| Since on the electron surface [p. 544]

$$\Phi' = \Phi_a{}' - \frac{M_g}{4\pi a} = \Phi_a{}' \cdot \text{const.},$$

one also has

$$a\Phi' = \text{const.} \tag{25a}$$

The elastic tension S in the electron surface varies also with Φ_a. From (a), p. 531 [p. 542 in the original], we find the law for this if we take into account that according to (17) and (25) g_a/a remains constant. We see that $a^3 S$ must be constant, and that therefore

$$\frac{S}{\Phi_a{}'^3} = \text{const.}$$

Mr. Einstein has proved that the dependence in the theory developed here of the length dimensions of a body on the gravitational potential must be a general property of matter. He has shown that otherwise it would be possible to construct an apparatus with which one could pump energy out of the gravitational field. In Einstein's example one considers a non-deformable rod that can be constrained to move between two vertical rails. One could let the rod fall while stressed, then remove the stress and raise it back up. The rod has a greater weight when stressed than when unstressed, and therefore it would do more work than would be consumed in raising the unstressed rod. However, because the rod lengthens while falling, the rails must diverge, and the excess work from the fall will be consumed again as the work done by the tensioning forces on the ends of the rod.

Let S be the total stress (stress times cross-sectional area) of the rod and l its length. Because of the stress, the gravitational mass of the rod is increased by

$$\frac{g(\Phi)}{c^2}Sl = \frac{1}{\Phi'}Sl.$$

In falling, this gravitational mass provides the extra work

$$-\frac{1}{\Phi'}sld\Phi'.$$

[p. 545] | However at the same time at the ends of the rod the work

$$Sdl$$

is lost. Equating these two expressions yields

$$-\frac{1}{\Phi'}d\Phi' = \frac{1}{l}dl,$$

which, on integration, gives

$$l\Phi' = \text{const.}$$

But this corresponds precisely to eq. (25a).[12]

The result found for the stressed rod, as well as other examples, shows that the eqs. (25) and (25a) possess a general validity for a material body's dimensions of length. Of course it is the real gravitational potential Φ' existing at a point, and not that of the external field, which influences the length; however, we can easily see that we may also generally use the potential $\Phi_a{}'$ of the external field when calculating changes in length, since Φ' and $\Phi_a{}'$ are proportional to one another. For a system of the kind considered in §1 at rest, we have from (4) and (17)

$$\Phi' = \Phi_a{}' - \frac{c^2}{4\pi}\int\frac{\nu dv}{\Phi' r}. \tag{a}$$

According to the results found earlier, the gravitational mass of the system does not change when the gravitational potential $\Phi_a{}'$ of the external field is changed. Therefore, upon such a change we have

12 If the rod is deformable, some work will be expended in stressing it, and the rest energy of the rod will be correspondingly increased. In this way too, the weight experiences an increase, which provides an added work dA in falling. However, since in falling the rest energy diminishes, the work recovered in relaxing the rod is smaller than that consumed in stressing and the difference amounts to exactly dA.

$$c^2 \int \frac{\nu dv}{\Phi'} = \text{const.}$$

The quantities ν and dv do change in a certain way with Φ', but in such a way that $\nu dv / \Phi'$ remains constant for each particular element of the system. If on the left in eq. (a) Φ' denotes the potential at a certain point of the material system, | the integral [p. 546] on the right varies inversely with the length r. We obtain

$$\Phi_a{}' = \Phi'(1 + \text{const.}),$$

that is, at each point of the system, Φ' changes in proportion to $\Phi_a{}'$.

For these reasons, the dependence of a body's linear dimensions l (at rest) on the gravitational potential is given generally by the two equivalent equations

$$l\Phi' = \text{const.}, \qquad l\Phi_a{}' = \text{const.}, \tag{26}$$

and further, corresponding to (21a)

$$l = \frac{l_0}{1 + \frac{g_0}{c^2}(\Phi_a - \Phi_0)}. \tag{26a}$$

For the volume dv of a particle of a body transformed to rest, one has, of course, from (26)

$$dv\Phi'^3 = \text{const.}$$

Since above we found $\nu dv / \Phi' = \text{const.}$, it follows that

$$\frac{\nu}{\Phi'^4} = \text{const.} \tag{27}$$

Since according to (14) $-c^2\nu$ is the sum of the diagonal components of the tensor $\boldsymbol{T}$, the components of $\boldsymbol{T}$, and thus in particular the elastic stresses $\boldsymbol{p}_{ab}$, depend on Φ' in the same way as ν:

$$\frac{\boldsymbol{p}_{ab}}{\Phi'^4} = \text{const.} \tag{28}$$

The result found earlier (p. 533) [p. 544 in the original] for the electron's surface tension S is thus in agreement with the above since S is a stress times a length.

5. THE DEPENDENCE OF A PROCESS'S TIME DEVELOPMENT ON THE GRAVITATIONAL POTENTIAL

The dependence of a body's linear dimensions on Φ' raises the question of whether a physical process's time development is also influenced by the gravitational potential. [p. 547] For a simple case we can answer the question | without difficulty. Due to the constancy of the speed of light, it is clear that the time during which a light signal propagates from one end of a rod to the other grows in the same proportion as the rod lengthens. This time is thus inversely proportional to the gravitational potential.

Another process that can be treated without difficulty is circular motion under the influence of the gravitational attraction of a central body. Let a mass point with gravitational mass M_2 move in a circular orbit around another mass point with gravitational mass M_1, with M_2 much smaller than M_1. Let M_1 be so large in relation to M_2 that we may consider the former as being at rest. In §7 we will further investigate the question of when a body may be viewed as a mass point, and derive the laws of its motion. Here, we only need to know that for a mass point the equations of motion (51), loc. cit. apply, where Φ denotes the gravitational potential of the field not produced by the mass point itself. This potential is

$$\Phi = \Phi_a - \frac{M_1}{4\pi r},$$

where Φ_a is the constant external potential and r is the distance from the mass M_1. Because Φ has the same value at all points on the circular orbit, the inertial mass of the moving point remains unchanged, and the equations of motion (51), loc. cit. yield:

$$-g(\Phi)\frac{\partial\Phi}{\partial x} = \frac{d\mathfrak{B}_x}{d\tau}, \text{ etc.}$$

According to (17), we can also write the equation as

$$-\frac{\partial\Phi}{\partial x} = \frac{\Phi'}{c_2}\frac{d\mathfrak{B}_x}{d\tau}$$

etc. Of course,

$$|\nabla\Phi| = \frac{M_1}{4\pi r^2}.$$

Furthermore, since the moving mass point has no tangential acceleration, we have

$$\frac{d\mathfrak{B}_x}{d\tau} = \frac{1}{1-\mathfrak{q}^2}\frac{d\mathfrak{v}_x}{dt}$$

etc. (e.g., compare eq. (62), loc. cit.). For the absolute value of the acceleration, |

$$\left|\frac{d\mathfrak{v}}{dt}\right| = \frac{\mathfrak{v}^2}{r}$$ [p. 548]

applies, where r is the radius of the circular orbit. By using all of these equations we obtain the following equation of motion

$$\frac{M_1}{4\pi r^2} = \frac{\Phi_a{}' - \frac{M_1}{4\pi r}}{c^2(1-\mathfrak{q}^2)} \cdot \frac{\mathfrak{v}^2}{r}.$$

A transformation yields

$$\frac{M_1}{4\pi r} = \Phi_a{}' \frac{\mathfrak{v}^2}{c^2}$$

or, if we introduce the period of rotation T,

$$\frac{M_1 c^2}{4\pi r \Phi_a{}'} = \frac{4\pi^2 r^2}{T^2}. \tag{a}$$

This equation connects the three quantities r, T and $\Phi_a{}'$ with one another. Setting $c^2/\Phi' = g$ according to (17), we obtain precisely the equation which classical mechanics would give.

We now imagine the two mass points M_1 and M_2 and also the measuring rods, with which we measure length, transferred to another location with a different external gravitational potential $\Phi_a{}'$. Then all lengths have changed in inverse proportion to $\Phi_a{}'$, and if we wish to re-establish the previous process, we measure the distance r such that $r \cdot \Phi_a{}'$ has the same value in both cases. Therefore, according to (a), r/T also has the same value, which means that the time of revolution changes in proportion to the bodies' linear dimensions. Therefore according to (26) one has,

$$T\Phi_a{}' = \text{const.} \tag{29}$$

We further want to investigate the behavior of the period of oscillation of a material point, oscillating about a fixed equilibrium position as a result of an elastic (or a "quasi-elastic") force. For a sufficiently small amplitude of oscillation we can use the usual harmonic oscillator equation:

$$m\frac{d^2x}{dt^2} = -ax,$$

where m is the inertial mass of the material point, x the | displacement from equilibrium, and a an elastic constant. For the period of oscillation T one obtains the well known expression (the easiest way to obtain this is by substituting $x = Ce^{2\pi i(t/T)}$) [p. 549]

$$T^2 = 4\pi^2\frac{m}{a}.$$

Upon a change of the gravitational potential one has from (21)

$$\frac{m}{\Phi'} = \text{const.}$$

Since ax is the elastic force, it is to be treated like a pull on a stretched string. Thus upon a change of Φ', it behaves like a stress times an area (the cross-section of the string), and according to (28) and (26)

$$\frac{ax}{\Phi'^2} = \text{const.}$$

Since x is a length, one has

$$\frac{a}{\Phi'^3} = \text{const.}$$

Hence for the period of oscillation

$$T\Phi' = \text{const.}$$

precisely in agreement with the two earlier results. It may be supposed that the course of all physical processes is influenced in a corresponding manner.

From the last example, it follows that the wavelength of a spectral line depends upon the gravitational potential. A numerical calculation shows that the wavelengths on the surface of the Sun must be greater by about one part in two million than those of terrestrial light sources. Several other recent theories of gravitation also give the same—perhaps even observable—displacement.

6. REMARKS ON THE DEFINITION OF FUNDAMENTAL UNITS

From the dependence of the linear dimensions and masses of the bodies as well as of the time development of phenomena on the gravitational potential, it follows that in
[p. 550] defining the fundamental units, the gravitational potential has to be taken | into account. By a centimeter, we thus understand the length of a reference rod at a certain temperature and at a certain gravitational potential. For the latter, one takes of course the potential present on the surface of the Earth. The same applies for the definition of the unit of time and the unit of inertial mass.

If the units of length and time have been established for a location with a certain gravitational potential, from this location one can in principle measure all lengths and times in the world by means of a telescope and the exchange of light signals, because light signals propagate in straight lines with a constant speed c. Hence, no transport of measuring rods and clocks from one place to another is necessary to compare lengths and times at different locations.

The gravitational potential Φ_0, in terms of which the fundamental units have been defined, is to be considered a universal constant, and the same applies to $g(\Phi_0)$. Thus the universal constant of the theory of gravitation developed here does not appear in the field eqs. (19), (20), but plays a part in the definition of the fundamental units.

7. THE EQUATIONS OF MOTION OF A BODY WHICH MAY BE TREATED AS A MASS POINT

In order to obtain the equations of motion of a material point, and additionally to gain clear insight into the conditions under which a body may be viewed as a material point, we consider a body that has the properties of the complete stationary system discussed in §1, moving freely in an external homogeneous gravitational field. We have

$$\Phi = \Phi_1 + \Phi_2,$$

where

$$\Phi_1 = -\frac{1}{4\pi}\int\frac{dv}{r}\{g(\Phi)\nu\}_{t-\frac{r}{c}}.$$

The gravitational potential Φ_2 of the external field does not need to be constant in time as long as $\partial\Phi_2/\partial t$ is constant. | We thus have [p. 551]

$$\Phi_2 = ax + by + fz + hu,$$

where the coefficients $a \dots h$ are constant in space and time.

According to the general foundations of relativistic mechanics, one has (compare § 1)

$$\left.\begin{aligned} &\frac{\partial}{\partial x}(\boldsymbol{G}_{xx} + \boldsymbol{T}_{xx} + \boldsymbol{L}_{xx}) + \frac{\partial}{\partial y}(\boldsymbol{G}_{xy} + \boldsymbol{T}_{xy} + \boldsymbol{L}_{xy}) \\ &+ \frac{\partial}{\partial z}(\boldsymbol{G}_{xz} + \boldsymbol{T}_{xz} + \boldsymbol{L}_{xz}) + \frac{\partial}{\partial u}(\boldsymbol{G}_{xu} + \boldsymbol{T}_{xu} + \boldsymbol{L}_{xu}) = 0, \end{aligned}\right\} \quad (a)$$

and another three equations obtained by interchanging the first index x with y, z, u. We integrate the expression (a) over all of xyz-space, and for the time being deal with the gravitation tensor $\boldsymbol{G}$ separately. According to (7) and (3) we have:

$$\left.\begin{aligned} &\int\left\{\frac{\partial\boldsymbol{G}_{xx}}{\partial x} + \frac{\partial\boldsymbol{G}_{xy}}{\partial y} + \frac{\partial\boldsymbol{G}_{xz}}{\partial z} + \frac{\partial\boldsymbol{G}_{xu}}{\partial u}\right\}dv \\ &= \int g(\Phi)\nu\frac{\partial\Phi_1}{\partial x}dv + \frac{\partial\Phi_2}{\partial x}\int g(\Phi)\nu dv. \end{aligned}\right\} \quad (b)$$

We wish to transform the first integral on the right. Since the second derivatives of Φ_2 are all zero, in the eq. (2) on the left we can set Φ_1 instead of Φ. If we insert the expression so obtained for $g(\Phi)\nu$, we obtain, after a transformation similar to the one that leads from eq. (3) to (7),

$$\int g(\Phi)\nu\frac{\partial\Phi_1}{\partial x}dv = \int\left\{\frac{\partial \boldsymbol{G}_{xx}{}^1}{\partial x}+\frac{\partial \boldsymbol{G}_{xy}{}^1}{\partial y}+\frac{\partial \boldsymbol{G}_{xz}{}^1}{\partial z}\right\}dv+\int\frac{\partial \boldsymbol{G}_{xu}{}^1}{\partial u}dv.$$

Here, $\boldsymbol{G}_{xx}{}^1$ etc. are the expressions one obtains by writing Φ_1 instead of Φ in the eqs. (6). We can transform the first integral on the right in the last equation into a surface integral over an infinitely large spherical surface using Gauss' theorem, and this integral becomes zero because expressions corresponding to eq. (5) apply for the first derivatives of Φ_1. Therefore, we obtain

$$\int g(\Phi)\nu\frac{\partial\Phi_1}{\partial x}dv = \frac{d}{du}\int\frac{\partial\Phi_1}{\partial x}\frac{\partial\Phi_1}{\partial u}dv. \tag{c}$$

[p. 552] | Upon integration of (a) the two remaining world tensors $\boldsymbol{T}$ and $\boldsymbol{L}$ give the result

$$\frac{d}{du}\int\{\boldsymbol{T}_{xu}+\boldsymbol{L}_{xu}\}dv,$$

since the remaining terms can also be transformed into a surface integral that becomes zero. Therefore the integration of (a) over all of space yields the following result, taking (b) and (c) into account,

$$-\frac{\partial\Phi_2}{\partial x}\int g(\Phi)\nu dv = \frac{d}{du}\int\left\{\frac{\partial\Phi_1}{\partial x}\frac{\partial\Phi_1}{\partial u}+\boldsymbol{T}_{xu}+\boldsymbol{L}_{xu}\right\}dv. \tag{d}$$

Of course

$$\mathfrak{G}_x = -\frac{i}{c}\int\left\{\frac{\partial\Phi_1}{\partial x}\frac{\partial\Phi_1}{\partial u}+\boldsymbol{T}_{xu}+\boldsymbol{L}_{xu}\right\}dv$$

is the x-component of the total momentum of the moving body. If the motion is *quasi-stationary*,[13] which we must now assume, we can calculate $\mathfrak{G}_x$ using the same formula that applies to uniform motion of the body, that is (compare p. 536)

$$\mathfrak{G}_x = \frac{m\mathfrak{v}_x}{\sqrt{1-\mathfrak{q}^2}} = m\mathfrak{V}_x.$$

m is the inertial mass of the body. If the body rotates or if stationary motions occur in its interior, then the three-dimensional velocity $\mathfrak{v}$ and the four-dimensional velocity vector $\mathfrak{V}$ relate to the body as a whole, according to §1, and therefore give the center of mass's change of position. We must assume further that the body is of such moder-

13 Compare M. Abraham, *Theorie der Elektrizität II*. Leipzig 1905, p. 183.

ate dimensions that one can view $g(\Phi_2)$ as spatially constant within the body, which is practically always the case. We then have from (11)

$$g(\Phi_2)m = \int g(\Phi)\nu dv_0 = \frac{1}{\sqrt{1-\mathfrak{q}^2}}\int g(\Phi)\nu dv,$$

because eq. (11) relates to a system of reference in which the velocity $\mathfrak{v}$ is momentarily zero, and for | a volume element dv_0 in this system of reference we have [p. 553]

$$dv_0 = \frac{dv}{\sqrt{1-\mathfrak{q}^2}}.$$

Since furthermore

$$d\tau = dt\sqrt{1-\mathfrak{q}^2} = \frac{du}{ic}\sqrt{1-\mathfrak{q}^2},$$

where τ denotes the proper time of the body, we finally obtain from (d)

$$-g(\Phi_2)m\frac{\partial\Phi_2}{\partial x} = \frac{d}{d\tau}m\mathfrak{V}_x.$$

This is the first of the equations of motion for the body; naturally one obtains the remaining three by exchanging y, z, u for x.

We have assumed that the external field is homogeneous. If this is not the case, then the equations are only valid to the degree of accuracy to which $\nabla\Phi_2$ and $\partial\Phi_2/\partial t$ are spatially constant within the body. (The field outside of the body can not, of course, act on the body itself.) To this degree of accuracy, we can consider the body as a material point, and thus the following equations of motion apply to it, according to the above considerations:

$$\left.\begin{aligned} -g(\Phi_a)m\frac{\partial\Phi_a}{\partial x} &= \frac{d}{d\tau}m\mathfrak{V}_x, \\ &\cdots\cdots\cdots \\ &\cdots\cdots\cdots \\ -g(\Phi_a)m\frac{\partial\Phi_a}{\partial u} &= \frac{d}{d\tau}m\mathfrak{V}_u, \end{aligned}\right\} \tag{30}$$

where the gravitational potential of the external field is denoted by Φ_a. This system of equations are equivalent to the eqs. (51), loc. cit.

If we multiply the eqs. (30) by $\mathfrak{V}_x, \mathfrak{V}_y, \mathfrak{V}_z, \mathfrak{V}_u$ and add them, we find the law for the change of the inertial mass in exactly the same manner as on p. 516, loc. cit. [p. 873 in the original], and we arrive at the formulas already derived in §2. Using these formulas, we can bring the equations of motion (30) into the following form: |

[p. 554]

$$\left.\begin{aligned} -c^2\frac{\partial}{\partial x}\ln\Phi_a{}' &= \frac{d\mathfrak{B}_x}{d\tau}+\mathfrak{B}_x\frac{d}{d\tau}\ln\Phi_a{}', \\ -c^2\frac{\partial}{\partial y}\ln\Phi_a{}' &= \frac{d\mathfrak{B}_y}{d\tau}+\mathfrak{B}_y\frac{d}{d\tau}\ln\Phi_a{}', \\ -c^2\frac{\partial}{\partial z}\ln\Phi_a{}' &= \frac{d\mathfrak{B}_z}{d\tau}+\mathfrak{B}_z\frac{d}{d\tau}\ln\Phi_a{}', \\ -c^2\frac{\partial}{\partial u}\ln\Phi_a{}' &= \frac{d\mathfrak{B}_u}{d\tau}+\mathfrak{B}_u\frac{d}{d\tau}\ln\Phi_a{}'. \end{aligned}\right\} \tag{31}$$

One sees from this that the motion of a body in a gravitational field is completely independent of the constitution of the body, if only it may be treated as a material point.

In a static, homogeneous field in particular, every quasi-stationarily moving body that satisfies the conditions in §1 may be treated as a material point, and thus all such bodies fall in the same manner. In the case of free fall, eqs. (64), loc. cit. apply; it is only to be noted that g depends on Φ, and that Φ denotes the potential of the external field. Since a body rotating about its axis of symmetry satisfies the conditions in §1, it must fall exactly like a non-rotating body. Therefore the assertion about rotating bodies made on p. 520, loc. cit. [p. 878 in the original] does not apply in our present theory; furthermore, molecular motions have no influence on the falling motion. In contrast, a body thrown horizontally falls slower than one which does not have an initial velocity, as demanded by eqs. (64), loc. cit.

Systems in an external field that do not satisfy the conditions in §1 generally move approximately according to the equations of motion (30), (31). For example, Mr. Einstein has shown that according to the theory developed here, an elastically oscillating system's gravitational acceleration must change with the phase of the oscillation, but that the mean acceleration is given by (64), loc. cit.

ALBERT EINSTEIN

ON THE PRESENT STATE OF THE PROBLEM OF GRAVITATION

Originally published as "Zum gegenwärtigen Stande des Gravitationsproblems" in Physikalische Zeitschrift 14 (1913), pp. 1249–1262. Reprinted in "The Collected Papers of Albert Einstein," Vol. 4, Doc. 17: An English translation is given in its companion volume.

1. GENERAL FORMULATION OF THE PROBLEM

The first domain of physical phenomena where a successful theoretical elucidation was achieved was that of the general attraction of masses. The laws of weight and of the motions of celestial bodies were reduced by Newton to a simple law of motion for a mass point and to a law of interaction for two gravitating mass points. These laws have proved to hold so exactly that, from an empirical point of view, there is no decisive reason to doubt their strict validity. If, despite this, one can scarcely find a physicist today who believes in the exact validity of these laws, this is due to the transformative influence of the development of our knowledge of electromagnetic processes over the last few decades.

Before Maxwell, electromagnetic processes were attributed to elementary laws built as closely as possible on the model of Newton's force law. According to these laws, electrical masses, magnetic masses, current elements, and so on, are supposed to exert actions-at-a-distance on each other, which require no time for propagation through space. Then 25 years ago, Hertz showed with his brilliant experimental investigation of the propagation of electrical force that electrical effects require time for their propagation. By doing so he contributed to the victory of Maxwell's theory, which replaced unmediated action-at-a-distance with partial differential equations. Following this demonstration of the untenability of action-at-a-distance theory in the area of electrodynamics, confidence in the correctness of Newton's action-at-a-distance gravitational theory was also | shaken. The conviction that Newton's law of [1250]
gravitation encompasses as little of the totality of gravitational phenomena as Coulomb's law of electrostatics and magnetostatics captures of the totality of electromagnetic phenomena had to come to light. Newton's law previously sufficed for calculating the motions of the celestial bodies due to the small velocities and accelerations of those motions. In fact, it is easy to demonstrate that the motion of celestial

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 3
Gravitation in the Twilight of Classical Physics: Between Mechanics, Field Theory, and Astronomy.

bodies determined by electrical forces acting on electrical charges they bear would not reveal Maxwell's laws of electrodynamics to us if their velocities and accelerations were of the same order of magnitude as in the motions of the celestial bodies with which we are familiar. One would be able to describe such motions with great accuracy on the basis of Coulumb's law.

Even though confidence in the comprehensiveness of Newton's action-at-a-distance law was thus shaken, there were still no direct reasons to force an extension of Newton's theory. However, today there is such a direct reason for those who adhere to the correctness of relativity theory. According to relativity theory, in nature there is no means that would permit us to send signals with a velocity greater than that of light. Yet on the other hand, it is clear that if Newton's law were strictly valid, we would be able to use gravitation to send instantaneous signals from a place A to a distant place B; since the motion of a gravitating mass at A would lead to simultaneous changes of the gravitational field, B– in contradiction to relativity theory.

The theory of relativity not only forces us to modify Newton's theory, but fortunately it also strongly constrains the possibilities for such a modification. If this were not the case, the attempt to generalize Newton's theory would be a hopeless undertaking. To see this clearly, one need only imagine being in the following analogous situation: suppose that of all electromagnetic phenomena, only those of electrostatics are known experimentally. Yet one knows that electrical effects cannot propagate with superluminal velocity. Who would have been able to develop Maxwell's theory of electromagnetic processes on the basis of these data? Our knowledge of gravitation corresponds precisely to this hypothetical case: we only know the interaction between masses at rest, and probably only in the first approximation. Relativity theory limits the bewildering manifold of possible generalizations of the theory, because according to it in every system of equations the time coordinate appears in the same manner as the three spatial coordinate, up to a difference in sign. This formal insight of Minkowski's, which is here only roughly foreshadowed, has been a tool of utmost importance in the search for equations compatible with relativity theory.

2. PLAUSIBLE PHYSICAL HYPOTHESES CONCERNING THE GRAVITATIONAL FIELD

In what follows we shall specify several general postulates, which can be employed by a gravitational theory, although it need not employ all of them:

1. Satisfaction of the laws of energy and momentum conservation.
2. Equality of the *inertial* and the *gravitational* mass for isolated systems.
3. Validity of the theory of relativity (in the restricted sense); i.e., the systems of equations are covariant with respect to linear orthogonal substitutions (generalized Lorentz transformations).
4. The observable laws of nature do not depend on the absolute magnitude of the gravitational potential (or gravitational potentials). Physically, this means the following: The embodiment of relations between observable quantities that one can

determine in a laboratory is not changed if I bring the whole laboratory into a region with a different (spatially and temporally constant) gravitational potential.

We note the following regarding these postulates. All theorists agree with one another that postulate 1 must be upheld. There is not such a general consensus regarding adherence to postulate 3. Thus, M. Abraham has developed a gravitational theory that does not comply with postulate 3. I could subscribe to this point of view, if Abraham's system were covariant with respect to transformations that turn into linear orthogonal transformations in regions of constant gravitational potential, but this does not appear to be the case with Abraham's theory. | Therefore this theory does not [1251] contain relativity theory, as previously developed without connection with gravitation, as a special case. All of the arguments that have been put forward in favor of relativity theory in its current form militate against such a theory. In my opinion, it is absolutely necessary to hold fast to postulate 3 as long as there are no compelling reasons against doing so; the moment we give up this postulate, the manifold of possibilities will become impossible to survey.

Postulate 2 calls for a more precise examination, and, in my opinion, we must hold on to it unconditionally until there is proof to the contrary. The postulate is initially based on the fact of experience that all bodies fall with the same acceleration in a gravitational field; we will have direct our attention to this important point again later on. Here it should only be said that the equality (proportionality) of gravitational and inertial mass was proved with great accuracy by Eötvös's investigation,[1] which is of highest significance to us; he proved this proportionality by establishing experimentally that the resultant of weight and of the centrifugal force due to the Earth's rotation is independent of the nature of the material (the relative difference between the two masses is $<10^{-7}$). In combination with one of the main results of the ordinary relativity theory, postulate 2 leads to a consequence that can already be drawn at this point. According to relativity theory, the inertial mass of a closed system (treated as a whole) is determined by its energy. From postulate 2, the same must also hold for *gravitational* mass. Therefore, if the state of the system undergoes an arbitrary change without altering its total energy, then the gravitational action-at-a-distance does not change, even if a part of the system's energy is converted into gravitational energy. The gravitational mass of a system is fixed by its total energy, including gravitational energy.

Finally, postulate 4 arguably cannot be grounded on experience. It is only justified by our confidence in the simplicity of the laws of nature, and we cannot have as much right to depend on it as we do with the three axioms named above.

I am fully aware that the postulates 2–4 resemble a scientific profession of faith more than a firm foundation. I am also far from claiming that the two generalizations of Newton's theory presented in the following are the only ones possible, but I dare say that given the current state of our knowledge they must be the *most natural* ones.

---

1 B. Eötvös, *Mathem. und naturw. Ber. aus Ungarn* 8, 1890; Supplement 15: 688, 1891.

3. NORDSTRÖM'S THEORY OF GRAVITATION

According to the familiar relativity theory in connection with gravitational theory, an isolated material point moves uniformly in a straight line in accord with Hamilton's equation

$$\delta\left\{\int d\tau\right\} = 0, \tag{1}$$

where we have set, in the usual way,

$$\left.\begin{aligned} d\tau &= \sqrt{-dx_1^2 - dx_2^2 - dx_3^2 - dx_4^2} = \\ &= \sqrt{c^2dt^2 - dx^2 - dy^2 - dz^2} = dt\sqrt{c^2 - q^2}. \end{aligned}\right\} \tag{2}$$

We can also write equation (1) as

$$\left.\begin{aligned} &\delta\left\{\int H dt\right\} = 0 \\ &\text{where} \\ &H = -m\frac{d\tau}{dt} = -m\sqrt{c^2 - q^2} \end{aligned}\right\} \tag{1a}$$

is the Lagrangian function of the moving point, and m is a constant characteristic of it, its "mass." The momentum (I_x, I_y, I_z) and the energy E of the point follows directly from this, as Planck has shown, in the familiar way.[2]

$$I_x = \frac{\partial H}{\partial \dot{x}} = m\frac{\dot{x}}{\sqrt{c^2 - q^2}}$$

$$E = \frac{\partial H}{\partial x}\dot{x} + \frac{\partial H}{\partial y}\dot{y} + \frac{\delta H}{\delta z}\dot{z} - H = m\frac{c^2}{\sqrt{c^2 - q^2}}.$$

From here it is easy to arrive at Nordström's theory if we make the following assumptions. The covariance of the equation with respect to linear orthogonal substitutions still stands, as is the case in the familiar relativity theory. The gravitational field can be described as a scalar. The motion of a material point in the gravitational field can be represented with an equation of Hamiltonian form. In that case one obtains the following equation for the motion for a mass point:[3] |

---

2 These expressions differ from the customary ones only by the constant factor $1/c$.

3 Taking into consideration the fact that the Hamiltonian integral must be an invariant.

$$\delta\left\{\int \varphi d\tau\right\} = 0, \tag{1'}$$ [1252]

such that (2) with constant c remains valid and φ is the scalar fixed by the gravitational field. For the propagation of this light ray we have $dt = 0$, and so $q = c$; i.e., the speed of light propagation is equal to the constant c. The light rays are not bent by the gravitational field.

In the place of equations (1a) we have

$$\left.\begin{aligned} &\delta\left\{\int H d\tau\right\} = 0, \\ &\text{whereupon} \\ &H = -m\varphi\frac{d\tau}{dt} = -m\varphi\sqrt{c^2-q^2}. \end{aligned}\right\} \tag{1a'}$$

The Lagrangian equations of motion read:

$$\frac{d}{dt}\left\{m\varphi\frac{\dot{x}}{\sqrt{c^2-q^2}}\right\} + m\frac{\partial\varphi}{\partial x}\sqrt{c^2-q^2} = 0 \text{ etc.}$$ [1]

From this it follows that the expressions for the impulse, energy, and the force $\mathfrak{K}$ exerted by the gravitational field at a point are:

$$\left.\begin{aligned} I_x &= m\varphi\frac{\dot{x}}{\sqrt{c^2-q^2}} \text{ etc.} \\ E &= m\varphi\frac{c^2}{\sqrt{c^2-q^2}} \\ \mathfrak{K}_x &= -m\frac{\partial\varphi}{\partial x}\sqrt{c^2-q^2} \text{ etc.} \end{aligned}\right\} \tag{2a}$$

Thus m is a constant characteristic of the mass point, independent of φ and q. The expression for $\mathfrak{K}$ shows that φ plays the role of the gravitational potential. Furthermore, the expressions for I_x and E show that according to Nordström's theory the inertia of a mass point is determined by the product $m\varphi$; the smaller φ is, i.e., the more mass we pile up in the region of the mass point, the smaller the inertial resistance the body exerts in response to a change in its velocity becomes. This is one of the most important physical consequences of the scalar theory of gravitation, to which we must return later.

In this theory, as well as in the theory explained below, the coordinate differences do not have as simple a physical meaning as they do in the usual relativity theory. Let us consider a given moveable unit measuring rod and a moveable clock, which ticks such that in a vacuum light traverses a distance equal to one unit measuring rod[4] during one unit of time, as measured by the clock. We will call the four-dimensional interval between two infinitely close spacetime points, which can be measured with these measuring tools in the same way as in the usual relativity theory, the "natural" four-dimensional interval $d\tau_0$ of the spacetime point. By definition this is an invariant, and hence in the case of the usual relativity theory it is equal to $d\tau$. We call the latter quantity the "coordinate interval," in contrast to the natural interval and according to its definition, or also briefly as the "interval" of the spacetime point. In our case it is possible, however, that the natural interval $d\tau_0$ differs from the coordinate interval $d\tau$ by a factor that is a function of φ. Thus we set

$$d\tau_0 = \omega d\tau. \tag{3}$$

We can further speak of the natural length l_0 and the natural volume V_0 of a body. These are the length and volume, respectively, that are measured using comoving unit measuring rods. The lengths l and volumes V measured in coordinates also play a role. It follows that the relation between the coordinate volume V and the natural volume V_0 is:

$$\frac{1}{V} = \frac{\omega^3 c dt}{V_0 \, d\tau} = \frac{\omega^3 c}{V_0\sqrt{c^2 - q^2}}. \tag{4}$$

In addition, by a unit mass we understand the mass of water enclosed in a natural volume of magnitude unity. The mass of a body is the ratio of its inertia to that of a unit mass, which is thus a scalar.We understand the natural density ρ_0 to be relative to the density of water or the mass in a natural volume with magnitude 1; ρ_0 is thus a scalar by definition.

We can derive further consequences from the results obtained above if we pass from material points to the continuum. We achieve this by treating the material point as a continuum of coordinate volume V and natural volume V_0. One multiplies the expressions for I_x, E and $\mathfrak{K}_x$ given above in (2a) by $1/V$, using (4), so that one obtains the impulse i_x etc., the energy η, and the pondermotive force $\mathfrak{k}_x$ etc., per unit volume for an incoherent mass stream. Taking the relation

$$\rho_0 = \frac{m}{V_0}$$

[1253] into account, one obtains |

4 We will make the assumption that this is achievable at all locations and at all times; this is a special case of postulate (4).

$$\left.\begin{aligned} ici_x &= \frac{icI_x}{V} = \rho_0 c\varphi\omega^3 \frac{dx_1}{d\tau}\frac{dx_4}{d\tau} \\ -\eta &= -\frac{E}{V} = \rho_0 c\varphi\omega^3 \frac{dx_4}{d\tau}\frac{dx_4}{d\tau} \\ \mathfrak{k}_x &= \frac{\mathfrak{K}_x}{V} = -\rho_0 c\omega^3 \frac{\partial\varphi}{\partial x}. \end{aligned}\right\} \tag{2b}$$

In the first equations i denotes the unit imaginary number. We now recall the expressions for the law of energy-momentum conservation in relativity theory. If X_x and etc. are the generalized pressure-stresses, and f_x etc. are the components of the energy flux density, then the quantities

$$\begin{matrix} X_x & X_y & X_z & ici_x \\ Y_x & Y_y & Y_z & ici_y \\ Z_x & Z_y & Z_z & ici_z \\ \frac{i}{c}f_x & \frac{i}{c}f_y & \frac{i}{c}f_z & -\eta \end{matrix}$$

form a symmetric tensor, that we will write $T_{\mu\nu}$ (μ and ν are indices running from 1 to 4). Furthermore, denoting the work transferred by external forces to the material per unit volume with $\mathfrak{l}$, then

$$\mathfrak{k}_x, \mathfrak{k}_y, \mathfrak{k}_z, \frac{i}{c}\mathfrak{l}$$

is a four vector, with its components referred to by k_μ. The law of energy-momentum conservation is then expressed by the equation

$$\sum_\nu \frac{\partial T_{\mu\nu}}{\partial x_\nu} = k_\mu \qquad (\mu = 1 \text{ to } \mu = 4). \tag{5}$$

As equations (2b) illustrate, this schema can find direct application in our case of an incoherent flow of matter in a gravitational field, insofar as one sets

$$\left.\begin{aligned} T_{\mu\nu} &= \rho_0 c\varphi\omega^3 \frac{dx_\mu}{dt}\frac{dx_\nu}{d\tau} \\ k_\mu &= -\rho_0 c\omega^3 \frac{\partial\varphi}{dx_\mu}. \end{aligned}\right\} \tag{5a}$$

So far we have treated only the question of how the gravitational field acts on matter, but not the question of by which law, in turn, the matter determines the gravitational field. According to Nordström's theory, the latter is given by a scalar φ; thus, what enters into the differential equation for φ we are seeking must also be a scalar associated with the field generating process. This scalar can only be the scalar

$$\sum_\sigma T_{\sigma\sigma}$$

whose existence and meaning was notably highlighted by von Laue. Setting up this scalar for the case of a case of an incoherent mass stream, we obtain with the help of (5a)

$$\sum_\sigma T_{\sigma\sigma} = -\rho_0 c\varphi\omega^3$$

$$k_\mu = \sum_\sigma T_{\sigma\sigma} \cdot \frac{1}{\varphi}\frac{\partial\varphi}{\partial x_\mu}.$$

Thus instead of (5) we have

$$\sum_\nu \frac{\partial T_{\mu\nu}}{\partial x_\nu} = \sum_\sigma T_{\sigma\sigma} \cdot \frac{1}{\varphi}\frac{\partial\varphi}{\partial x_\mu}. \tag{5b}$$

This equation is particularly important in that there is nothing in it to remind us of the case of an incoherent mass stream discussed so far. According to Nordström's theory, equation (5b) expresses the energy balance of an arbitrary material process, if the stress-energy tensor corresponding to this process is substituted for $T_{\mu\nu}$.

From equation (5b) it follows that Nordström's theory satisfies postulate 2. If one were to observe a system on such a small scale that one could regard the $\partial \lg\varphi / \partial x_\mu$ as clearly constant for the spatial extent of the system, then one obtains for the total force exerted on the system by the gravitational field in the X-direction:

$$\frac{\partial \lg\varphi}{\partial x_\mu}\int\sum T_{\sigma\sigma} dv = \frac{\partial \lg\varphi}{\partial x_\mu}\int T_{44} dv = -\frac{\partial \lg\varphi}{\partial x_\mu}\int \eta\, dv,$$

where dv is the three-dimensional volume element. This reformulation is based on Laue's theorem, that for a closed system

$$\int \Sigma T_{11} dv = \int \Sigma T_{22} dv = \int \Sigma T_{33} dv = 0.$$

This proves that what determines the gravity of a closed system is its total quantity.

Equation (5b) further allows us to determine the function φ, which has been left undetermined so far, on the basis of the physical assumption that no work can be extracted from a static gravitational field via cyclic processes. In section 7 of my paper on gravitation published jointly with Mr. Grossmann, I obtained a contradiction between a scalar theory and this basic principle, based, however, on the tacit assumption that $\omega = \text{const.}$ But it is easy to show that the contradiction vanishes if one sets

$$l = \frac{l_0}{\omega} = \frac{\text{const.}}{\varphi}$$

or

$$\omega = \text{const.} \cdot \varphi. \tag{6}$$

| Later we will give yet another justification for this stipulation. [1254]

Now it is simple to establish the general equation for the gravitational field, which is to be regarded as a generalization of Poisson's equation for the gravitational field. That is, one has to set Laue's scalar equal to a scalar differential expression of the quantity φ such that the conservation laws hold for the material process and the gravitational field taken together. One achieves this by setting,

$$-\kappa \Sigma T_{\sigma\sigma} = \varphi \Box \varphi, \tag{7}$$

where κ is a universal constant (the gravitational constant), and $\Box$ denotes the operator

$$\sum_\tau \frac{\partial^2}{\partial x_\tau^2} \quad (\tau \text{ from 1 to 4}).$$

The fact that the conservation laws are satisfied follows from the equations (5b) and (7), by virtue of the identity which follows from (7)

$$\sum T_{\sigma\sigma} \frac{1}{\varphi} \frac{\partial \varphi}{\partial x_\mu} = -\frac{1}{\kappa} \frac{\partial \varphi}{\partial x_\mu} \sum \frac{\partial^2 \varphi}{\partial x_\sigma^2} = -\sum \frac{\partial t_{\mu\nu}}{\partial x_\nu},$$

where we set

$$t_{\mu\nu} = \frac{1}{\kappa} \left\{ \frac{\partial \varphi}{\partial x_\mu} \frac{\partial \varphi}{\partial x_\nu} - \frac{1}{2} \delta_{\mu\nu} \sum \frac{\partial \varphi^2}{\partial x_\tau} \right\}. \tag{8}$$

$\delta_{\mu\nu}$ denotes 1 respectively 0, depending on whether $\mu = \nu$ or $\mu \neq \nu$. The component of the stress-energy tensor of the gravitational field is $t_{\mu\nu}$; then it follows from the penultimate equation and (5b) that

$$\sum_\nu \frac{\partial}{\partial x_\nu} (T_{\mu\nu} + t_{\mu\nu}) = 0. \tag{9}$$

Thus postulate 1 is satisfied. It can also be shown that, in accord with postulate 2, the number of gravitational lines emanating from a closed stationary system to infinity depends on the total energy of the system.

Furthermore, the following is in conformity with postulate 4. If one places two mirrors at the ends of a natural length l_0 facing each other, and allows a light ray to go back and forth between them in a vacuum, then this system represents a clock (light clock). If two masses m_1 and m_2 circle each other at the natural distance l_0 under the influence of their gravitational interaction, then this system also represents a clock (gravitational clock). With the help of the equations derived above, one can easily show that the relative rate of these two clocks, supposing that they are found in the same gravitational potential, is independent of the absolute value of the potential. This is an indirect confirmation of the expression given for ω in equation (6).

In conclusion we can say that Nordström's scalar theory, which holds firmly onto the postulate of the constancy of the speed of light, satisfies all the requirements for a theory of gravitation that can be imposed on the basis of current experience. Only one unsatisfactory circumstance remains, namely that according to this theory the inertia of bodies seems to not be *caused* by other bodies, even though it is *influenced* by them, because the inertia of a body is greater the farther other bodies are from it.

4. IS THE ATTEMPT TO EXTEND RELATIVITY THEORY JUSTIFIED?[5]

If we wish to show a neophyte the extent to which the formulation of relativity theory is empirically justifiable, we can point out the following to him. For a person located in a railway car travelling uniformly in a straight line with its windows covered it is not even possible to decide what direction and at what speed the car travels; if we abstract from the inevitable shaking of the car, it is not even possible to decide whether the car is moving or not. Expressed abstractly: the laws of events described with respect to the system moving uniformly with respect to the original coordinate system (the Earth's surface) are the same as with respect to the original coordinate system (the Earth's surface). We call this proposition the relativity principle for uniform motion.

Yet one might be apt to add: it is surely different if the railway car moves non-uniformly; if the car changes its velocity, the passenger gets a jolt through which he detects the acceleration of the car. Speaking abstractly, there is no relativity principle for nonuniform motion. But concluding in this way is not irreproachable, because it is, after all, not certain whether the occupant of the car must necessarily ascribe the jolt he felt to the acceleration of the wagon. From the following example one sees that this conclusion is premature.

Two physicists, A and B, wake from a drug-induced slumber and discover that [1255] they are in a closed box with opaque walls, equipped with | all of their instruments. The have no idea how the box is situated, and how and whether it is moving. Now they determine that all bodies that they bring to the middle of the box and release fall in the same direction—let's say downward—with the same acceleration γ. What can the physicists conclude from this? A concludes that the box sits still on a celestial body, and that the downward direction must be towards the center of the celestial body, if it is taken to be spherical. But B adopts the point of view that the box could be moving with constant acceleration upward with the acceleration γ; due to an externally applied force, and there need not be a celestial body nearby. Is there a criterion that the two physicists could use to determine who is correct? We do not know of any such criterion, but we also do not know whether there is no criterion. *In any*

---

5 Cf. A. Einstein, *Ann. d. Phys.* (4) 35: 898, 1911.

case, Eötvös's exact experimental result regarding the equality of inertial and gravitational mass supports the view that there is no such criterion. One sees that, in this regard, Eötvös's experiment plays a similar role to that of Michelson's experiment with respect to the physical verifiability of *uniform* motion.

If it is really in principle impossible for the two physicists to decide which of the two views is correct, then *acceleration* has as little absolute physical meaning as *velocity*.[6] The same reference system can be taken to be accelerating or non-accelerating with equal justice, but then, according to the view chosen, one must postulate the presence of a gravitational field that determines the motion of freely moving bodies with respect to the reference system together with the possible acceleration of the system.

The circumstance that bodies behave in exactly the same in what is, according to our view, a nonaccelerated reference system in the presence of a gravitational field, as in an accelerating reference system, forces us to seek an extension of the principle of relativity to the case of accelerating reference systems.

From a mathematical standpoint, this amounts to demanding covariance of the equations expressing laws of nature not only under linear orthogonal substitutions, but also with respect to other, in particular non-linear, transformation; for only the non-linear substitutions correspond to a transformation between relatively *accelerated* systems. But then we face the difficulty that our scant empirical knowledge of the gravitational field permits no reliable deduction of the substitutions for which the covariance of the equations must be demanded. In an investigation undertaken with my friend Grossmann,[7] it turned out that it is possible and expedient to initially demand covariance with respect to *arbitrary* substitutions.

One further comment before proceeding to dispel a natural misunderstanding. An adherent of current relativity theory has some right to call the velocity of a point mass "apparent." In fact he can choose a coordinate system, such that the velocity is zero at the instant in question. But if he is dealing with a system of points whose mass points have different velocities, he cannot introduce a reference system such that all of the velocities of the mass points vanish with respect to it. Analogously, a physicist sharing our point of view can call the gravitational field "apparent," for by a suitable choice of the state of acceleration he can achieve the result that there is no gravitational field present at a given spacetime point. But it is clear that for extended gravitational fields this elimination of the gravitational field by a transformation cannot be achieved, in general. For example, it would not be possible to make the Earth's gravitational field vanish by choosing an appropriate reference system.

6 This point of view will be modified in section 6; but for the time being we will stick with it firmly.

7 *Entwurf einer verallgemeinerten Relativitätstheorie und einer Theorie der Gravitation,* Leipzig: B. G. Teubner, 1913.

5. CHARACTERIZATION OF THE GRAVITATIONAL FIELD; ITS EFFECT ON PHYSICAL PROCESSES

Since we are uncertain about the class of admissible spacetime substitutions, the most natural thing, as already mentioned above, is to consider arbitrary substitutions of the spacetime variables x, y, z, t, which we can more conveniently write as x_1, x_2, x_3, x_4. It turns out to be pointless to introduce an imaginary time coordinate in the case of the generalization considered below.

First we consider a spacetime region, in which there is no gravitational field in an
[1256] appropriately chosen coordinate system. | We are then faced with the case that is familiar from the usual relativity theory. A free mass point moves uniformly and in a straight line according to the equation

$$\delta\left\{\int\sqrt{-dx^2-dy^2-dz^2+c^2dt^2}\right\} = 0.$$

Introducing new coordinates x_1, x_2, x_3, x_4 through an arbitrary substitution, it then follows that the motion of the point relative to the new system obeys the equation

$$\left.\begin{array}{l}\delta\left\{\int ds\right\} = 0 \\ \text{where we set} \\ ds^2 = \sum_{\mu\nu} g_{\mu\nu}dx_\mu dx_\nu.\end{array}\right\} \tag{1b}$$

From this we can also assume that

$$\left.\begin{array}{l}\delta\left\{\int H dt\right\} = 0 \\ \text{where we set} \\ H = -m\dfrac{ds}{dt}.\end{array}\right\} \tag{1b'}$$

H is the Hamiltonian function.

In the new system the quantities $g_{\mu\nu}$, determine the motion of the mass point, which according to the general observations of the foregoing section can be conceived of as the components of the gravitational field, as long as we treat the new system as "at rest." In general each gravitational field is defined by the ten components $g_{\mu\nu}$, which are functions of x_1, x_2, x_3, x_4. The motion of material points will always be determined by equations of the given form. Given its physical meaning, the element ds must be an invariant with respect to all substitutions. Through this the transformation laws for the components $g_{\mu\nu}$ is established if the coordinate transformation is given. ds is the only invariant associated with the four-dimensional line element (dx_1, dx_2, dx_3, dx_4). We call it the value or magnitude of the line element. If there is no gravitational field, then given a suitable choice of variables the system of $g_{\mu\nu}$'s reduces to the system

$$\begin{matrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & c^2. \end{matrix}$$

Thus we have come back to the case of the usual relativity theory.

The following equation determines the law for the velocity of light:

$$ds = 0.$$

With this one recognizes that in general the velocity of light depends not only on the spacetime point but also on the direction. The reason why we do not notice anything like this is that in the region of spacetime accessible to us the $g_{\mu\nu}$ are almost constant, and we can choose the reference system such that, up to small deviations, the $g_{\mu\nu}$ will have the constant values given above.

We can speak here of the natural length of a four-dimensional element exactly as in Nordström's theory. This is the element's length as measured by a moveable unit measuring rod and moveable clock. By definition this natural length is a scalar, and must therefore be equal to the magnitude ds up to a constant, which we set to 1. This gives the relation between coordinate differentials, on one hand, and measurable lengths and times, on the other; since they have this dependence on the quantities $g_{\mu\nu}$, the coordinates by themselves have no physical meaning. The stipulations regarding mass and natural density remain applicable without modification

Now we can set up the Lagrangian equations of motion for a material point, just as in our analysis of Nordström's theory, starting with equations (1b) and (1b'). From them we borrow the expressions for the momentum I and the energy E of a mass point, and the force $\mathfrak{K}$ exerted by a gravitational field on the mass point. Just as above, we can derive the corresponding expressions for the unit volume, and we obtain

$$\left.\begin{aligned} i_x &= -\rho_0\sqrt{-g}\sum_{\nu s} g_{1\nu}\frac{dx_\nu}{dx_s}\frac{dx_4}{dx_s} \\ -\eta &= -\rho_0\sqrt{-g}\sum_{\nu s} g_{4\nu}\frac{dx_\nu}{dx_s}\frac{dx_4}{dx_s} \\ \mathfrak{k}_x &= -\frac{1}{2}\rho_0\sqrt{-g}\sum_{\nu s}\frac{dg_{\mu\nu}}{dx_1}\frac{dx_\mu}{dx_s}\frac{dx_\nu}{dx_s}. \end{aligned}\right\} \tag{2c}$$

From this we obtain, as above, the law of energy-momentum conservation for the incoherent mass stream:

$$\left.\begin{aligned} \sum_{\mu\nu}\frac{\partial}{\partial x_\nu}(\sqrt{-g}g_{\sigma\mu}\Theta_{\mu\nu}) - \frac{1}{2}\sum_{\mu\nu}\sqrt{-g}\frac{\partial g_{\mu\nu}}{\partial x_\sigma}\Theta_{\mu\nu} &= 0 \qquad (\sigma = 1, 2, 3, 4) \\ \Theta_{\mu\nu} &= \rho_0\frac{dx_\mu}{ds}\frac{dx_\nu}{ds}. \end{aligned}\right. \tag{5b}$$

Here g denotes the determinant of the $g_{\mu\nu}$. The first three equations of (5b) express the law of momentum conservation, and the last states the law of energy conservation. We can give this system of equations a somewhat more perspicuous form if we introduce the quantities |

[1257]

$$\left.\begin{aligned} &\mathfrak{T}_{\sigma\nu} = \sqrt{-g}g_{\sigma\mu}\Theta_{\mu\nu} \\ &\text{It follows that} \\ &\sum_\nu\frac{\partial\mathfrak{T}_{\sigma\nu}}{\partial x_\nu} = \frac{1}{2}\sum_{\mu\nu\tau}\frac{\partial g_{\mu\nu}}{\partial x_\sigma}\gamma_{\mu\tau}\mathfrak{T}_{\tau\nu}, \end{aligned}\right\} \tag{5c}$$

where $\gamma_{\mu\tau}$ denotes the subdeterminant of $g_{\mu\nu}$ divided by g. The physical meaning of the quantities $T_{\sigma\nu}$ emerges from the following schema:

$$\begin{matrix} \mathfrak{T}_{11} & \mathfrak{T}_{12} & \mathfrak{T}_{13} & \mathfrak{T}_{14} \\ \mathfrak{T}_{21} & \mathfrak{T}_{22} & \mathfrak{T}_{23} & \mathfrak{T}_{24} \\ \mathfrak{T}_{31} & \mathfrak{T}_{32} & \mathfrak{T}_{33} & \mathfrak{T}_{34} \\ \mathfrak{T}_{41} & \mathfrak{T}_{42} & \mathfrak{T}_{43} & \mathfrak{T}_{44} \end{matrix} = \begin{matrix} X_x & Y_y & Z_z & i_x \\ Y_x & Y_y & Y_z & i_y \\ Z_x & Z_y & Z_z & i_z \\ f_x & f_y & f_z & \eta, \end{matrix}$$

where the relations given on the right-hand side have the same meaning as in section 3. The right-hand side of (5c) expresses the momentum $(\delta = 1-3)$ and energy $(\delta = 4)$ given off by the gravitational field per unit volume and time.

The equations (5b) and (5c), without a doubt, have a meaning that extends far beyond the case of incoherent mass streams we have considered; they probably express the energy-momentum balance between a physical process and the gravitational field in general. But for each particular physical domain the quantities $\Theta_{\mu\nu}$ and $\mathfrak{T}_{\mu\nu}$ must be expressed in a specific manner.

6. COMMENTS ON THE MATHEMATICAL METHOD

In the theory we have just sketched the conventional theory of vectors and tensors cannot be applied, since according to it Σdx_ν^2 is not an invariant. The fundamental invariant, which we have called the magnitude of the line element, is rather

$$ds^2 = \Sigma g_{\mu\nu} dx_\mu dx_\nu.$$

The theory of covariance of a four-dimensional manifold defined by its line element has already been developed under the name "absolute differential calculus," by Ricci and Levi-Civita[8] in particular, who based their work primarily on a fundamental paper by Christoffel.[9] One can find a concise account of the most important theorems in the part of our work cited above penned by Mr. Grossmann.

In this theory one distinguishes several kinds of tensors, namely covariant, contravariant, and mixed, which are governed by algebraic rules similar to the well-known case characterized by the Euclidean line element. Differential operators that, when applied to tensors, produce tensors again have also been worked out, so that one can specify algebraic and differential relations for the general line element corresponding to those of the conventional theory of vectors and tensors.

It should be noted that dx_ν is the ν^{th} component of a contravariant tensor of the first rank (i.e., with one index). $g_{\mu\nu}$ and $\gamma_{\mu\nu}$, respectively, are components of a covariant and a contravariant tensor of the second rank, which we call the "fundamental tensor" based on its significance for the line element. $\Theta_{\mu\nu}$ is a second rank contravariant tensor, and $\frac{1}{\sqrt{-g}}\mathfrak{T}_{\sigma\nu}$ is a second rank mixed tensor.

Equation (5b) expresses the vanishing of the "divergence" of the tensor $\Theta_{\mu\nu}$. From this it follows that equation (5b) is covariant with respect to arbitrary substitutions, which naturally must also be demanded from a physical point of view.

By replacing the equations of relativity theory with the corresponding equations by means of the absolute differential calculus, one obtains a system of equations that account for the effect of the gravitational field on the domain of phenomena under consideration. This problem has already been solved by Köttler for the case of electromagnetic processes in a vacuum.[10]

From what has been said it follows that the question of the influence of the gravitational field on physical processes has been satisfactorily solved in principle, and in

---

8 Ricci and Levi-Civita, "Méthodes de calcul différentiel absolu et leurs applications," *Math. Ann.* 54: 125, 1900.

9 Christoffel, "Über Transformation der homogenen Differentialausdrücke zweiten Ranges," *Journ. f. Math.* 70: 46, 1869.

10 Köttler, "Über die Raumzeitlinien der Minkowskischen Welt," *Wien. Ber.* 121, 1912.

such a way that the equations are covariant under arbitrary substitutions. With that the spacetime variables are reduced to intrinsically meaningless, auxiliary variables that can be chosen arbitrarily. The whole problem of gravitation would thus be satisfactorily solved if one could find equations *covariant under arbitrary substitutions* that are satisfied by the quantities $g_{\mu\nu}$ fixed by the gravitational field itself. However, we have not succeeded in solving the problem in this manner.[11] We have obtained a solution by instead subsequently restricting the reference system. One is led to this
[1258] method naturally by the following considerations. It is clear that for any | material process by itself (i.e., without its gravitational field) the conservation theorems for momentum and energy cannot be satisfied. This situation corresponds to the appearance of the term on the right-hand side of (5c). On the other hand, we certainly must demand that the conservation theorems are satisfied for the material process and the gravitational field together. From this it follows that we must demand the existence of an expression $\mathrm{t}_{\sigma\nu}$ for the stress, momentum, and energy flux and energy density of the gravitational field that, together with the corresponding quantity $\mathfrak{T}_{\sigma\nu}$ for the material process, fulfills the relation

$$\sum_{\nu} \frac{\partial(\mathfrak{T}_{\sigma\nu} + \mathrm{t}_{\sigma\nu})}{\partial x_{\nu}} = 0.$$

If $\mathrm{t}_{\sigma\nu}$ should have the same character as $\mathfrak{T}_{\sigma\nu}$, according to the theory of invariants, then the left-hand side of this equation cannot be covariant under arbitrary transformations; it is probably so only with respect to arbitrary *linear* transformations.

Therefore by demanding the validity of the conservation theorems, we restrict the reference systems to a great extent, and thereby relinquish the construction of gravitational equations in generally covariant form.

Thus, here is where the limit of applicability of the arguments given in section 4 lies. If one begins with a reference system with respect to which the conservations laws in the form given above hold and introduces a new reference system through an acceleration transformation, then with respect to the latter the conservation theorems are no longer satisfied. Nevertheless, I believe that the equations derived on the basis of the considerations in section 1 do not lose their footing because of this. On the one hand, it is certainly possible to describe the processes with respect to arbitrary reference systems; on the other hand, I do not see how the specialization of the reference system introduced here could bring about the specialization of the equations.

---

11 A short time ago I found a proof to the effect that such a generally covariant solution cannot exist at all.

7. SYSTEM OF EQUATIONS FOR THE GRAVITATIONAL FIELD

The sought-after system of equations should be a generalization of Poisson's equation

$$\Delta\varphi = 4\pi k\rho .$$

Since in our theory the gravitational field is determined by the 10 quantities $g_{\mu\nu}$ in place of φ, we will obtain 10 equations in place of this *one*. By the same token, $\Theta_{\mu\nu}$ appears on the right-hand side of the field equations as the field source instead of ρ, so that the sought-after equation will be of the form

$$\Gamma_{\mu\nu} = \kappa\Theta_{\mu\nu}.$$

$\Gamma_{\mu\nu}$ is a differential expression built up from the quantities $g_{\mu\nu}$, from which we know that it must be covariant with respect to linear transformations. I further assume that $\Gamma_{\mu\nu}$ does not contain anything higher than second derivatives. Furthermore, the conservation theorem necessitates the following: if we replace the second term of (5b) $\Theta_{\mu\nu}$ with $(1/\kappa)\Gamma_{\mu\nu}$, then we must allow this term to be transformed such that, like the first term of (5b), it can be written as a sum of derivatives. So far as I can see, these considerations gave me a unique way of identifying the $\Gamma_{\mu\nu}$ and hence the sought-after equations. These read:

$$\Delta_{\mu\nu}(\gamma) = \kappa(\Theta_{\mu\nu} + \vartheta_{\mu\nu}), \tag{7a}$$

where we set

$$\Delta_{\mu\nu}(\gamma) = \sum_{\alpha\beta}\frac{1}{\sqrt{-g}}\frac{\partial}{\partial x_\alpha}\left(\gamma_{\alpha\beta}\sqrt{-g}\frac{\partial\gamma_{\mu\nu}}{\partial x_\beta}\right) - \sum_{\alpha\beta\tau\rho}\gamma_{\alpha\beta}g_{\tau\rho}\frac{\partial\gamma_{\mu\tau}}{\partial x_\alpha}\frac{\partial\gamma_{\nu\rho}}{\partial x_\beta}$$

and

$$-2\kappa\vartheta_{\mu\nu} = \sum_{\alpha\beta\tau\rho}\left(\gamma_{\alpha\mu}\gamma_{\beta\nu}\frac{\partial\gamma_{\tau\rho}}{\partial x_\alpha}\frac{\partial\gamma_{\tau\rho}}{\partial x_\beta} - \frac{1}{2}\gamma_{\mu\nu}\gamma_{\alpha\beta}\frac{\partial g_{\tau\rho}}{\partial x_\alpha}\frac{\partial\gamma_{\tau\rho}}{\partial x_\beta}\right). \text{[2]}$$

The energy-momentum equation for material process and the gravitational field together assume the form

$$\sum\frac{\partial}{\partial x_\nu}\{\sqrt{-g}g_{\sigma\mu}(\Theta_{\mu\nu} + \vartheta_{\mu\nu})\} = 0. \tag{9a}$$

From (9a) one sees that $\vartheta_{\mu\nu}$ plays the same role for the gravitational field that $\Theta_{\mu\nu}$ plays for material processes. $\vartheta_{\mu\nu}$ is a covariant tensor with respect to linear transformations, and we will call it the stress-energy tensor of the gravitational field. In accord with postulate 2, $\vartheta_{\mu\nu}$ appears like $\Theta_{\mu\nu}$, as a field-generating cause.

The equations become simpler when one introduces the stress components themselves in the equations:

$$\mathfrak{T}_{\sigma\nu} = \sqrt{-g}g_{\sigma\mu}\Theta_{\mu\nu}$$

and

$$\mathfrak{t}_{\sigma\nu} = \sqrt{-g}g_{\sigma\mu}\vartheta_{\mu\nu}.$$

The equations then take the form: |

[1259]
$$\sum_{\alpha\beta\mu}\frac{\partial}{\partial x_\alpha}\left(\sqrt{-g}\gamma_{\alpha\beta}g_{\sigma\mu}\frac{\partial\gamma_{\mu\nu}}{\partial x_\beta}\right) = \kappa(\mathfrak{T}_{\sigma\nu} + \mathfrak{t}_{\sigma\nu}) \tag{7b}$$

$$-2\kappa\mathfrak{t}_{\sigma\nu} = \sqrt{-g}\left(\sum_{\beta\tau\rho}\gamma_{\beta\nu}\frac{\partial g_{\tau\rho}}{\partial x_\sigma}\frac{\partial\gamma_{\tau\rho}}{\partial x_\beta} - \frac{1}{2}\sum_{\alpha\beta\tau\rho}\partial_{\sigma\nu}\gamma_{\alpha\beta}\frac{\partial g_{\tau\rho}}{\partial x_\alpha}\frac{\partial\gamma_{\tau\rho}}{\partial x_\beta}\right).$$

Then the conservation theorem assumes the form

$$\sum\frac{\partial}{\partial x_\nu}(\mathfrak{T}_{\sigma\nu} + \mathfrak{t}_{\sigma\nu}) = 0. \tag{9b}$$

Equation (7b) allows us to conclude that the equations obtained above satisfy postulate 2.[12]

8. THE NEWTONIAN GRAVITATIONAL FIELD

The gravitational equations we have established are certainly very complicated. But several important consequences can be easily derived from them based on the following considerations. If the usual relativity theory in its familiar form were exactly correct, the components of $g_{\mu\nu}$ respectively $\gamma_{\mu\nu}$ would be given by the following tables:

Table of the $g_{\mu\nu}$

$$\begin{matrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & c^2 \end{matrix}$$

Table of the $\gamma_{\mu\nu}$

$$\begin{matrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & \frac{1}{c^2} \end{matrix}$$

The gravitational field equations do not allow that the components of the fundamental tensor could actually have these values in a finite region, if some physical process occurs in it. However, it appears that the departures of the tensor components from the given constant values can be taken to be very small quantities for the region

12 Because from equation (7b) one can see, for example, that the quantities $\mathfrak{t}_{\sigma\nu}$ of the gravitational field, which play the same role for this field that the quantities $\mathfrak{T}_{\sigma\nu}$ do for the material process, have the same field-inducing effect as the quantities $\mathfrak{T}_{\sigma\nu}$, in conformity with postulate (2).

of the world accessible to us. We obtain a far-reaching approximation if we take these deviations, which we will write with $g^*_{\mu\nu}$ and respectively $\gamma^*_{\mu\nu}$, along with their derivatives into consideration only when they enter linearly, and disregard all terms in which two such quantities are multiplied together. Then the equations (7a) and (7b) assume the form:

$$\Box g^*_{\mu\nu} = \frac{\partial^2 g^*_{\mu\nu}}{\partial x^2} + \frac{\partial^2 g^*_{\mu\nu}}{\partial y^2} + \frac{\partial^2 g^*_{\mu\nu}}{\partial z^2} - \frac{1}{c^2}\frac{\partial^2 g^*_{\mu\nu}}{\partial t^2} = \kappa T_{\mu\nu}, \tag{7c}$$

where the $T_{\mu\nu}$ gives an incoherent mass flow according to the schema

$$\left.\begin{array}{cccc} \frac{\rho_0}{c^2-q^2}\dot{x}\dot{x} & \frac{\rho_0}{c^2-q^2}\dot{x}\dot{y} & .. & -\frac{\rho_0 c^2}{c^2-q^2}\dot{x} \\ \frac{\rho_0}{c^2-q^2}\dot{y}\dot{x} & .. & .. & -\frac{\rho_0 c^2}{c^2-q^2}\dot{y} \\ .. & .. & .. & -\frac{\rho_0 c^2}{c^2-q^2}\dot{z} \\ -\frac{\rho_0 c^2}{c^2-q^2}\dot{x} & .. & .. & \frac{\rho_0 c^4}{c^2-q^2}. \end{array}\right\}. \tag{8}$$

We obtain the Newtonian system insofar as we introduce the following approximations:

1. Only the mass flow is regarded as the field source.
2. The influence of the velocity of the field-generating masses is neglected, and hence the field is treated as static.
3. The velocity and acceleration components in the equations of motion of a material point are treated as small quantities, and only quantities of the lowest order are retained.

Finally, we also have to assume that at infinity the $g^*_{\mu\nu}$ vanish.

It then follows from (7c) and (8) that, if we write the Laplacian operator as Δ,

$$\left.\begin{array}{ll} \Delta g^*_{\mu\nu} = 0 & \text{(unless } \mu = \nu = 4) \\ \Delta g^*_{44} = \kappa c^2 \rho_0. & \end{array}\right\} \tag{7d}$$

From this, as is well known, it follows that

$$\left.\begin{array}{ll} g^*_{\mu\nu} = 0 & \text{(except in the case where } \mu = \nu = 4) \\ g^*_{44} = \frac{\kappa c^2}{4\pi}\int\frac{\rho_0 dv}{r}, & \end{array}\right\} \tag{10}$$

where the integration extends over three-dimensional space, and r is the distance from dv to the origin. It follows from (1b) and (1b'), taking the approximation postulated above taken into account, that

$$\ddot{x} = -\frac{1}{2}\frac{\partial g_{44}^{*}}{\partial x}. \tag{1c}$$

Equations (9) and (1c) contain Newton's gravitational theory, where our constant κ is connected to the usual gravitational constant K by the relation

$$K = \frac{\kappa c^2}{8\pi}, \tag{11}$$

from which it follows that

$$K = 6,7 \cdot 10^{-8} \qquad \kappa = 1,88 \cdot 10^{-27}. \;|$$

[1260] In the approximation considered here, for the "natural" four-dimensional volume element ds, we have

$$ds = \sqrt{-dx^2 - dy^2 - dz^2 + g_{44}dt^2},$$

whereby

$$g_{44} = c^2\left(1 - \frac{\kappa}{4\pi}\int\frac{\rho_0 dv}{r}\right).$$

One can recognize that the coordinate length is identical to the natural length $(dt = 0)$; hence measuring rods undergo no distortion in a "Newtonian" gravitational field. By contrast, the rate of a clock depends upon the gravitational potential. For ds/dt gives a measure of the clock's rate, if one sets $dx = dy = dz = 0$. One obtains:

$$\frac{ds}{dt} = \sqrt{g_{44}} = \text{const.}\left(1 - \frac{\kappa}{8\pi}\int\frac{\rho_0 dv}{r}\right).$$

Thus, the greater the mass placed in its vicinity, the slower the clock ticks.[13] It is interesting that the theory has this result in common with Nordström's theory.

For the propagation of light $(ds = 0)$ one obtains the velocity

$$\mathfrak{L} = \left|\sqrt{\frac{dx^2 + dy^2 + dz^2}{dt^2}}\right|_{ds = 0} = \sqrt{g_{44}} = c\left(1 - \frac{\kappa}{8\pi}\int\frac{\rho_0 dv}{r}\right).$$

Thus according to the foregoing theory, and in contradiction with Nordström's theory, light rays are bent by the gravitational field. This is the only consequence of the theory find so far that is accessible to experience.

13 According to postulate (4), this result holds for the rate of any process whatsoever.

Without continuing this consideration of the use of approximations in field calculations, we will give the exact equations of motion for a point in the field considered here. From the general equations of motion (1b') we obtain

$$\frac{d}{dt}\left\{-m\sum_{\nu} g_{\sigma\nu}\frac{dx_{\nu}}{ds}\right\} = -\frac{1}{2}m\sum_{\mu\nu}\frac{\partial g_{\mu\nu}}{\partial x_{\sigma}}\frac{dx_{\mu}}{ds}\frac{dx_{\nu}}{dt}. \qquad (1b'')$$

For the special case of the Newtonian field this yields

$$\frac{d}{dt}\left\{m\frac{\dot{x}}{\sqrt{g_{44}-q^2}}\right\} = -\frac{1}{2}m\frac{\frac{\partial g_{44}}{\partial x}}{\sqrt{g_{44}-q^2}}. \qquad (1c')$$

9. ON THE RELATIVITY OF INERTIA

From (1c') it follows that the momentum I and the energy E of a mass point moving slowly in the Newtonian gravitational field are given by the equations:

$$\left.\begin{aligned} I_x &= m\left(1+\frac{\kappa}{8\pi}\int\frac{\rho_0 dv}{r}\right)\frac{x}{c} \text{ etc.} \\ &\text{and} \\ E &= mc\left(1-\frac{\kappa}{8\pi}\int\frac{\rho_0 dv}{r}\right)+\frac{1}{2}\frac{m}{c}\left(1+\frac{\kappa}{8\pi}\int\frac{\rho_0 dv}{r}\right)q^2. \end{aligned}\right\} \qquad (12)$$

Thus, although the energy of a point mass at rest decreases with the accumulation of masses in its vicinity, as the first term of the expression for E shows, the same accumulation leads to an *increase* of the inertia of the point mass under consideration. This result is of great theoretical interest. For if the inertia of a body can *increase* due to the piling up of mass in its vicinity, then we have no choice but to regard the inertia of a point as being caused by the presence of the other masses.Thus, inertia appears to be *caused* by a kind of interaction between the point mass to be accelerated and all of the other point masses.

This result appears quite satisfactory if one reflects on the following. It makes no sense to speak of the motion, and hence also the acceleration, of a body A by itself. One can only speak of the motion or acceleration of a body A relative to other bodies B, C, etc. Whatever holds true kinematically regarding acceleration must also hold true for the inertial resistance with which bodies oppose acceleration; it is to be expected a priori, if not exactly necessarily,[14] that inertial resistance is nothing but a resistance of the designated body A to relative acceleration with respect to the totality of all other bodies B, C, etc. It is well known that E. Mach first defended this point of view, with perfect acuity and clarity, in his history of mechanics, so that here I can

simply refer to his arguments. Let me also refer to a witty brochure by the Viennese mathematician W. Hoffman that independently argues for the same position. I will
[1261] call the | conception sketched here the "hypothesis of the relativity of inertia."

To avoid misunderstandings, let me say once more that, like Mach, I do not think that the relativity of inertia is a logical necessity. But a theory which grants the relativity of inertia is more satisfactory than our current theory, because in the latter theory, inertial systems are introduced which, on the one hand, have a state of motion that does not depend on the states of observable objects, and thus is not caused by anything accessible to observation, but, on the other hand, are supposed to determine the behavior of material points.

The concept of the relativity of inertia requires, however, not only that the inertia of a mass A increases when masses at rest pile up in its surroundings, but also that this increase of inertial resistance will not take place if the masses BC ... are accelerated with the mass A. One can express this point as follows: the acceleration of the masses BC ... must induce an accelerative force on A that is in the same direction as the acceleration. With this one can see that this accelerating force must overcompensate for the increase of inertia produced by the mere presence of BC ..., for according to the relation between the inertia and energy of systems, the system ABC ... as a whole must have less inertia the smaller its gravitational energy.

In order to see that our theory fulfills this requirement, we must take into account the terms on the right-hand side of the system of equations (7c), which are proportional to the first power of the velocity of the field-producing masses. We then obtain the following instead of the system of equations (7d):

$$\left.\begin{aligned} \Box g^*_{\mu\nu} &= 0 \qquad (\text{if } \mu \neq 4 \text{ and } \nu \neq 4) \\ \Box g^*_{14} &= -\kappa\rho_0\dot{x} \\ \Box g^*_{24} &= -\kappa\rho_0\dot{y} \\ \Box g^*_{34} &= -\kappa\rho_0\dot{z} \\ \Box g^*_{44} &= -\kappa c^2\rho_0. \end{aligned}\right\} \tag{7e}$$

The equations of motion of the material point (1b'') differ from (1c'), in that now g_{14}, also differ from zero. They read in full:

$$\left(\frac{d}{dt}\left\{m\left(\frac{dx}{ds} - g_{14}\frac{dt}{ds}\right)\right\} = -\frac{1}{2}m\left(2\frac{\partial g_{14}}{\partial x}\frac{dx}{ds} + 2\frac{\partial g_{24}}{\partial x}\frac{dy}{ds} + 2\frac{\partial g_{34}}{\partial x}\frac{dz}{ds} + \frac{\partial g_{44}}{\partial x}\frac{dt}{ds}\right)\right) \text{ etc.}$$

14 One typically avoids the consequences of such arguments by introducing reference systems (inertial systems) with respect to which freely moving mass points are in rectilinear uniform motion. What is unsatisfactory is that it remains unexplained how the inertial systems can be privileged with respect to other systems.

For a slowly moving point one can write these equation as follows in the usual three-dimensional vector notation:

$$\ddot{\mathfrak{r}} = -\frac{1}{2}\operatorname{grad} g_{44} + \dot{\mathfrak{g}} - [\dot{\mathfrak{r}}, \mathfrak{o}]. \tag{1d}$$

Here

$$\begin{aligned} \mathfrak{r} &= \text{radius vector of the mass point} \\ \dot{\mathfrak{r}} &= \frac{d\mathfrak{r}}{dt} \text{ etc.} \\ \mathfrak{g} &= \text{a vector with the components } g_{14}, g_{24}, g_{34}, \\ \mathfrak{o} &= \operatorname{curl} g \end{aligned}$$

If we denote the velocity of the field-producing masses (comp. $\dot{x}, y, \dot{z}$) with $\mathfrak{v}$, then we can write (7e) in more concisely:

$$\left.\begin{aligned} \Box\mathfrak{g} &= -\kappa\rho_0\mathfrak{v} \\ \Box g_{44}^{*} &= \kappa c^2\rho_0. \end{aligned}\right\} \tag{7e'}$$

The equations (7e') and (1d) show how slowly moving masses influence each other according to the new gravitational theory. To a great extent, the equations correspond to those in electrodynamics, g_{44} corresponds to the scalar potential of electrical masses up to the sign and up to the circumstance that the factor $1/2$ appears in the first term of the right-hand side of (1d). $\mathfrak{g}$ corresponds to the vector potential of electric currents; the second term on the right-hand side of (1d), which corresponds to an electric field strength resulting from a temporal change of the vector potential, yields precisely those induction effects, directed like the acceleration, that we must expect based on our ideas regarding the inertia of energy. The vector $\mathfrak{o}$ corresponds to the magnetic field strength (curl of the vector potential), so the last term in (1d) corresponds to the Lorentz force.

It should further be remembered that a term of the form $(\dot{\mathfrak{r}}, \mathfrak{o})$ occurs in the theory of relative motion in mechanics, and is known as the Coriolis force. One can show from (7e') that a field with vector $\mathfrak{o}$, exists on the inside of a rotating spherical shell, which leads to the result that the plane of oscillation of a pendulum set up inside the spherical shell does not stay fixed in space, but rather, due to the sphere's rotation, must take part in a precessional motion in the same direction as this rotation. This result is also to be expected from the meaning of the concept of the relativity of inertia, and has long been anticipated. It is noteworthy that the theory also agrees with the above conception with regard to this point, but unfortunately | the expected [1262]
effect is so slight that we cannot hope to confirm it via terrestrial experiments or astronomy.

10. CONCLUDING REMARKS

In the foregoing discussion we have sketched the most natural paths that a gravitational theory can follow. One either stands by the usual relativity theory, i.e., one assumes that the equations expressing laws of nature remain covariant only under linear orthogonal substitutions. Then one can develop a scalar theory of gravitation (Nordström's theory), which is fairly simple and sufficiently satisfies the most important requirements to be imposed on a gravitational theory, although it does not include the relativity of inertia as a consequence. Or one augments the relativity theory in the way sketched above. One certainly attains equations of considerable complexity, but, in exchange, the sought after equations follow from the basic principles with the help of surprisingly few hypotheses, and the conception of the relativity of inertia is satisfied.

Whether the first or the second way corresponds essentially to nature must be decided by observations of stars appearing near the Sun during solar eclipses. Hopefully the solar eclipse of 1914 will already resolve this important decision.

EDITORIAL NOTES

[1] In the original, the following equation was mistakenly given the equation number (2), which appeared already on p. 546 [p. 1251 in the original].

[2] In the original, the first $\gamma_{\tau\rho}$ and $\gamma_{\alpha\beta}$ are misprinted as $y_{\tau\rho}$ and $\gamma_{\beta\nu}$ respectively.

FROM HERETICAL MECHANICS
TO A NEW THEORY OF RELATIVITY

JULIAN B. BARBOUR

EINSTEIN AND MACH'S PRINCIPLE

INTRODUCTION

Einstein's attempt to realize Machian ideas in the construction of general relativity was undoubtedly a very major stimulus to the creation of that theory. Indeed, the very name of the theory derives from Einstein's conviction that a theory which does justice to Mach's critique of Newton's notion of absolute space must be generally relativistic, or covariant with respect to the most extensive possible transformations of the spacetime coordinates.

The extent to which general relativity is actually Machian is, however, the subject of great controversy. During the last six months, I have been examining closely all of Einstein's papers that concern the special and general theory of relativity together with a substantial proportion of his correspondence related to relativity. There were several things that I wished to establish: 1) What precisely was the defect (or defects) in the Newtonian scheme that Einstein sought to rectify in his general theory of relativity? 2) How did Einstein propose to rectify the perceived defect(s)? 3) What relation does Einstein's work on his Machian ideas bear to the other ideas and work of his predecessors and contemporaries on the problem of absolute and relative motion? 4) Finally, to what extent did general relativity solve that great and ancient problem of the connection between and status of absolute and relative motion?

In this paper, which addresses the first three issues and gives my main conclusions (which are being presented in more detail together with my attempt at an answer to the fourth question in a forthcoming book (Barbour, in preparation)), I begin by reviewing the most important contributions to the discussion of absolute and relative motion made by Einstein's predecessors and contemporaries. As we shall see, this work identified certain key problems and went some way to providing the solutions to them. In particular, in 1902 Poincaré (1902; 1905, 75–78 and 118) provided a very valuable criterion for when a theory could be said to be Machian. Moreover, Mach (1883, 1960), Hofmann (1904), and Reissner (1914, 1915) made definite proposals of non-relativistic models of particle mechanics that meet this criterion. The examination of Einstein's entire relativity opus shows that this work made virtually no impact on him. Moreover, there is rather strong evidence which indicates a surprising lack of awareness on Einstein's part of the central problem with which the absolute-relative debate is concerned—*the problem of defining velocity*, i.e., change of position (and, more generally, *change* of any kind). For reasons that can be at least partly under-

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 3
Gravitation in the Twilight of Classical Physics: Between Mechanics, Field Theory, and Astronomy.

stood, Einstein saw this as a relatively trivial matter and regarded *acceleration* as more problematic.

In fact, Einstein associated with Mach's name two specific problems.

The first may be called the **absolute-space problem**, but it could equally well be called the problem of the *distinguished frames of reference*. Einstein initially presented it as the great mystery of why there seem to exist distinguished frames of reference for the expression of the laws of nature, though later he often spoke of the unacceptability of there being a thing (absolute space) that could influence the behavior of matter without itself being affected by matter.

The second may be called the **inertial-mass problem**. This problem was first mentioned explicitly by Einstein in 1912, when he asserted that Mach had sought to explain the *inertial mass* of bodies through a kind of interaction with all the masses of the universe.

In the years up to the definitive formulation of general relativity in 1915 and a little beyond, Einstein repeatedly mentioned these two problems. However, in 1918, following a critique by Kretschmann (Kretschmann 1917), Einstein (Einstein 1918a) said that he had not hitherto distinguished properly between these two problems (and between the means by which he proposed to resolve them). He then gave a formal definition of what he called *Mach's Principle*, which took the form of the requirement that all the local inertial properties of matter should be completely determined by the distribution of mass-energy throughout the universe. He said that this was "a generalization of Mach's requirement that inertia should be derived from an interaction of bodies." At the same time, Einstein gave a definition of the relativity principle that took from it all the specific empirical content it had previously seemed to possess in Einstein's work and transformed it into a very general necessary condition on the very possibility of stating any laws of nature: "The laws of nature are merely statements about spacetime coincidences; they therefore find their only natural expression in generally covariant equations."

Towards the end of his life, Einstein admitted (not very publicly but explicitly in a letter to Felix Pirani)[1] that his 1918 formulation of Mach's Principle made no sense mathematically and from the physical point of view had been made obsolete by the development of physical notions that had displaced material bodies from the pre-eminence they had possessed in Newtonian theory. However, to the end of his life he retained the 1918 formulation of the relativity principle, which he admitted carried little real physical content. However, he asserted that in conjunction with a requirement of simplicity it possessed great heuristic value, namely that, in a choice between rival theories, preference should be given to those theories that, when expressed in generally covariant form, took a simple and harmonious form.

This faith in *simplicity* as a criterion for selecting physical theories is extremely characteristic of Einstein and gives expression to his deep faith in the ultimate rationality of physics. It is, however, a notoriously slippery criterion. It is also a fact that

1 Einstein to Felix Pirani, 1954 (EA 17-447).

when, in the years up to and including 1916, Einstein said that a satisfactory theory of gravity and inertia must be generally covariant he undoubtedly thought that this requirement had a deep physical significance going far beyond the bland 1918 formulation of the relativity principle.

Mach made the comment that the creators of great theories are seldom the best people to present those theories in a logically concise and consistent form. In this book devoted to alternative strategies that could have been adopted (and in some cases were) to the development of relativity theory, I hope that the following attempt to establish what Einstein was trying to do, actually did, and might have done will help to cast light on the extremely tangled story of the creation of one of the wonders of theoretical physics: the general theory of relativity. In particular, I hope this paper will complement the articles by Jürgen Renn and John Norton[2] (both of which I found very useful in my own work) by looking at Einstein's work closely from the perspective of the specific problem of absolute *vs* relative motion. John Norton has done a splendid technical and conceptual job in comparing Einstein's approach with the more conventional 'Lorentz-invariant field theoretical' approach (to use Norton's useful anachronism) that virtually all his contemporaries adopted to the finding of a relativistic field theory of gravitation. Jürgen Renn, for his part, has emphasized the vital importance of Einstein's more wide-ranging approach and the inclusion of epistemological problems from the foundations of mechanics in the set of issues to be resolved in a satisfactory theory of gravitation. He brings out the value of Einstein's philosophical and integrative outlook. Examination of Einstein's work from the specific absolute *vs* relative perspective brings to light some further issues and aspects of Einstein's work that are not so readily revealed in their approaches.

I hope and believe that nearly all the articles in this book will have not only historical and philosophical interest but also serve a useful purpose for current research. It is widely agreed that the greatest current problem that has to be solved in theoretical physics is that of the relationship between quantum theory and the general theory of relativity. It is my conviction (Barbour 1994, 1995, in preparation) that general relativity is deeply Machian in a sense that unfortunately Einstein never managed to pinpoint accurately and that precisely this very Machian nature of general relativity is the main cause of the difficulties that stand in the way of its quantization. I therefore hope that the present article will have not only historical relevance but also help to clarify some central issues of current research.

In this article, it will not be possible to give a comprehensive account. I aim merely to identify some of the most important issues and ask the reader to consult my forthcoming monograph for a more detailed account. See also the *Notes Added in Proof* at the end of this article.

---

2 See *The Third Way to General Relativity* and *Einstein, Nordström, and the Early Demise of Scalar, Lorentz Covariant Theories of Gravitation* (both in this volume).

1. THE ORIGIN AND EARLY HISTORY OF THE ABSOLUTE *VS* RELATIVE DEBATE

The whole absolute *vs* relative debate arose from Descartes's claim in his *Principles of Philosophy* (1644) that *motion is relative* (Barbour 1989). Descartes argued that position can only be defined relative to definite reference bodies. Since there is evidently no criterion for choosing certain reference bodies in preference to others, Descartes argued that there can be no unique definition of motion—a given body has as many different motions as there are reference bodies (which, in general, will, of course, be moving relative to each other) with which it can be compared.

Despite this rather cogent argument, Descartes then proceeded, in a manifest *non sequitur*, to formulate definite laws of motion, the first two of which were identical in their content to the law that Newton subsequently adopted as his first law: Any body free of disturbing forces will either remain at rest or move in a straight line with uniform speed. It is evident that such a law presupposes a definite frame of reference—a reference space—and an independent time (an external clock) if it is to make any sense. About this mysterious reference space Descartes said not a word.

We know from Newton's tract *De Gravitatione* (Hall and Hall 1962), written around 1670 but published only in 1962, that Newton was intensely aware of the flagrant contradiction between Descartes's espousal of relativism and the vortex theory, on the one hand, and his anticipation and formulation of the law of inertia, on the other. In a world in which all matter is in ceaseless relative motion (as it is in accordance with Cartesian vortex theory or the atomistic theories so prevalent in the 17th century), Cartesian relativism seems to make it utterly impossible to define a definite motion; in particular, it would appear to be impossible to say that any given body is moving in a straight line. Commenting sarcastically on Descartes's law, Newton said: "That the absurdity of this position may be disclosed in full measure, I say that thence it follows that a moving body has no determinate velocity and no definite line in which it moves." This may truly be called the *fundamental problem of motion*: If all motion is relative and everything in the universe is in motion, how can one ever set up a determinate theory of motion?

The entire story of the absolute *vs* relative debate flows from this dilemma that Newton posed so clearly in around 1670. For completeness, one should also add the temporal part of the story: Motion can never be measured by *time* in the abstract but only by a definite comparison motion. For scientific purposes, the comparison motion was for millennia the rotation of the Earth, though more recently a global network of atomic clocks has been introduced as the official standard of time. Thus, statements in physics involving time are really statements about physical clocks, for which a theory based on first principles is needed (given the fundamental importance of time).

Having formed the deep conviction that no sensible mathematically well-defined dynamics could be based upon Cartesian relativism, Newton insisted on the introduction of a rigidly fixed absolute space and a uniformly flowing external absolute time as the kinematic framework for the definition of motion. However, he was still very conscious of the cogency of Descartes's relativism and in the famous Scholium in the

Principia on absolute and relative motion admitted freely the need to show how absolute motions, which cannot be observed directly ("because the parts of that immovable space, in which those motions are performed, do by no means come under the observation of the senses"), could be deduced from the observed relative motions. This task may be appropriately called the *Scholium problem*: Given observed relative motions, find the corresponding absolute motions. Although Newton actually claimed at the end of the Scholium that he wrote the *Principia* specifically in order to show how that problem is to be solved, he never spelled out the solution explicitly and in the Scholium merely advanced some first qualitative arguments designed to show that absolute space must exist. Even less effort was made to demonstrate the existence of absolute time.

Despite eloquent criticism of the notions of absolute space and time by Newton's contemporaries Huygens, Leibniz, and Berkeley, the absolute *vs* relative problem remained effectively in a state of limbo for very nearly 200 years until it was taken up again by the mathematician Carl Neumann in 1870 (Neumann 1870) and by Ernst Mach in 1872 (Mach 1872, 25; 1911) at the end of an extended essay on the conservation of energy and then again in his famous book on mechanics in 1883 (Mach 1883, 1960). Parallel but less influential work was done in Britain (Scotland to be precise) by William Thomson (later Lord Kelvin) and Tait (Thomson and Tait 1867, §§208ff.; Tait 1883) and also Lord Kelvin's brother James Thomson (Thomson 1883). The interventions of Neumann and Mach brought two issues to the fore.

The first was essentially the Scholium problem: under the assumption that Newton's scheme is in essence correct, how can one make correct epistemological sense of his notions of absolute space and time? Important and significant contributions to the resolution of this problem were made by Neumann (Neumann 1870), Tait (in an unfortunately little noted elegant piece of work (Tait 1883)), Ludwig Lange (Lange 1884, 1885, 1886), the logician Frege (Frege 1891), and above all Poincaré (Poincaré 1898 and 1902; 1905, 75–78 and 118). This work will be considered in Sec. 3.

The second issue brought to the fore was Mach's proposal, made already in 1872 and then repeated (though not quite so clearly or unambiguously as one might wish) in his 1883 *Mechanik* and all its subsequent editions, to the effect that Newton's mechanics might actually be *physically incorrect* and should be replaced by a dynamics of a different form in which only relative separations of bodies occur. The physical cogency of this proposal was made much more impressive by Mach's ability to counter Newton's bucket argument from the undoubted existence of centrifugal force to the need for an absolute space to explain it. Mach observed that the distant masses of the universe rather than some absolute space could be the ultimate origin of the centrifugal forces and that if this were the case local material bodies, such as the wall of Newton's bucket, could be expected to have only a minuscule and unobservable effect.

2. DIRECT ATTEMPTS TO IMPLEMENT MACH'S PROPOSAL AND THEIR LACK OF IMPACT ON EINSTEIN

Although they have attracted very little notice, attempts at a direct implementation of Mach's proposal were made throughout the twentieth century. The first such attempts were made early enough for them to have influenced Einstein in his work on general relativity. In this section, this work and its very marginal impact on Einstein will be considered.

A proposal for a new, non-Newtonian mechanics was already advanced by Mach, in a very tentative and mathematically rather unsatisfactory form, in the *Mechanik* in 1883.[3] His ideas were advanced in several interesting ways by the Friedlaender brothers in a rather obscure booklet published in 1896 (Friedlaender and Friedlaender 1896). In a simple and beautiful example,[4] Benedict Friedlaender showed how distant rotating masses (the 'stars' as seen from someone rotating with Newton's bucket) could very well generate centrifugal forces away from the axis of rotation and thus make absolute space unnecessary. In his contribution to this volume, Renn discusses the various interesting points and also anticipations of Einstein's later work that can be found in the Friedlaenders' booklet.

A rather general way of generating (nonrelativistic) relational theories of the kind envisaged by Mach was found by a certain Wenzel Hofmann of Vienna, who in 1904 (Hofmann 1904) published an even more obscure booklet[5] than the Friedlaenders' which would surely have been lost forever had it not been for fleeting references to it by Mach in the 5th and 6th editions of the *Mechanik* and by Einstein in 1913 (Einstein 1913a). In modern terms, the essence of Hofmann's proposal was to replace the Newtonian kinetic energy T, which occurs in the Lagrange function $T - V$ of the classical mechanics of n point particles and consists of a sum over individual masses of the form

$$1/2\sum m_i\dot{\boldsymbol{r}}_i \cdot \dot{\boldsymbol{r}}_i,\ i = 1, \ldots, n, \tag{1}$$

where m_i is the mass of particle i, $\boldsymbol{r}_i$ is its position vector in absolute space, and the dot denotes the time derivative, by a sum over all pairs of the n particles of the form

$$\sum_{i<j} m_i m_j f(r_{ij})\dot{r}_{ij}^2, \tag{2}$$

where r_{ij} is the (Euclidean) separation of particles i and j, $f(r_{ij})$ is some function of this separation, and the dot has the same meaning as in (1).

Hofmann was able to show qualitatively that in a realistic cosmological model, in which there are many stars distributed more or less uniformly over a large area,

---

3 See (Mach 1960, §VI.7, 286–7) and the discussion of this section by Norton (who questions whether it is a proposal for a new mechanics) and myself in (Barbour and Pfister 1995).

4 Translated in part in (Barbour and Pfister 1995).

5 Mach's proposal reduced essentially to the special case $f = 1$ of Hofmann's general proposal (2).

masses such as those in the solar system would behave in accordance with laws that approximated quite well Newton's laws but in an effective space determined explicitly by the matter distribution in the universe.

Hofmann's idea has since been independently rediscovered many times. The first person to do that was Reissner in 1914 and 1915 (Reissner 1914, 1915), when he chose the particular form $1/r_{ij}$ for $f(r_{ij})$ in (2). This choice is physically plausible and has some remarkably interesting consequences as was shown in part by Reissner himself and also Schrödinger (Schrödinger 1925) in a very beautiful paper at least partly inspired by Reissner's work.

More recently, Bertotti and I (Barbour and Bertotti 1977, 1982) considered a very general framework for constructing relational theories of this kind, including a relational treatment of time. The basic idea is taken straight from Mach. One assumes that dynamics must be formulated for the universe as a whole[6] and, in a variational formulation, insists that only the relative quantities r_{ij} and their rates of change may appear on the Lagrangian that describes the dynamics of the universe. Time is treated relationally by insisting that all changes are measured, not by comparison with some abstract external time t but always by comparison with other actual changes in the universe. This has the effect that Newton's abstract time is replaced by an appropriate average of the totality of changes in the universe.

It turns out that within this large class of possible Machian theories there exist at least two distinct subclasses. One is essentially the class discovered by Hofmann, but it has the disadvantage that it leads to an effective inertial mass that is anisotropic in the presence of nearby accumulations of mass. Schrödinger, in particular, was well aware of this anisotropy and knew that it could lead to an experimental refutation of such theories. He attempted to investigate the effect of the Galaxy and found it to be just below the then existing observational accuracy. He was however using a much too low value for the mass of the Galaxy, and modern data rule out such a theory completely. Such theories are therefore of interest mainly as examples of what Machian theories might look like. In contrast, in the theories of the second class, which Bertotti and I base on a notion called the intrinsic derivative (or *best matching*), mass anisotropy is completely absent. Indeed, one can construct intrinsic models of Machian mechanics that in their locally (but not globally) observable consequences are completely indistinguishable from Newtonian mechanics. I shall return to this briefly at the end of the next section.

The fact that the basic idea of relational mechanics was rediscovered many times[7] indicates that it is a very natural and direct way of realizing Mach's ideas and thereby eliminating absolute motions (and with them absolute space and time) from the foundations of physics. Given Einstein's passionate desire to implement Mach's ideas, it

---

6 This is implicit in the proposal of Mach and is made explicit by the appearance of the crucial summation in Hofmann's expression (2).

7 Apart from Hofmann, Reissner, and Schrödinger in the early part of this century, at least five other people besides Bertotti and myself hit on the same basic idea in the period 1960–1990, as noted in the articles by myself and Assis in (Barbour and Pfister 1995).

has always seemed to me most surprising that the basic idea—the insistence that only relative quantities should appear in the laws of nature—never seems to have been considered seriously by Einstein. All of Einstein's work on relativity—from 1905 right through to his death in 1955—has a quite different 'flavour.' In fact, it is quite difficult to find evidence that Einstein was even aware of the possibility.

Unless more evidence comes to light in the as yet unpublished correspondence, the only really clear statement of Einstein which does show that he was aware of what might be done along these lines comes from a paper published in 1918 (Einstein 1918b) with the title "Dialogue on objections to the theory of relativity," which includes the following:

> We want to distinguish more clearly between quantities that belong to a physical system as such (are independent of the choice of the coordinate system) and quantities that depend on the coordinate system. Ones initial reaction would be to require that physics should introduce in its laws only the quantities of the first kind. However, it has been found that this approach cannot be realized in practice, as the development of classical mechanics has already clearly shown. One could, for example, think—and this was actually attempted—of introducing in the laws of classical mechanics only the distances of material points from each other instead of coordinates; *a priori* one could expect that in this manner the aim of the theory of relativity should be most readily achieved. However, the scientific development has not confirmed this conjecture. It cannot dispense with coordinate systems and must therefore make use in the coordinates of quantities that cannot be regarded as the results of definable measurements

In the absence of definite references, it is impossible to know for sure whose work Einstein had in mind with his "this was actually attempted" but it is plausible to suppose that he was referring to Mach's original proposal of 1883, Hofmann's 1904 booklet, which he had mentioned briefly in 1913 (Einstein 1913a), describing it as "ingenious," and also perhaps Reissner's two papers.[8] It must also be said that, if he was thinking of the work of Hofmann and Reissner, Einstein had clearly failed to grasp what had been achieved in that work. Both authors had in fact succeeded in finding a genuine alternative to Newtonian inertia governed by absolute space. Moreover, the alleged difficulty to which Einstein refers, that of dispensing with coordinate systems, is simply nonexistent. Both Hofmann and Reissner *did* dispense with coordinate systems in the formulation of their proposed law and worked directly with "only the distances of material points from each other instead of coordinates."

Since these last cited words of Einstein do perfectly encapsulate what Mach had advocated, and since also Einstein repeatedly expressed the greatest admiration for Mach's critique of Newtonian mechanics, his remarks in 1918 present something of a

---

8 No correspondence from Einstein to Reissner survives. There is one letter from Reissner to Einstein in the Einstein Archives. It dates from 1915 but concerns Reissner's work on general relativity. Reissner makes no mention of his Machian papers. In September 1925, Einstein (Einstein to Schrödinger, September 26, 1925 (EA 22-003) thanked Schrödinger for sending him a copy of his 1925 paper on the relativity principle. Einstein merely said it was "interesting." Had the work of Hofmann and Reissner truly made any impact on him, one might have expected Einstein to point out to Schrödinger that his work had been anticipated by them.

puzzle, as I noted a little earlier: Why did Einstein take so little interest in a serious and direct attempt to implement Mach's proposal? To this query one may add the observation that Einstein's frequent references to Mach in his papers in the period 1912 to 1923 seldom reflect accurately what Mach actually said and sometimes even represent a serious distortion. The most serious distortion concerns a straight confusion between two quite distinct meanings of the word *inertia*. It is worth saying something about this.

Both in Mach's time and now, the word inertia meant two things: first, as expressed in Newton's first law, the law of inertia, namely the tendency of a body to continue in rest or in uniform motion in a straight line unless acted upon by some force; second, the quantitative measure of resistance to acceleration as expressed by the presence of m, the *inertial mass*, in Newton's second law $\boldsymbol{F} = m\boldsymbol{a}$. Mach (Mach 1872, 25; 1883) pointed out that Newton had failed to give a meaningful definition of inertial mass and proceeded to supply one himself. He believed that his definition removed all difficulty surrounding the use of the concept of inertial mass in Newtonian dynamics. In contrast, he felt that Newton's formulation of the law of inertia was very seriously deficient and probably incapable of being given adequate expression without some actual change in its physical content. Mach insisted that genuine content must be given to expressions like "uniform motion in a straight line": uniform with respect to what and straight with respect to what? He considered it absolutely impermissible to invoke invisible time and space to answer these questions, and his discussion of these issues takes us straight back to the problems with which Newton grappled in *De Gravitatione*.

Very careful examination of *all* of Einstein's numerous comments on issues related to Mach have led me to a very surprising conclusion. Einstein *never once* even mentioned this problem—the fundamental problem of motion—at the heart of dynamics. He seems to have been more or less completely blind to its existence. He very often used the word inertia but never once made the distinction between the two meanings of it. When he was most explicit about Mach and inertia, he incorrectly attributed to Mach the idea that the *inertial mass* should arise in some manner from a kind of interaction of all the bodies in the universe (Einstein 1912, 1917). Now it is true that the m_i's that appear in Hofmann's proposal (2) are best interpreted as inertial *charges*. In the theory to which (2) and other similar proposals give rise, one then obtains effective *inertial masses*, which are indeed determined by interaction with all the bodies in the universe. This was clearly demonstrated by both Reissner and Schrödinger, but it was already qualitatively clear to Hofmann.

Einstein may very well have had a correct intuitive appreciation that some such effect could come out of a Machian theory of motion, but his repeated assertions that this was what Mach had called for are unfortunate on several counts: 1) They are historically inaccurate. 2) The effect arises in a certain class of Machian theories—the class considered by Hofmann, Reissner and Schrödinger—but not in another, which Bertotti and I discovered (Barbour and Bertotti 1982). This second class of theories is impeccably Machian and actually includes general relativity as a special and remarkably interesting example (Barbour 1995, see also the *Notes Added in Proof*). 3) Einstein's concentration on the inertial mass deflects attention away from the true and profound problem that underlies the absolute *vs* relative debate: How are time and motion to be defined?

This is the fundamental question that, very surprisingly, Einstein never addressed directly. In the final section of this paper, I shall try to establish why this was so. However, before then, in the following section, I want to complete the review of the work of Einstein's predecessors and contemporaries. As noted earlier, the critique of Neumann and Mach raised two issues: 1) Can Newtonian theory be recast in an epistemologically satisfactory manner without change of its essential physical content? 2) Can Newtonian theory be replaced by a physically different theory based on Machian ideas?

This section has essentially considered the answer to the second question. In the next section, we shall consider the answer to the first.

3. THE EPISTEMOLOGICAL WORK OF NEUMANN, LANGE, AND POINCARÉ AND ITS IMPACT ON EINSTEIN

In his habilitation lecture of 1870, Neumann posed a general problem and provided a partial solution to a small part of it. The general problem was this: As formulated by Newton, the laws of mechanics simply cannot be tested because absolute space and time are invisible and inaccessible to experimentalists. The question then was: Is it nevertheless possible to make epistemological sense of Newton's laws by identifying operational surrogates of absolute space and time?

To begin to make progress in this direction, Neumann assumed that particles moving freely of all forces (force-free particles) exist and could be identified as such and that also by some means absolute space (or a suitable surrogate of it) could be observed directly. If the second assumption is satisfied, one can then observe the motion of some chosen force-free particle. Neumann pointed out that, in the absence of an external clock, it is meaningless to say that such a particle is moving uniformly (though, if absolute space has been 'made visible', one can verify that it is moving in a straight line). However, what one can do is observe further force-free particles and see how they behave relative to the original particle, which is taken as a reference body. One can use the distance traversed by this reference body as a measure of time (inertial clock) and see if, relative to this inertial clock, a second force-free body moves uniformly. In this way, Neumann was able to give genuine operational content to the part of Newton's first law which asserts the uniformity of the motion of a force-free body. However, Neumann admitted that he was unable to solve the problem of making absolute space 'visible.'

This problem was taken up by the youthful Ludwig Lange (he was only 21) in 1884. He proceeded very much in the spirit of Neumann and assumed the existence of force-free particles that could be identified as such. His basic idea was to use *three* such particles to define a spatial frame of reference. Just as in the case of Neumann's inertial clock, for which it is meaningless to say that the clock itself is moving uniformly, Lange noted that it would be meaningless to say that his three reference bodies are moving rectilinearly. Instead, they *define* a frame of reference, with respect to which one can then verify that other bodies are moving rectilinearly. Moreover, using any one of the three chosen reference bodies as a Neumann inertial clock, one can simultaneously verify that further bodies are moving uniformly as well as rectilinearly.

Lange's actual construction of the spatial frame of reference using three force-free bodies is in fact rather awkward and clumsy, so I shall not attempt to describe it here, especially since I shall shortly describe a much neater construction due to Tait (Tait 1883). However, it is worth emphasizing the crucial point of the construction, which Lange was the first to recognize clearly and for which he deserves great credit. It will be recalled that Newton criticized Cartesian relativism because it made the motion of a considered body dependent on the choice of the reference bodies used to determine its motion. Since the choice of reference bodies is entirely arbitrary, it would appear that motion itself cannot be defined in any unique way. However, the situation is radically altered if one insists that the reference bodies—no matter which are chosen—*are themselves moving in accordance with Newton's laws*. This is the crucial stipulation that takes the seemingly fatal arbitrariness out of a relational definition of motion. Once this basic fact has been recognized, precise definitions merely reduce to a working out of details.

One severe problem with the Neumann-Lange approach—Lange never succeeded in overcoming it—was that of recognizing when bodies are free of forces. The construction depends crucially on the existence of unambiguously identifiable force-free bodies. This raises *two* problems: 1) How can one tell if a body is free of forces? 2) What can one do if nature fails to provide *any* force-free bodies? In fact, this is exactly the case with gravity, to which all bodies are subject. These serious difficulties were pointed out clearly by the logician Frege (Frege 1891) in an otherwise positive review of Lange's work. Frege correctly emphasized that the axioms of dynamics form a closed system and can only be tested in their totality. Since forces are an integral part of dynamics, their existence must be taken into account in the foundations of any method used to determine the distinguished frames of reference that play such an important role in Newtonian dynamics.

As it happens, the requirement that Frege raised was (in its essentials) met in three studies that unfortunately received very little attention. The first was actually the work of Tait in 1883 that I already mentioned. The other two were published in 1898 and 1902 by Poincaré.

Tait did not solve the problem of finding the dynamical frame of reference in full generality in the case when no force-free bodies are available. He did, however, give a solution to the problem for purely inertial motion that yields the Newtonian frames of reference given purely relative data and simultaneously confirms that all the considered bodies are actually free of all forces.

Tait solved the following problem, which had been posed by James Thomson (Thomson 1883). Suppose that at certain unknown instants of time we are given all the relative separations r_{ij} between a set of n point particles. Thus, we are, as it were, given 'snapshots' of the relative configurations of the particles. Using these snapshots and nothing else, can we verify if there exist a frame of reference and a measure of time, both of which must be deduced from the snapshots, in which all the particles are moving in accordance with Newton's first law?

To solve this problem, Tait supposed that the answer is yes. I shall consider the solution he gave for the case of three particles, since it fully illustrates the underlying principle. If all the particles are moving in accordance with Newton's first law, then one can certainly always choose the frame of reference in such a way that one of the particles is permanently at rest at the origin of the frame. If we exclude the special cases in which there are collisions of the particles, then if we consider some second particle there must exist a time at which it passes the first one at a distance a of closest approach. We can then choose x and y axes of the frame of reference in such a way that at $t = 0$ this second particle is at the point $(a, 0, 0)$ and at time t is at the point $(a, t, 0)$. Thus, we choose the unit of time such that particle 2 has unit velocity. It becomes a Neumann inertial clock. The spatiotemporal framework is then uniquely defined (up to reflections). At $t = 0$, the third particle will have some initial position (x_3, y_3, z_3) and initial velocity $(\dot{x}_3, \dot{y}_3, \dot{z}_3)$ Thus, this three-body problem will have *seven* essential unknowns. The problem of inertial motion is more or less trivial and one can find an analytical solution for the observable separations r_{ij} in terms of these seven unknowns. Given observed values of r_{ij}, these can be compared with the analytical solution and the seven unknowns determined.

As Tait noted, the most interesting point concerns the number of snapshots needed to find the seven unknowns. Each snapshot yields three independent data—the three sides of the triangle—but each snapshot is taken at an unknown time, so that only *two* useful data are supplied with each. It is thus clear that to determine the spatiotemporal framework and test whether all three particles are moving inertially in accordance with Newton's first law one needs at least four snapshots, since they give eight data, from which the seven unknowns can be determined and one verification made of the conjecture. Each extra snapshot yields a further two verifications.

Several important points emerge from Tait's analysis. First, contrary to a very widespread opinion engendered by Lange's work, three particles are already sufficient to establish the spatiotemporal framework and to test whether Newton's first law is satisfied. Lange, and many of his followers, believed three particles were needed to define the framework and that only a fourth would permit a nontrivial verification of Newton's law. Second, attention should be drawn to the central importance of the complete configurations of the three particles, which, in a sense, *define* the instants of time, and to the fact that both time and the spatial reference frame are best and mostly effectively determined together from the raw observational data—the relative separations. Third, knowledge of the spatial frame of reference is a vital prerequisite for determination of all quantities of primary concern in dynamics, above all time, which in the Tait procedure is read off from distance traversed in the spatial inertial frame of reference, and velocity and momentum, both of which can only be found once the complete spatiotemporal framework has been determined.

Two further points should be made here. In analytical mechanics, great emphasis is placed on the possibility of representing dynamics in completely arbitrary frames of reference. However, this does not alter the fact that somehow or other the primary dynamical quantities such as momentum and energy must be found in an inertial spatiotemporal framework. It is only then that a transformation to an arbitrary framework can be performed. Many people, even experts, are quite unaware of this fact. The second remark concerns the definition of a clock. It is widely believed that the essential basis of a clock is a strictly periodic process, the 'ticks' of which measure time. This belief is wrong on *two* scores. First, the Neumann-Lange-Tait procedure shows that linear distance traversed in an inertial frame of reference by a force-free particle is a perfectly good measure of time. Thus, a periodic process is not needed. Second, the inertial frame of reference and distance traversed in it are (in mechanics at least) always the ultimate source of a scientifically meaningful definition of time. Ironically, a pendulum clock, the rate of which depends upon the strength of the gravitational field in which it is set up, is not really a good clock, since its rate is not exclusively determined by its local inertial frame of reference. Thus, a pendulum clock goes faster near sea level than on the top of a mountain, but (as Einstein's general theory of relativity established) clocks that measure proper time go slower at sea level. This highlights the salient point: A clock, to function properly, must 'lock onto' or 'tap' processes directly and exclusively governed by the local inertial frame of reference.

We still have to consider the realistic general case in which no force-free particles are available at all. How is the inertial spatiotemporal framework to be determined in that case? As preparation to the answer to this question, it is worth noting that in the case of Tait's problem in the general case of n point particles, the number of unknowns to be determined is $1 + 6(n-2) = 6n - 11$ (giving our 7 for $n = 3$). On the other hand, each snapshot of n particles yields $3n - 6$ independent mutual separations or $3n - 7$ useful bits of information (since the time of the snapshot is unknown). Thus, two snapshots can only yield $6n - 14$ data, while $6n - 11$ are needed to determine the inertial spatiotemporal framework and, from it, the dynamically relevant quantities. Two snapshots are therefore never enough information but, if n is large, three are comfortably more than enough. The reason why two snapshots always fail to yield enough information is that, in Newtonian terms, they contain no data at all on the change of the *orientation* of the system of n particles as a whole in absolute space.

This fundamental fact was made the point of departure of a very interesting analysis of the problem of absolute *vs* relative motion made by Poincaré in his *La Science et l'Hypothèse* in 1902 (Poincaré 1902; 1905, 75–78 and 118). Before considering this, it is worth mentioning that unfortunately Poincaré never, so far as I know, published a single comprehensive study of the problem of determining the complete spatiotemporal framework of dynamics from observable relative quantities. He considered the temporal and spatial problems separately (the former in his "*Mèsure* *du temps*” in (Poincaré 1898) and the latter in 1902). Both studies were rather qualitative in nature, and both attracted much less attention than they might otherwise have done on account of the creation in 1905 of the special theory of relativity. This then attracted most of the serious attention of scientists concerned with foundational problems and also introduced a host of new issues. This was unfortunate, since a solid authoritative study by Poincaré, of which he was undoubtedly capable, would have become an important landmark in the absolute *vs* relative debate. As it is, his work has very largely passed unnoticed (in part, at least, because Einstein did not notice it, as we shall see).

In his *La Science et l'Hypothèse*, Poincaré asked what if anything was ‘wrong’ with Newton's use of absolute rather than relative quantities in the foundations of dynamics. Instead of asking the epistemological question—how do we find the absolute quantities given the relative quantities?—Poincaré posed a very interesting question, which was this: If, in the case of the n- body problem of celestial dynamics, one has access to only *relational* initial data (which will be the mutual separations r_{ij} of bodies and their various derivatives $\dot{r}_{ij}, \ddot{r}_{ij}, \ldots,$ with respect to the time t (Poincaré assumed t known for the purposes of his discussion)), what *initial data* must be specified if one is to be able to predict the observable future evolution of the system uniquely? Since the ability to predict the future is the acid test of dynamical theory, Poincaré's question could not be better designed to cast much needed light on the role of absolute and relative quantities in dynamics.

Poincaré then noted that if, like the relationists, one believed the relative quantities were truly fundamental and all that counted, one might then suppose that (given known masses of the bodies and under the assumption that they were moving in accordance with Newton's laws, including the law of universal gravitation) knowledge of the r_{ij} at one instant together with the rates of change of these r_{ij}, i.e., the $\dot{r}_{ij}$'s, would be sufficient to determine the future uniquely. However, he then drew attention to the fact with which we are already familiar from Tait's analysis of the inertial case, namely, that even in that simplest of cases two snapshots are not sufficient to determine the absolute quantities, which, as Poincaré pointed out, are needed to make dynamical calculations. (The initial-value problem of celestial mechanics is well posed if, in addition to the masses and specification of the law of interaction, one is given initial positions and initial velocities in *absolute space*.) The situation is no different if interactions occur. In Poincaré's view, this failure of the initial-value problem if one is given only relative quantities is the clearest indication that dynamics involves something more than just relations of bodies among themselves—and that ‘something more’ is what Newton called absolute space.

It is important to realize, as Poincaré was careful to emphasize, that it is perfectly possible to express the entire content of Newtonian mechanics in purely relational terms. However, the resulting equations, unlike Newton's equations, which contain at the highest *second* derivatives with respect to the time, must contain at least some *third* derivatives. Although he did not explicitly mention him by name, Poincaré almost certainly had in mind here Lagrange's famous study of the three-body problem of celestial mechanics made in 1772 (Lagrange 1772). Lagrange (1772) had assumed the validity of Newton's equations in absolute space and, in an outstanding piece of work, had then proceeded to find equations that govern the variation in time of the *sides of the triangle* formed by the three particles, i.e., precisely the r_{ij}'s for this problem. Lagrange had found three equations, each containing the r_{ij}'s and their derivatives symmetrically and all containing first, second, and *third* derivatives of the r_{ij} with respect to the time. He was also able to show that two of the equations could be integrated once, giving two equations of the form

$$F_1(r_{ij}, \dot{r}_{ij}, \ddot{r}_{ij}) = E, \tag{3}$$

$$F_2(r_{ij}, \dot{r}_{ij}, \ddot{r}_{ij}) = M^2, \tag{4}$$

where E is the total energy of the system and M^2 is the square of the total angular momentum of the system (both in the center-of-mass system). These equations show very graphically that whereas the fundamental dynamical quantities such as energy and angular momentum are functions of the coordinates and their *first* time derivatives *in absolute space*, the expressions for the same quantities in relative quantities also necessarily contain the *second* derivatives.

Poincaré considered this a decidedly mysterious and unsatisfying feature of Newtonian mechanics and felt that it was the only thing one could fault in the Newtonian scheme. He felt, repugnant though this state of affairs was to a philosophically minded person, that one still had to accept it as a fact. He was however prepared to speculate as to how things might be in an ideal world, and this led him to a very interesting speculation as to the form that the relativity principle might have taken.

He noted that the ordinary Galilean relativity principle of classical mechanics had very interesting consequences for the initial data that had to be specified in mechanics. An n-particle system requires formally the specification of $3n$ initial positions and $3n$ initial velocities in absolute space. However, because of the fundamental symmetries of classical mechanics, it is sufficient to specify these quantities with respect to the center of mass of the system. This reduces the number of data that need to be given by 6. In addition, the initial orientation of the system in absolute space has no physical significance, so three more data are redundant. However, essentially that is as far as the reduction to relative quantities can go. It remains crucially important to know at the initial instant how the orientation of the system as a whole in absolute space is changing. This cannot be obtained from purely relative quantities and is the reason why third derivatives of the r_{ij} occur in one of Lagrange's equations.

Such considerations then led Poincaré to comment that "for the mind to be fully satisfied" the law of relativity would have to be formulated in such a way that the initial-value problem of dynamics would hold for a *completely relational* specification of the initial data. One should not be left with the curious absolute-relative mixture just described.

This analysis and suggestion of Poincaré are both extremely valuable. They show that the problem with Newtonian dynamics is not that it cannot be cast into relational form—Lagrange's work is the clearest demonstration of the incorrectness of that belief (which is actually quite widely held).[9] The problem is that when Newtonian theory is recast in a relational (or generally covariant) form it turns out to be *less predictive* than one would like it to be. In addition, Poincaré's analysis also shows what a Machian theory, expressed solely in relative quantities as Mach required, must achieve if it is to represent any improvement on Newtonian theory: It must be able to predict the future uniquely given only r_{ij} and $\dot{r}_{ij}$ at an initial instant. Mach's critique of Newtonian mechanics was unfortunately couched in rather vague terms and the same goes for his proposal for a relational alternative. Poincaré's analysis provides a most welcome clarification and sharpening of the issues involved.

It should be mentioned that all the Machian models of the Hofmann-Reissner-Schrödinger type together with the alternative (intrinsic) type considered by Bertotti and myself meet the requirement of the relativity principle in the stronger form as formulated by Poincaré. It is also the case that the special set of *Newtonian* solutions of an n- body universe for which the total angular momentum in the center-of-mass system vanishes are described by equations of a form different from those that hold in the general case. In this special case, the constants E and M^2 disappear from the right-hand sides of Eqs. (3) and (4) and the third derivative also disappears from Lagrange's third equation. Therefore, the corresponding set of equations for this special case satisfy Poincaré's requirement. Indeed, it is a very interesting fact that when Newton's equations are expressed in a generally covariant form (as Lagrange in effect did, using quantities completely independent of all coordinate systems), the complete set of possible solutions breaks up into distinct classes corresponding to the general case with both $E, M^2 \neq 0$ and the various special cases with either one or both of E and M^2 equal to zero. The most interesting special case

$$E = 0, M^2 = 0 \tag{5}$$

arises very naturally from the intrinsic Machian dynamics developed by Bertotti and myself and referred to in the previous section.

In fact, such a situation was foreseen to quite an extent by Poincaré, who pointed out that, when one is considering the complete universe, it is appropriate to consider

9 Lagrange's work does in fact represent the complete solution (for the three-body case) of the problem that Newton posed in the Scholium: Given relative observations, how can one find the absolute quantities? First, Lagrange found equations that govern the evolution of the sides of the triangle. Second, he showed how, once these equations for the sides of the triangle had been solved, one could find the position of the triangle in absolute space (the position of its center of mass and—a much greater problem—its orientation) by quadrature (i.e., by straightforward integration of functions known from the solution of the problem for the sides). A good account of all this is given by Dziobek (Dziobek 1888, 1892). It is somewhat ironic that Lagrange was evidently much more interested in practical problems of celestial mechanics than Newton's Scholium problem and did his work at a time when absolute space had ceased to be a problematic issue. Its importance for the Scholium problem was not noted and escaped Neumann, Lange, and Mach. It is truly a great pity that Poincaré did not flesh out his very perceptive remarks in *La Science et l'Hypothèse* and draw explicit attention to Lagrange's work and its bearing on the Scholium problem.

these various different cases as actually corresponding to fundamentally different dynamical laws of the universe. An important point to note is that if an n- body universe as a whole does satisfy the condition (5) isolated subsystems of it can still perfectly well have nonvanishing values of their energy and angular momentum. They would then appear to be governed by perfectly standard Newtonian dynamics, even though the universe as a whole is governed by a more powerful and more predictive dynamics. This is the reason why Bertotti and I were able to recover Newtonian behavior exactly for local observations. It may also be mentioned that the formalism of intrinsic dynamics is completely general and is not restricted to nonrelativistic mechanics. Unlike the Hofmann-Reissner-Schrödinger approach, it can readily be applied to field theory and even to dynamic geometry. Indeed, it turns out that general relativity is itself of the general type of intrinsic theories, and this is the reason why Bertotti and I have concluded that it is actually perfectly Machian (Barbour 1995).

Let me now go back to Poincaré's earlier paper of 1898 on the topic of time. This paper has received significantly more attention than the analysis of the absolute *vs* relative question in *La Science et l'Hypothèse*, but its Machian implications have nevertheless been completely missed.

Poincaré noted that in recent years there had been considerable discussion of the problem of measuring time. What does it mean to say that a second today has the same *duration* as a second tomorrow? What criterion is to be used to choose the unit of time and identify clocks? Poincaré noted that these questions had become especially topical and acute for the astronomers, who had been finding anomalies in the observed motion of the Moon, one possible explanation of which could be irregularities in the rotation rate of the Earth. (This has since been confirmed. It is due to tidal effects of the Moon.) Since for millennia the rotation of the Earth had constituted the sole reliable clock for use in astronomy, this placed the astronomers in a serious quandary.

Poincaré then proceeded to outline the solution to which the astronomers were moving. Their point of departure was that Newtonian theory was in fact correct, namely, that there did exist a frame of reference and time for which Newton's laws were correct. The entire problem consisted of finding the invisible frame of reference and time from things that could actually be observed. The only material on which they could work was the motions of the bodies making up the solar system. Fortunately, this could, on account of the immense distance of the stars, be treated as an effectively isolated dynamical system. However, in contrast to the gedanken experiments considered by Lange and Tait, the bodies of the solar system were certainly not free of forces, since they all interacted with one another through universal gravitation. The astronomers were therefore confronted with the task that Frege a few years earlier had said needed to be solved by Lange.

The solution proposed by the astronomers, and endorsed in principle by Poincaré, was to seek a frame of reference and time in such a way that the observed motions did indeed accord with Newton's laws when referred to the obtained frame of reference and time. This is a rather obvious generalization of the method initiated by Neumann, Lange, and Tait, but, of course, entailed much greater mathematical difficulties on account of the need to take into account interactions. Fortunately for the astronomers, they did not have to start completely from scratch, since excellent approximations to the conjectured Newtonian frame of reference and time already existed.

A very significant difference of this astronomical procedure from the Tait-Lange procedure is that in the latter time and the frame of reference can in principle be found from just *three* bodies, but the astronomical procedure entails consideration of *all* the dynamically significant bodies in the solar system. If accuracy adequate for astronomical purposes is to be achieved, it is in principle necessary to take into account even relatively small asteroids. This means that effectively the only clock available to the astronomers is the complete solar system.

About forty years after Poincaré wrote his 1898 paper, the astronomers did indeed go over to such a definition of time (which by then had to take into account small relativistic corrections as well). It was initially called *Newtonian* time, but is now known as *ephemeris time* (Clemence 1957). A rather beautiful feature of ephemeris time is that it is actually a weighted average of all the dynamically significant motions of the bodies in the solar system in its center-of-mass inertial frame. Were the solar system to consist of a system of point particles, the expression for the infinitesimal increment of ephemeris time would be given as follows. Let the position of particle $i, i = 1, \ldots, N$, at one instant of time be given by x_i and at a slightly later instant by $x + dx_i$, the positions being measured in the inertial frame of reference. Then the increment dt of ephemeris time is given by

$$dt = \frac{k\sqrt{\sum m_i dx_i \cdot dx_i}}{\sqrt{E-V}},$$

where E is the total energy of the system, V is the instantaneous potential energy, and k is a constant.

Note also that but for the fortunate fact that the solar system is almost perfectly isolated an accurate determination of time would require the summation in the above expression to be extended to the complete universe. Ultimately, the only reliable clock is the complete universe!

I have gone into this detail about ephemeris time (the theory of which was outlined rather more sketchily by Poincaré in his 1898 paper) because, first, it rectifies the shortcoming of Newton's treatment in the Scholium, and, second, it has passed almost without notice for over a century. This remarkable state of affairs has arisen because a quite different aspect of time—the problem of defining simultaneity at spatially separated points—came to dominate discussions once Einstein had created the special theory of relativity.

As it happens, Poincaré also mentioned this problem of simultaneity in his 1898 paper and noted that in some respects it was a more immediate problem than that of defining duration but that hitherto it had hardly been noted. It is on account of this remarkable early anticipation of the key problem of special relativity that Poincaré's

1898 paper is mentioned relatively often today, but I am not aware of any discussion of the duration problem even though it is certainly very fundamental.

The reason for this lack of notice is, I suspect, to be traced to the immense influence of Einstein, and this is an appropriate point at which to consider how his own work on special and, more particularly, general relativity relates to the topics discussed in this and the previous section. At the end of the previous section, I noted that Einstein seems to have had not much accurate knowledge of the work done by Hofmann and Reissner and to have taken little interest in it. The same comment is true of the epistemological work reported in this section. So far as I can judge from his published papers and the correspondence I have examined, *none* of the work described in this section made any significant impact on him. In the remainder of this section, I shall substantiate this claim; in the following section, I shall try to establish why Einstein seems to have been remarkable insensitive to what might be called the classical issues in the absolute *vs* relative debate.

Let me start with the topic last discussed—the definition of duration and a clock. To the best of my knowledge, this question was never once discussed by Einstein (in striking contrast to his numerous discussions of the definition of simultaneity). Throughout his entire work on relativity, Einstein simply assumed, as a phenomenological fact, that clocks (like rods to measure distance) exist and can be used to measure the fundamental interval ds of relativity theory.

Already in the 1920s (Einstein 1923) and then again in the *Autobiographical Notes* (Einstein 1949) written towards the end of his life, Einstein noted that his consistently phenomenological treatment of rods and clocks, which made it necessary to introduce them formally as separate entities in the framework of his theory, was a logical defect of the theory that ought to be eliminated. Rods and clocks should be constructed explicitly from the truly fundamental physical quantities in the theory—preferably fields alone, but, if particles could not be eliminated as fundamental entities, then from fields and particles together.

From the way Einstein wrote about this, I get the strong impression that he did not think anything particularly interesting would come out of this exercise. However, I think it can be argued that he was actually insensitive to a fundamental issue. This is reflected in the fact that he invariably described a clock as being realized through some strictly periodic process. However, this immediately begs the question that Neumann set out to answer with his inertial clock: How can one say of a single motion that it is uniform? I have not seen anything in Einstein's writings which shows an awareness of the fact that a measure of time can be extracted only from the totality of the motions within a dynamically isolated system and that, if it is to give true readings, a clock must somehow 'lock onto' and reflect the inertial spatiotemporal framework. I shall return to this.

A similar rather perfunctory attitude characterizes Einstein's references to the determination of inertial frames of reference. In his published papers, he never once referred to the procedures of Lange or Tait or drew attention to the difficulties that Newton 250 years earlier had already recognized so clearly. Generally, he simply says that an inertial frame of reference is one in which a force-free particle moves rectilinearly and uniformly, giving no indication at all how such a frame of reference is to be found. Very characteristic of his approach is the following passage written in the early 1920s (Einstein 1923):

> In classical mechanics, an inertial system and time are best determined together by means of a suitable formulation of the law of inertia: It is possible to establish a time and give the coordinate system a state of motion (inertial system) such that relative to it material points not subject to the action of forces do not undergo acceleration.

A little later, Einstein noted that such a definition had a logical weakness "since we have no criterion to establish whether a material point is free of forces or not; therefore the concept of an 'inertial system' remains to a certain degree problematic" This passage (with its incorrect conclusion) suggests to me that Einstein never gave much serious thought to the issue of the determination of inertial frames of reference.

Confirmation that this is the case can be found in some remarkably interesting late correspondence between Einstein and his old friend Max von Laue. Among the leading relativists, von Laue is the only one who mentions the work of Lange. In 1948 (von Laue 1948), he wrote an appreciation of Lange and his work, in which he stated: "Ludwig Lange progressed so far in the solution of the problem of the physical frame of reference, which Copernicus, Kepler, and Newton did not completely solve, that only Einstein's theory of relativity added something new." In 1951, he published a new edition of his book on the theory of relativity (von Laue 1955), which opens with the definition of the inertial time scale and inertial system as given by Lange, calling it a great achievement. Not surprisingly, he sent Einstein a copy of the new edition. In response,[10] Einstein commented:

> I was surprised that you find Lange's treatment of the inertial system significant. It merely says that there exists a coordinate system (with time) in which 'uninfluenced' material points move rectilinearly and uniformly. This is Newton's '*absolute* space.' It is not absolute because no transformations exist that conserve the law of inertia but because it must be prescribed in order to give the concept of acceleration a clear meaning.

In the same letter, Einstein remarked: "Provided one considers action-at-a-distance forces that decrease with r sufficiently rapidly, the word 'uninfluenced' has a direct meaning." This comment implies, like the one made in the 1920s, that inertial frames of reference can only be determined if force-free bodies are available. As we have noted earlier, this is simply not true, though unfortunately the correct state of affairs had never been clearly stated in the literature (see footnote 9). However, I am convinced that had Einstein really made a serious effect to find out the truth he would certainly have succeeded. What we must try to establish (in the next section) is why he was insensitive to the issue.

To conclude this section, it is worth mentioning a connection between Einstein's lack of concern about the definition of the inertial frame of reference and his belief

---

10 Einstein to Max von Laue, 17 January 1952 (EA 16–168).

that a Machian theory of motion should provide some kind of cosmic derivation of inertial mass (rather than a cosmic derivation of the law of inertia). It is a very striking fact that the expressions 'relativity of position' and especially 'relativity of velocity' (the truly fundamental problem of the absolute *vs* relative question) hardly ever occur in Einstein's writings, whereas he frequently mentions the relativity of acceleration. In fact, almost the only case in which relativity of velocity occurs is in the following passage (Einstein 1913b), in which Einstein is discussing his first attempt at a general theory of relativity undertaken with Grossmann in 1913:

> The theory sketched here overcomes an epistemological defect that attaches not only to the original theory of relativity, but also to Galilean mechanics, and that was especially stressed by E. Mach. It is obvious that one cannot ascribe an absolute meaning to the concept of acceleration of a material point, no more so than one can ascribe it to the concept of velocity. Acceleration can only be defined as relative acceleration of a point with respect to other bodies. This circumstance makes it seem senseless to simply ascribe to a body a resistance to an acceleration (inertial resistance of the body in the sense of classical mechanics); instead, it will have to be demanded that the occurrence of an inertial resistance be linked to the relative acceleration of the body under consideration with respect to other bodies. It must be demanded that the inertial resistance of a body could be increased by having unaccelerated inertial masses arranged in its vicinity; and this increase of the inertial resistance must disappear again if these masses accelerate along with the body.

Einstein then proceeds to claim that the 1913 theory does indeed contain an effect of the desired kind.

The above passage is remarkable on two scores. First, there is the already noted incorrect claim that Mach was concerned about the definition of inertial resistance. Second, Einstein states that both velocity and acceleration are relative and presents this as a major problem. However, he never once in his papers attempted to show how general relativity attacked the fundamental kinematic problem of the relativity of velocity. The idea that inertial resistance, like acceleration, must be relative, is expressed very prominently in Einstein's writings from 1912 through to about 1922. However, Einstein never once attempted to show how such an idea (and still less the even more fundamental relativity of motion alluded to above) was implemented in the basic kinematic and dynamic structure of the theory he was constructing.

This is in very striking contrast to the epistemological work of Neumann, Tait, Lange, and Poincaré and the manifestly relational proposals of Hofmann and Reissner. All of these authors attacked the relativity of motion head on. What are the reasons for Einstein's conspicuous failure to follow their example?

4. EINSTEIN'S PRIORITIES WHEN CREATING GENERAL RELATIVITY

Let me now attempt to begin to answer the question with which the previous section ended by considering the evidence that can be gleaned from Einstein's early papers and correspondence. It is quite clear that by the time he had left school and commenced university studies Einstein had set himself a supremely ambitious task. He was going to attack and make an extremely serious attempt to solve the great topical problems of physics. In later years, he may have liked to cultivate the image of a somewhat indolent student, but a very different picture emerges from his correspondence. There were certain fundamental issues that he followed avidly, above all anything related to Maxwellian field theory and also anything that could provide evidence for the existence of atoms. These were the burning topics of the time, and he followed them closely.

It seems to me that with regard to the absolute *vs* relative debate, the situation was somewhat different. There is no doubt that it was a topic of genuine widespread interest; Poincaré's inclusion of it in *La Science et l'Hypothèse* is clear evidence of that. However, it was a topic with relatively few (but by no means none at all) opportunities for decisive experimental tests;[11] both Mach and Poincaré tended to treat the topic in a rather passive manner, drawing attention to problems but without proposing an energetic programme for their resolution. For an ambitious young man like Einstein, with a strong awareness of the importance of experiment and clearly determined to make a name for himself as quickly as he could, the problems of electromagnetism and atomism must surely have appeared to offer far better prospects. This could well explain why Einstein's imagination was clearly caught, through his reading of Mach's *Mechanik* around 1898 (CPAE 1), by the great issue of absolute space without this leading him on to a more detailed consideration of the details. Whatever the reason, in the period 1898 to 1905 (and, indeed, up to the end of his life) Einstein had the opportunity to go into the details and really come to grips with the central problems of defining time, clocks, and motion. He did not or, at least, not directly (except, of course, with regard to simultaneity).

There are, I believe, at least three clearly identifiable reasons for Einstein's *indirect* attack on the problem of absolute space. All three are important and interrelated and already played a decisive role in his creation of the *special* theory of relativity.

The first, and surely the most important, is that the principle of Galilean relativity suggested to Einstein an indirect but extremely effective way of making absolute space redundant in physics. He saw the success of special relativity as an important first step in that direction and then attempted, with great consistency, to generalize the relativity principle to the maximum extent possible. He believed that this would make absolute space completely redundant as a concept in physics.

The second reason for Einstein's indirect strategy is to be found in the phenomenological concept of the rigid body and the important work done by Helmholtz on the empirical foundations of geometry. The phenomenological rigid body played a vitally important role in both special and general relativity but, as we shall see, made it extremely difficult to address directly the relativity of motion in any obviously Machian manner.

The third reason for Einstein's indirect approach may seem somewhat surprising at the first glance—it was Planck's discovery of the quantum of action in 1900. We shall see that this discovery greatly diminished Einstein's confidence in the possibil-

11 For a discussion of early experiments, see (Norton 1995).

ity of finding quickly any explicit and detailed dynamical equations that could be taken to describe the behavior of particles and fields at the fundamental microscopic level. Instead, he consciously sought general principles such as those established in phenomenological thermodynamics by means of which he could obtain constraints on the behavior of matter. This strengthened his faith in the value of the relativity principle and his indirect approach to implementation of Mach's ideas. It also persuaded him that it would be useless to attempt to construct a microscopic theory of rods and clocks.

Let me now expand on these three points in more detail.

It seems to me entirely possible that an overall strategy for eliminating absolute space from physics started to take shape in Einstein's mind very soon after he had read Mach's *Mechanik* around 1898. The basic idea arose from consideration of a problem that Mach had not considered at all: electrodynamics. Much of the later development of relativity theory is clearly prefigured in a comment of Einstein to his future wife in a letter written in August 1899 (CPAE 1):

> I am more and more convinced that the electrodynamics of moving bodies, as presented today, is not correct, and that it should be possible to present it in a simpler way. The introduction of the term "aether" into the theories of electricity led to the notion of a medium of whose motion one can speak without being able, I believe, to associate a physical meaning with this statement.

This train of thought then led on to the clear formulation in 1905 of the relativity principle, in accordance with which uniform motion relative to the supposed aether is completely undetectable. As Einstein (Einstein 1905) famously remarked, this then meant that "the introduction of a 'luminiferous aether' will prove to be superfluous". Moreover, by the end of the 19th century, the aether had more or less come to be identified with absolute space, a rigid substrate that besides being the carrier of electromagnetic excitations also served as the ultimate standard of rest for all bodies in the universe. In his famous 1895 paper on electrodynamics with which Einstein was certainly familiar, Lorentz said of the aether (Lorentz 1895, 4): "When for brevity I say that the aether is at rest this means merely that no part of this medium is displaced relative to any other part and that all observable motions of the heavenly bodies are relative motions with respect to the aether."[12]

Having banished the aether from the foundations of physics, Einstein felt that he had made an important first step on the way to the complete elimination of the notion of absolute space. Einstein felt that a thing could only be said to exist if it had observable effects. The 1905 relativity principle showed that *uniform* motion relative to the putative aether (or absolute space) had no observable consequences. If the relativity principle could be extended further, to all accelerated motions, then all residual arguments for the existence of absolute space would be eliminated. Einstein's 1933 Gibson lecture (Einstein 1933) suggests rather strongly that this train of thought had

---

12 It is worth noting that this is a remarkably naive concept of motion compared with the subtlety of Lange's construction.

taken shape in Einstein's mind already by 1905, but that at that stage he was unable to take the idea any further. It was only in autumn 1907 (Einstein 1907) that the potential of what he later called the equivalence principle struck him; for it suggested that the restricted principle of relativity could be extended from uniform motions to *uniformly accelerated* motions as well. This then opened up the prospect of extension of the relativity principle even further—to all motions whatsoever.

This logic is spelled out very clearly in the Gibson lecture, from which the following quotation is taken:

> After the special theory of relativity had shown the equivalence for formulating the laws of nature of all so-called inertial systems (1905) the question of whether a more general equivalence of coordinate systems existed was an obvious one. In other words, if one can only attach a relative meaning to the concept of velocity, should one nevertheless maintain the concept of acceleration as an absolute one? From the purely kinematic point of view the relativity of any and every sort of motion was indubitable; from the physical point of view, however, the inertial system seemed to have a special importance which made the use of other moving systems of coordinates appear artificial.
>
> I was, of course, familiar with Mach's idea that inertia might not represent a resistance to acceleration as such, so much as a resistance to acceleration relative to the mass of all the other bodies in the world. This idea fascinated me; but it did not provide a basis for a new theory.

Note how Einstein insists that the idea of a more general equivalence of coordinate systems "was an obvious one". It certainly was not so to his contemporaries. If there is one aspect of Einstein's work on gravitation that most clearly distinguished him from them all, it was his insistence on the need to generalize the relativity principle and on the equivalence principle as the means to do so. All of the truly original steps which eventually led Einstein to the general theory of relativity sprang from this conviction. It is certainly the case that Mach's vehement opposition to Newton's absolute space as a nonexistent monstrosity was completely shared by Einstein and served as the main stimulus to the creation of general relativity.

However, it is important to note that the two men disliked absolute space for rather different reasons. Mach tended very much to concentrate on the things in the world that could be directly observed—bodies—and on the relationships between them, which were expressed in the first place by the mutual separations between them. This gut instinct is expressed very clearly in Mach's famous comment (Mach 1960): "The world is not *twice* given, with an earth at rest and an earth in motion, but only *once*, with its *relative* motions, along determinable." Given Einstein's great enthusiasm for Mach, it is remarkably difficult to find evidence which shows unambiguously that Einstein understood what Mach really wanted to do: base mechanics solely on the relative separations of bodies. As I have already noted, many of Einstein's remarks about Mach actually represent a distortion of the older man's thought. The passage from 1918 quoted earlier is a clear expression of what Mach wanted to do ("introducing in the laws of classical mechanics only distances of material points from each other"), but there is no direct attribution to Mach. The solitary direct attribution I have found is in a very late letter to Pirani,[13] in which Einstein says [my translation]:

> There is much talk of Mach's Principle. It is, however, not easy to associate a clear notion with it. Mach's stimulus was this. It is unbearable [unerträglich] that space (or the inertial system) influences all ponderable things by determining the inertial behaviour without the ponderable things exerting a determining reaction back on space. Mach rediscovered what Leibniz and Huygens had correctly faulted in Newton's theory. He sought to eliminate this evil by attempting to abolish space and replace it by the relative inertia of the ponderable bodies with respect to each other. Space should be replaced by the distances between the bodies taken in pairs (with these distances as independent concepts). This evidently did not work, quite apart from the fact that the time with its absolute nature remained.

Several comments can be made about this opening paragraph of Einstein's letter.

First, there is Einstein's admission that it is not easy to associate a clear notion with Mach's Principle. This, however, is what the criterion of Poincaré considered earlier does do.

Second, the idea that something should not be able to influence another thing without suffering a back reaction on itself is not, so far as I know, to be found anywhere in Mach's writings. It is, however, an idea that Einstein himself frequently advanced from around 1914, mainly I think as a result of his work on Nordström's theory, in which the propagation of light is governed by an absolute structure and is not subject to the influence of gravitation. For example, in 1914 (Einstein 1914) he wrote: "It seems to me unbelievable that the course of any process (e.g., that of the propagation of light in a vacuum) could be conceived of as independent of all other events in the world."

Next, it should be noted that even the account of what Mach proposed is not strictly correct, since Mach did not propose to eliminate the relations of Euclidean space and regard the distances between bodies as completely independent. It should be noted that for n bodies in Euclidean space there are $n(n-1)/2$ mutual separations, of which only $3n-6$ are independent (for $n \geq 3$). In his proposal for a new law of inertia, Mach did not include any suggestion that this basic fact of three-dimensional Euclidean geometry should be relaxed. However, in a very early paper he had speculated (Mach 1872, 25; 1911, 51–53) that such a relaxation might occur in the interior of atoms and play an important role in the formation of spectral lines. He later explicitly withdrew (Mach 1911, 94) this theoretical speculation, which hinged on a putative representation of atoms and molecules in Euclidean spaces of more than three dimensions. Einstein read Mach's booklet on the *Conservation of Energy*, where the idea is discussed, in 1909,[14] so it is possible that in his old age he muddled it up with Mach's proposals for inertia.

Finally, we note in Einstein's "this evidently did not work" an echo of the comment in 1918 that the proposal to found mechanics solely on relative separations had not proved feasible. However, the papers of Hofmann, Reissner, and Schrödinger had shown the approach to be perfectly feasible. With the possible exception of Reiss-

13 Einstein to Felix Pirani, 1954 (EA 17–447).

14 As we know from a letter Einstein wrote to Mach in August 1909 (CPAE 5E, Doc. 174).

ner's work, Einstein had read these papers. It therefore seems that they made very little impression on him; he was certainly confused about their content, since all demonstrated the Machian approach *was* feasible. Had Einstein from the beginning shared Mach's gut instinct that only relative separations count and that the central problem was to reflect this in the foundations of mechanics, then surely he might have been expected to have taken more notice of what had been achieved.

The fact that he did not suggests that Einstein's objection to absolute space had a somewhat different psychological origin. For this conclusion, there is much evidence. Rather than consider objects in space, Einstein was evidently wont to contemplate the notion of empty space by itself. Evidence for this can be found, for example, in (Einstein 1921). Given the perfect uniformity of space, Einstein then found it an affront to the principle of sufficient reason that such a featureless thing should contain within it *distinguished frames of reference* for the formulation of the laws of nature. Numerous arguments on such lines can be found in Einstein's papers from 1913 up to the *Autobiographical Notes*. They always invoke the point mentioned in the 1933 Gibson lecture—that "from the purely kinematic point of view the relativity of any and every sort of motion [in space] was undubitable." Thus, the only way to create a theory perfectly in accord with the principle of sufficient reason was through generalization of the relativity principle to the absolutely greatest extent possible.

The difference between Mach and Einstein can be summarized very simply: Mach wanted to eliminate coordinate systems entirely, Einstein wanted to show that all coordinate systems were equally valid. Given the tremendous success of the special theory of relativity, which established the equivalence of all coordinate systems in uniform motion relative to each other, and the promise offered by the equivalence principle for extension to accelerated motion, it is very easy to see why Einstein became so totally committed to his approach and took virtually no notice of the alternative.

The question of whether one (and then which one) or both of these two approaches are valid is very complex. It is a subject that I cannot follow further in this paper, in which I have set myself the more modest task of identifying some characteristic differences between the approaches of Mach (and his contemporaries) and Einstein, finding the reasons for Einstein's choices, and placing his work in the perspective of other work on the absolute *vs* relative debate. Let me just say that, in my opinion, Mach's approach (augmented by Poincaré's analysis) is deeper and more consistent than Einstein's but that nevertheless Einstein's theory, when properly analyzed as a dynamical theory, does perfectly implement Mach's ideas. However, the reason for this has more to do with deep intrinsic properties of the absolute differential calculus, which Einstein took over 'ready made' from the mathematicians, than with Einstein's covariance arguments. All this will be spelled out in my forthcoming monograph (Barbour, in preparation). See also my *Notes Added in Proof* at the end of this article.

Because it ties in very well with Einstein's conception of space that we have just been considering, let me now turn to the role played by the rigid body and Helmholtz's work on the empirical foundation of geometry (Helmholtz 1868) in Einstein's development of both special and general relativity. Helmholtz's study was made about a decade after Riemann's famous habilitation lecture of 1854 (Riemann 1867). Initially Helmholtz was unaware of Riemann's work, which was not published until 1867, and found that he had largely rediscovered already known results. However, there was one respect in which Helmholtz went significantly beyond Riemann. This concerned the hypothesis that Riemann had made for the form of the line element.

Riemann had assumed, more or less on grounds of simplicity and to match Pythagoras's theorem, that the fundamental line element ds of his generalized geometry should be the square of a *quadratic* form in the coordinate differences. He noted, however, that *a priori* one could not rule out, say, taking the fourth root of a quartic form. In contrast, Helmholtz considered the empirical realization of geometry by rigid bodies and congruence relations between them. If such bodies are to be brought to congruence, they must satisfy certain conditions of mobility and remain congruent in different positions and different orientations. Their congruence must also be independent of the paths by which they are brought to congruence. Helmholtz was able to show that if these conditions are to be met then the quadratic form of the line element adopted by Riemann as a simplicity hypothesis is indeed uniquely distinguished. This established a very beautiful connection between empirical geometry based on physical measuring rods and a particular mathematical formalism. Helmholtz concluded his important paper with the following words:

> the whole possibility of the system of our space measurements ... depends on the existence of natural bodies that correspond sufficiently closely to the concept of rigid bodies that we have set up. The fact that congruence is independent of position, of the direction of the objects brought to congruence, and of the way in which they have been brought to each other—that is the basis of the measurability of space.

The influence of Helmholtz's study is manifest throughout Einstein's entire relativity opus. In a newspaper article published in 1926, Einstein (Einstein 1926) described the practical geometry of the experimental physicist in which "rigid bodies with marks made on them realize, provided certain precautions are taken, the geometrical concept of interval" and said [my italics]:

> Then the geometrical "interval" corresponds to a definite object of nature, and thus all the propositions of geometry acquire the nature of assertions about real bodies. This point of view was particularly clearly expressed by Helmholtz; *one may add that without this viewpoint it would have been practically impossible to arrive at the theory of relativity.*

The Helmholtzian conception was crucial for two reasons in particular: It provided a definite framework in which Einstein could comprehend length contraction and simultaneously gave a method for *position determination*. It also gave Einstein a way of 'making space visible' that perhaps made him less concerned than Mach about the problems of position determination. Taken together, these factors led Einstein to a method of position determination that appears to be decidedly un-Machian.

Indeed, a complete method of position coordination appeared already in the famous Kinematical Part of his 1905 paper (Einstein 1905). Einstein opens that section as follows: "Let us take a system of coordinates in which the equations of Newtonian mechanics hold good." Thus, he simply *presupposes* the outcome of a Tait-Lange procedure for finding an inertial frame of reference. He then continues:

> If a material point is at rest relatively to this system of coordinates, its position can be defined relatively thereto by the employment of rigid standards of measurement and the methods of Euclidean geometry, and can be expressed in Cartesian coordinates.

The standards of measurement (Helmholtz's rigid bodies) serve two supremely important purposes. First, they can be imagined to "fill" the whole space of the inertial system; since one can also suppose that the bodies carry marks permanently scratched on them, other bodies can be unambiguously located by means of these marks. Space has been made visible. Then, second, comes the really great convenience of such rigid bodies—intervals defined by the marks on them satisfy the congruence conditions required in Helmholtz's phenomenological foundation of geometry. Thus, the coordinates can be associated with the marks on the rigid bodies in such a way that they simultaneously *give physical distances directly.*

Convenient as all this is, it still does not contain anything that goes beyond Helmholtz's scheme. However, the scheme turned out to be wonderfully adapted to the exigencies of relativity theory and length contraction. Here the important thing is the underlying conception Einstein had of what might be called the true physical nature of rigid bodies (and therefore of measuring rods). He certainly did not think of them as ultimate elements incapable of further explication. On the contrary, Einstein was a convinced atomist and he conceived of a measuring rod as being made up of a definite number of atoms governed by quite definite laws of nature. Provided external circumstances (pressure, temperature, etc.) remained the same, such a system of atoms could be expected to 'settle into' a unique equilibrium configuration. Two such systems constituted by identical collections of atoms would settle into the same configuration and therefore be congruent to each other. Thus, Helmholtz's phenomenological foundation of geometry would have a theoretical underpinning in atomism and the laws governing it.

A vital part in this overall picture was played by the notion of an inertial system coupled with Einstein's formulation of the (restricted) relativity principle, in accordance with which the laws of physics must have the identical form in all inertial systems obtained from each other by a uniform translational motion. Coupled with Einstein's (long tacit but later explicit (Einstein 1923)) atomistic conception, the relativity principle ensured that the identical phenomenological Helmholtzian geometry must be realized in each inertial system. However, it left open the connection between the geometries (and chronometry, which I have not considered here) in different inertial systems. The Helmholtzian scheme had *just* enough flexibility to allow and accommodate those marvellous *bombes surprises* of relativity: length contraction and time dilation. Moreover, the underlying atomistic conception meant that one could still talk about 'the same measuring rod' in two different inertial systems. One merely had to suppose two rods constituted of the same atoms and subject to the same external conditions. Then in their respective inertial systems they would settle into identical configurations, and comparison of these configurations between inertial systems would become epistemologically valid. One would be talking about the 'same things.'

Another extraordinary convenience of this whole conception was that it could be generalized almost unchanged to the framework of the general theory of relativity. It was merely necessary to repeat the entire exercise, not globally, but 'in the small.' Very important here was the presumed existence of local approximations to inertial systems in the neighborhood of every point of spacetime; for the distinguished 'equilibrium' configurations into which rods could be assumed to 'settle' exist only in an inertial system.

Einstein's attitude to his phenomenological treatment of rods and clocks is summarized very clearly in his *Autobiographical Notes*:[15]

> One is struck [by the fact] that the theory (except for the four-dimensional space) introduces two kinds of physical things, i.e., (1) measuring rods and clocks, (2) all other things, e.g., the electro-magnetic field, the material point, etc. This, in a certain sense, is inconsistent; strictly speaking measuring rods and clocks would have to be represented as solutions of the basic equations (objects consisting of moving atomic configurations), not, as it were, as theoretically self-sufficient entities. However, the procedure justifies itself because it was clear from the very beginning that the postulates of the theory are not strong enough to deduce from them sufficiently complete equations for physical events sufficiently free from arbitrariness, in order to base upon such a foundation a theory of measuring rods and clocks. If one did not wish to forego a physical interpretation of the coordinates in general (something which, in itself, would be possible), it was better to permit such inconsistency—with the obligation, however, of eliminating it at a later stage of the theory. But one must not legalize the mentioned sin so far as to imagine that intervals are physical entities of a special type, intrinsically different from other physical variables ("reducing physics to geometry," etc.).

Before commenting on this passage and its bearing on the central issues of the absolute *vs* relative question, let us consider the third factor that I identified as shaping Einstein's overall strategy in the creation of both relativity theories: the quantum. In the above passage, Einstein merely says "it was clear from the very beginning that the postulates of the theory are not strong enough to deduce ... a theory of measuring rods and clocks." However, while this statement is obviously true it is at the same time something of an inversion of the actual historical development. The fact is that Einstein *deliberately*, already in the period 1904/5, chose to develop his ideas on the basis of very general postulates. His reasons for doing so are very well known and were explained by Einstein himself in the *Autobiographical Notes*.

The truth is that Planck's discovery of the quantum of action in 1900 made a tremendous impression on Einstein and quickly persuaded him that some very strange things indeed must be happening at the microscopic level. In particular, he was certain that Maxwell's equations (for which he had the very greatest respect) could not be valid in their totality for microscopic phenomena. This comes out graphically in a letter that Einstein wrote to von Laue in January 1951:[16]

---

15 He had already made very similar comments in (Einstein 1923).

16 Einstein to Max von Laue, January 1951 (EA 16-154).

> If one goes through your collection of the verifications of the special theory of relativity, one gets the impression that Maxwell's theory is secure enough to be grasped. But in 1905 I already knew for certain that it yields false fluctuations of the radiation pressure and thus an incorrect Brownian motion of a mirror in a Planck radiation cavity.

It is well known that Einstein's complete conviction that Maxwell's theory could be at best partly right was a decisive factor in his selection, as a postulate in the foundations of his special theory of relativity, of the one minimal part of Maxwellian theory in which he did retain confidence: the light propagation postulate.

More generally, it led him at the same time to formulate consciously the idea of a theory based on principles that wide experience had demonstrated had universal validity. The classic paradigms of such principles were the denials of the possibility of construction of perpetual motion machines of the first and second kind, which played such a fruitful role in the phenomenological thermodynamics that had been created around the middle of the 19th century. The great strength of such theories was that they enabled one to make many important predictions without attempting to find a detailed theory at a fundamental microscopic level. Such a theory Einstein called a *constructive* theory, in contrast to a theory based on principles. In a very clear account of the distinction between the two kinds of theory that he included in a piece that he wrote for *The Times* (Einstein 1919), Einstein said that the theory of relativity was one of the latter kind, which he called *fundamental*.

Let me now conclude by considering some of the consequences of these three aspects of Einstein's overall strategy—the programme to eliminate absolute space by generalization of the relativity principle, the use of Helmholtzian rigid bodies to define position, and the eschewal of constructive theories. Both together and separately, they had the consequence that virtually all the issues that one might have expected to feature prominently in a frontal attack on the absolute *vs* relative question—and that was certainly a very major part of Einstein's undertaking—were actually missing as explicit elements in Einstein's work. It is almost a case of Hamlet without the Prince of Denmark.

One of the biggest problems with Einstein's approach is that distinguished frames of reference figured crucially in his work, but he never explicitly considered their status and origin. For example, in his work on special relativity he would invariably start by assuming the existence of inertial frames of reference and then postulate the existence of laws of nature that had to be expressed relative to these frames of reference. Once this step had been taken, the relativity principle could come into play—the laws of nature must take the same form in all frames of reference related by Lorentz transformations. It is however legitimate to ask what determines the frames of reference: Are there laws of nature that determine them? The whole logic of Einstein's approach made it impossible for him to pose, let alone answer, this question, since the frames of reference had to be there before he could formulate the laws of nature. Einstein bequeathed us an unresolved vicious circle at the heart of his theory.

It should not be thought that the transition to general relativity, in which all frames of reference are purportedly allowed, eliminates this problem. The fact is that the *local* existence of distinguished frames of reference (locally inertial frames) is an absolute prerequisite of the theory, since it is only when such frames exist that Einstein's phenomenological treatment of rods and clocks can be taken over from special relativity. But Einstein never once seems to have considered seriously how the local frames of reference and his rods and clocks could arise from first principles.

Of course, we know that he was at least partly aware of this issue, since he more than once said that one should elaborate a theory of rods and clocks. However, I suspect that Einstein did not quite appreciate the true nature of the problem, which is already evident from Neumann's theory of the inertial clock. This showed clearly that there is a *twofold* problem. First, one must have access to an inertial system; second, one must track some physical object or process whose behavior stands in a known one-to-one relationship to the framework defined by the inertial system. In the simplest case of Neumann's inertial clock, this is done directly by the motion of a force-free particle. Of these two problems, the first is by far the most difficult; indeed, the second problem is effectively solved together with the first, the very posing of which is a decidedly subtle matter (as the long and painful elaboration of the problem demonstrates).

If we now examine Einstein's relatively terse comments about rods and clocks, rather strong evidence emerges that his understanding of the issue never really advanced beyond the level of Neumann's inertial clock defined in a *known* inertial system. For example, in the same *Autobiographical Notes* from which the earlier quotation about rods and clocks was taken, Einstein refers to a clock as an 'in itself determined periodic process realized by a system of sufficiently small spatial extension' and then shortly afterwards comments:

> The presupposition of the existence (in principle) of (ideal, viz, perfect) measuring rods and clocks is not independent of each other; since a light signal, which is reflected back and forth between the ends of a rigid rod, constitutes an ideal clock, provided that the postulate of the constancy of the light-velocity in vacuum does not lead to contradictions.

From this it is clear that Einstein already presupposed knowledge of the positions of objects constituting his model clocks in an inertial frame of reference—for if the rigid rod is not moving strictly inertially, Einstein's ideal light clock is useless. All the evidence I have examined is consistent with my conclusion that Einstein never grasped this fact and that he did not properly understand the problem posed by determination of the inertial frames of reference. His disparaging remarks to von Laue about Lange's work are strong support for this view.

This is not deny the correctness of Einstein's supposition that the quantum problems made it premature to try and make truly realistic microscopic models of rods and clocks. But what Einstein had in mind was the theory of such things in *a known inertial frame of reference*, whereas the more fundamental problem concerned the origin of the frame. And to address that problem he did not really need such advanced physics. The quantum bogey gave Einstein a valid excuse for not constructing microscopic theories of actual physical clocks, but may have misled him by seeming to locate the problem in the wrong place.

Mention should also be made here of the rather vague way in which Einstein formulated the relativity principle. He invariably simply said that the laws of nature, the form of which he deliberately left vague, must take the same form in all frames of reference. He never attempted anything like the subtly refined formulation proposed by Poincaré based on identification of the kind of initial data needed to predict the future uniquely. A strength of Poincaré's approach is that it avoids the vicious circle inherent in Einstein's approach whereby the status and origin of the necessary distinguished frames (the local existence of which is still needed in general relativity despite the general covariance of that theory) is left open. In Poincaré's approach, the question of the distinguished frames of reference does not arise, since he formulated the initial-value problem deliberately in such a way that they do not enter into it at all. For some reason or other, Einstein never seems to have thought of general relativity as a dynamical theory and Poincaré's comments seem to have made no impression on him.

Finally, we must consider the effect of Einstein's Helmholtzian method of position determination. It was this above all that precluded any directly Machian implementation of relativity of position and velocity after the manner of Hofmann and Reissner. For them, like Mach, position and velocity were determined by the set of distances to all other bodies in the universe. But Einstein was completely and irrevocably tied to local position determination by means of Helmholtzian rigid bodies that 'filled' the space of inertial systems, which Mach insisted must be understood as arising from relations to other matter in the universe, whereas Einstein simply took them as given. Thus, Einstein's technique was doubly un-Machian. In the 1918 paper quoted in Sec. 2, Einstein said that "the historical development" had shown that it was not possible to "dispense with coordinate systems." For 'historical development' we must here understand the foundations of Einstein's own work: coordination by Helmholtzian rigid bodies and relativity transformations between such coordinates. Note also that in the passage cited earlier from his 1913 paper Einstein said:

> It is obvious that one cannot ascribe an absolute meaning to the concept of acceleration of a material point, no more so than one can ascribe it to the concept of velocity. Acceleration can only be defined as relative acceleration of a point with respect to other bodies.

In this passage, Einstein is quite clearly saying that position and velocity are defined relative to other bodies in exactly the same sense as did Mach (and all the other relationists). Yet he did not give any indication how that requirement was implemented in his own theory. He merely pinned his hopes on a *resistance* to acceleration induced by distant matter. These hopes came to nothing—and meanwhile the Machian issues were never directly addressed.

However, general relativity is an immensely rich and sophisticated theory, and the same can be said of the veritable odyssey of its discovery by Einstein. One can find ironies, serendipity, and utter brilliance throughout the entire saga. Just because Einstein's chosen route to the creation of general relativity did not directly address certain central issues of the absolute *vs* relative debate, this does not necessarily mean that his wonderful theory fails to solve them. After all, Newton posed some critical problems in the Scholium at the time he wrote the *Principia*, but some two hundred years passed before they were more or less completely resolved within the framework of Newtonian theory. As already indicated, I am convinced that the problems considered in this paper, which has considered alternative issues that Einstein might have addressed (and his contemporaries, above all Poincaré, did address), are actually resolved within the heart of general relativity by virtue of the exquisite mathematics on which it is based. However, that is quite a long story too and will have to be considered elsewhere (Barbour, in preparation).

NOTES ADDED IN PROOF

Since writing this article, I have continued to research and write about Mach's principle and related issues. This activity has generated much material that relates to the topics discussed in this article and envisaged for (Barbour, in preparation). First, there is my book *The End of Time* (Barbour 1999a), which considers the quantum cosmological implications of a relational treatment of time and motion. Second, the three review papers (Barbour 1997, 1999b, 2001) complement the present paper. Third, collaboration with Edward Anderson, Brendan Foster, Bryan Kelleher and Niall Ó Murchadha has resulted in the publication of about ten research papers, of which I mention (Barbour, Foster and Ó Murchadha 2002, Anderson and Barbour 2002, Barbour 2003a, Anderson, Barbour, Foster and Ó Murchadha 2003, Barbour 2003b).

These research papers take very much further the approach to Mach's principle initiated Hofmann, Reissner and Schrödinger as modified in (Barbour and Bertotti 1982) through the use of best matching (Sec. 2) to avoid anisotropy of inertial mass. The relational treatment of time, implemented through reparametrization invariance, also plays a central role. If one assumes that space is Riemannian and that all interactions are local, the two principles of *best matching* and *reparametrization invariance* lead almost uniquely to Einstein's general theory of relativity. Quite unexpectedly, they also enforce the existence of a universal lightcone and gauge theory. One starts with the notion of Riemannian space (*not* pseudo-Riemannian spacetime) and three-dimensional fields (scalar, spinor and vector) defined on it. Then implementation of the Machian principles of the relativity of motion (through best matching) and time (through reparametrization) creates a four-dimensional spacetime with all the basic features of modern physics. I believe that my claim that general relativity is perfectly Machian (as regards the relativity of time and motion) is strongly vindicated.[17]

Another issue that we have investigated is relativity of scale (Barbour 2003a, Anderson, Barbour, Foster and Ó Murchadha 2003). The same intuitive convictions that lead one to require relativity of time and motion suggest that physics ought to be scale invariant too. In the two cited papers, we have extended the notion of best matching to scale transformations. We have shown that general relativity is almost scale invariant but not quite perfectly so. Specifically, in the case of a spatially closed universe one can change the (spatial) scale arbitrarily at all points. However, this must be done subject to the solitary requirement that these local transformations do

17 See, especially (Barbour, Foster and Ó Murchadha 2002).

not change the spatial volume of the universe. This remarkably weak restriction is, in fact, what permits expansion of the universe to be a meaningful concept. Modern cosmology depends crucially on this single residual 'Machian defect.' Work on this topic and our general approach, which we call the 3-space approach, is continuing. My personal feeling is that we are close to a definitive formulation of the principles and main conclusions of the 3-space approach. I may at last be in the position to complete (Barbour, in preparation)! In fact, I put aside work on it because I felt that it would not be complete without a proper Machian treatment of the relativity of scale.

ACKNOWLEDGMENTS

I am grateful to Domenico Giulini for drawing my attention to Frege's paper and to the Albert Einstein Archives at the Hebrew University of Jerusalem for permission to quote from the Einstein correspondence.

REFERENCES

Anderson, Edward and Julian B. Barbour. 2002. "Interacting Vector Fields in Relativity without Relativity." *Classical and Quantum Gravity* 19: 3249–3261 (gr-qc 0201092).

Anderson, Edward, Julian B. Barbour, Brendan Z. Foster, and Niall Ó Murchadha. 2003. "Scale-Invariant Gravity: Geometrodynamics." *Classical and Quantum Gravity* 20: 3217–3248 (gr-qc 0211022).

Barbour, Julian B. 1989. *Absolute or Relative Motion,* Vol. 1. *The Discovery of Dynamics*. Cambridge: Cambridge University Press. Reprinted 2001 as the paperback *The Discovery of Dynamics*. New York: Oxford University Press.

———. 1994. "The Timelessness of Quantum Gravity. I. The Evidence from the Classical Theory. II. The Appearance of Dynamics in Static Configurations." *Classical and Quantum Gravity* 11: 2853–2897.

———. 1995. "General relativity as a Perfectly Machian Theory." In (Barbour and Pfister 1995).

———. 1997. "Nows Are All We Need." In H. Atmanspacher and E. Ruhnau (eds.), *Time, Temporality, Now.* Berlin: Springer-Verlag.

———. 1999a. *The End of Time*. London: Weidenfeld & Nicholson: New York: Oxford University Press (2000).

———. 1999b. "The Development of Machian Themes in the Twentieth Century." In J. Butterfield (ed.), *The Arguments of Time*. Oxford: Oxford University Press.

———. 2001. "On General Covariance and Best Matching." In C. Callender and N. Huggett (eds.), *Physics Meets Philosophy at the Planck Scale*. Cambridge: Cambridge University Press.

———. 2003a. "Scale-Invariant Gravity: Particle Dynamics." *Classical and Quantum Gravity* 20: 1543-1570 (gr-qc 0211021).

———. 2003b. "Dynamics of Pure Shape, Relativity and the Problem of Time." In H.-T. Elze (ed.), *Decoherence and Entropy in Complex Systems. Lecture Notes in Physics*, Vol. 663. Berlin: Springer-Verlag (gr-qc 0308089).

———. (in preparation). *Absolute or Relative Motion?* Vol. 2: *The Frame of the World*. New York: Oxford University Press.

Barbour, Julian B. and Bruno Bertotti. 1977. "Gravity and Inertia in A Machian Framework." *Nuovo Cimento* 38B: 1–27.

———. 1982. "Mach's Principle and the Structure of Dynamical Theories." *Proceedings of the Royal Society of London,* Series A 382: 295–306.

Barbour, Julian B., Brendan Z. Foster and Niall Ó Murchadha. 2002. "Relativity without Relativity." *Classical and Quantum Gravity* 19: 3217–3248 (gr-qc 0012089).

Barbour, Julian B. and Herbert Pfister (eds.). 1995. *Mach's Principle: From Newton's Bucket to Quantum Gravity. Einstein Studies,* Vol. 6. Boston: Birkhäuser.

Clemence, G.M. 1957. "Astronomical Time." *Reviews of Modern Physics* 29, 2–8.

CPAE 1: John Stachel, David C. Cassidy, Robert Schulmann, and Jürgen Renn (eds.), *The Collected Papers of Albert Einstein*. Vol. 1. *The Early Years, 1879–1902*. Princeton: Princeton University Press, 1987.

CPAE 2. 1989. John Stachel, David C. Cassidy, Jürgen Renn, and Robert Schulmann (eds.), *The Collected Papers of Albert Einstein*. Vol. 2. *The Swiss Years: Writings, 1900–1909*. Princeton: Princeton University Press.

CPAE 4. 1995. Martin J. Klein, A. J. Kox, Jürgen Renn, and Robert Schulmann (eds.), *The Collected Papers of Albert Einstein*. Vol. 4. *The Swiss Years: Writings, 1912–1914*. Princeton: Princeton University Press.

CPAE 5E: *The Collected Papers of Albert Einstein*. Vol. 5. *The Swiss Years: Correspondence, 1902–1914*. English edition translated by Anna Beck, consultant Don Howard. Princeton: Princeton University Press, 1995.

CPAE 6. 1996. A. J. Kox, Martin J. Klein, and Robert Schulmann (eds.), *The Collected Papers of Albert Einstein*. Vol. 6. *The Berlin Years: Writings, 1914–1917*. Princeton: Princeton University Press.

CPAE 7. 2002. Michel Janssen, Robert Schulmann, József Illy, Christoph Lehner, and Diana Kormos Buchwald (eds.), *The Collected Papers of Albert Einstein*. Vol. 7. *The Berlin Years: Writings, 1918–1921*. Princeton: Princeton University Press.

Dziobek, Otto. 1888. *Die mathematischen Theorien der Planeten-Bewegungen*. Leipzig: J. A. Barth.

———. 1892. *Mathematical Theories of Planetary Motions*. Register Publishing Company. Reprinted by Dover in 1962.

Einstein, Albert. 1905. "Zur Elektrodynamik bewegter Körper." *Annalen der Physik* 17, 891–921, (CPAE 2, Doc. 23). English translation in (Lorentz et al. 1923).

———. 1907. "Über das Relativitätsprinzip und die aus demselben gezogenen Folgerungen." V. §17. *Jahrbuch der Radioaktivität und Elektronik* 4: 411–462, (CPAE 2, Doc. 47).

———. 1912. "Gibt es eine Gravitationswirkung, die der elektrodynamischen Induktionswirkung analog ist?" *Vierteljahrschrift für gerichtliche Medizin und öffentliches Sanitätswesen* 44: 37–40, (CPAE 4, Doc. 7).

———. 1913a. "Zum gegenwärtigen Stande des Gravitationsproblems." *Physikalische Zeitschrift* 14: 1249–1266, (CPAE 4, Doc. 17). (English translation in this volume.)

———. 1913b. "Physikalische Grundlagen einer Gravitationstheorie." *Vierteljahrsschrift der Naturforschenden Gesellschaft Zürich* 58: 284–290, (CPAE 4, Doc. 16).

———. 1914. "Die formale Grundlage der allgemeinen Relativitätstheorie." *Sitzungsberichte der Preussischen Akademie der Wissenschaften*, 1030–1085, Part 2, (CPAE 6, Doc. 9).

———. 1917. "Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie." *Sitzungsberichte der Preussischen Akademie der Wissenschaften*: 142–145, (CPAE 6, Doc. 43).

———. 1918a. "Prinzipielles zur allgemeinen Relativitätstheorie." *Annalen der Physik* 55: 241–244, (CPAE 7, Doc. 4).

———. 1918b. "Dialog über Einwände gegen die Relativitätstheorie." *Die Naturwissenschaften* 6: No. 48, 697–702, (CPAE 7, Doc. 13).

———. 1919. "What is the Theory of Relativity?" *The Times,* 29 November 1919. Republished in (Einstein 1954).

———. 1921. "Geometrie und Erfahrung." *Sitzungsberichte der Preussischen Akademie der Wissenschaften* Vol. 1, 123–130.

———. 1923. "Grundgedanken und Probleme der Relativitätstheorie." In *Nobelstiftelsen, Les Prix Nobel en 1921–1922*. Stockholm: Imprimerie Royale.

———. 1926. "Nichteuklidsche Geometrie in der Physik." *Neue Rundschau,* January.

———. 1933. "Notes on the Origin of the General Theory of Relativity." In *Albert Einstein. Ideas and Opinions,* 285–290. Translated by Sonja Bargmann. New York: Crown, 1954.

———. 1949. "Autobiographical Notes." In P.A. Schilpp (ed.), *Albert Einstein, Philosopher-Scientist*. Evanston, Illinois: The Library of Living Philosophers, Inc., Illinois, 29.

———. 1954. *Ideas and Opinions*. New York: Crown Publishers.

Frege, Gustav. 1891. "Über das Trägheitsgesetz." *Zeitschrift für Philosophie und philosophische Kritik* 98, 145–161.

Friedlaender, Benedict and Immanuel Friedlaender. 1896. *Absolute oder relative Bewegung?* Berlin: Leonhard Simion. (English translation in this volume.)

Hall, Alfred R. and Marie B. Hall. 1962. *Unpublished Scientific Papers of Isaac Newton*. Cambridge: Cambridge University Press.

Helmholtz, Hermann L. F. 1968. "Über die Tatsachen, die der Geometrie zum Grunde liegen." *Nachrichte von der Königlichen Gesellschaft der Wissenschaften zu Göttingen,* No. 9, June 3rd.

Hofmann, W. 1904. *Kritische Beleuchtung der beiden Grundbegriffe der Mechanik: Bewegung und Trägheit und daraus gezogene Folgerungen betreffs der Achsendrehung der Erde und des Foucault'schen Pendelversuchs*. Wien und Leipzig: M. Kuppitsch.

Kretschmann, Erich. 1917. "Über den physikalischen Sinn der Relativitätspostulate, A. Einsteins neue und seine ursprüngliche Relativitätstheorie." *Annalen der Physik* 53: 575–614.

Lagrange, Joseph-Louis. 1772. "Essai sur le problème des trois corps." Republished in: *Oeuvres de Lagrange*, Vol. 6, Paris: Gauthier-Villars, 229 (1873).

Lange, Ludwig. 1884. "Über die wissenschaftliche Fassung des Galilei'schen Beharrungsgesetz." *Philosophische Studien* 2: 266–297.

———. 1885. "Über das Beharrungsgesetz." *Berichte der Königlichen Sächsischen Gesellschaft der Wissenschaften, Math.-Physik*. Klasse 333–351.

———. 1886. *Die geschichtliche Entwicklung des Bewegungsbegriffs und ihr voraussichtliches Endergebnis. Ein Beitrag zur historischen Kritik der mechanischen Prinzipien*. Leipzig: W. Engelmann.

Laue, Max von. 1948. "Dr. Ludwig Lange. 1863–1936. (Ein zu unrecht Vergessener.)" *Die Naturwissenschaften* 35, 193.

———. 1955. *Die Relativitätstheorie,* Vol. 1. *Die Spezielle Relativitätstheorie*. Braunschweig: Vieweg.

Lorentz, Hendrik Anton. 1895. *Versuch einer Theorie der electrischen und optischen Erscheinungen in bewegten Körpern*. Leiden: Brill.

Lorentz, Hendrik Antoon et al. 1923. *The Principle of Relativity.* London: Methuen.

Mach, Ernst. 1872. *Die Geschichte und die Wurzel des Satzes von der Erhaltung der Arbeit*. Prague: Calve.

———. 1883. *Die Mechanik in ihrer Entwickelung. Historisch-kritisch dargestellt*. Leipzig: F.A. Brockhaus.

———. 1911. *History and Root of the Principle of the Conservation of Energy.* Chicago: Open Court.

———. 1960. *The Science of Mechanics. A Critical and Historical Account of Its Development*. LaSalle: Open Court.

Neumann, Carl. 1870. *Ueber die Prinzipien der Galilei-Newtonschen Theorie*. Leipzig: Teubner.

Norton, John. 1995. "Mach's Principle before Einstein." In (Barbour and Pfister 1995).

Poincaré, Henri. 1898. "La Mesure du Temps." Republished in Poincaré, Henri (1905). *La Valeur de la Science*.

———. 1902. *La Science et l'Hypothèse*. Paris: Flammarion.

———. 1905. *Science and Hypothesis*. London: Walter Scott Publ. Co.

Reissner, Hans. 1914. "Über die Relativität der Beschleunigungen in der Mechanik." *Physikalische Zeitschrift* 15: 371–75.

———. 1915. "Über eine Möglichkeit die Gravitation als unmittelbare Folge der Relativität der Trägheit abzuleiten." *Physikalische Zeitschrift* 16: 179–85.

Riemann, Bernhard. 1867. "Über die Hypothesen, welche der Geometrie zu grunde liegen." *Abhandlungen der Königlichen Gesellschaft der Wissenschaften zu Göttingen* 13.

Tait. Peter G. 1883. "Note on Reference Frames." *Proceedings of the Royal Society of Edinburgh,* Session 1883–84: 743–745.

Thomson, James. 1883. "On the law of inertia; the principle of chronometry; and the principle of absolute clinural rest, and of absolute rotation." In *Proceedings of the Royal Society of Edinburgh,* Session 1883–84, 568 and 730.

Thomson, William and Peter G. Tait. 1867. *Elements of Natural Philosophy.* Cambridge: Cambridge University Press.

ALBERT EINSTEIN

ON THE RELATIVITY PROBLEM

Originally published as "Zum Relativitätsproblem" in Scientia 15, 1914, pp. 337–348. Reprinted in "The Collected Papers of Albert Einstein," Vol. 4, Doc. 31: An English translation is given in its companion volume.

After two eminent specialists have presented their objections to relativity theory in this journal, it must be not undesirable for the readers if an adherent of this new theoretical direction expounds his view. This shall be done as concisely as possible in the following.

Currently we have to distinguish two theoretical systems, both of which fall under the name "relativity theory." The first of these, which we will call "relativity theory in the narrower sense," is based on a considerable body of experience and is accepted by the majority of theoretical physicists to be one of the simplest theoretical expressions of these experiences. The second, which we will call "relativity theory in the broader sense," is as yet by no means established on the basis of physical experience. The majority of my colleagues regard this second system either sceptically or dismissively. It should be said immediately that one can certainly be an adherent of the relativity theory in the narrow sense without admitting the validity of relativity theory in the wider sense. For that reason we will discuss the two theories separately.

I. RELATIVITY THEORY IN THE NARROWER SENSE

It is well known that the equations of the mechanics established by Galileo and Newton are not valid with respect to an arbitrarily moving coordinate system, if one adheres to the requirement that the description of motion admits only central forces satisfying the | law of equality of action and reaction. But if the motion is referred to [338]
a coordinate system K, such that Newton's equations are valid in the indicated sense, then that coordinate system is not the *only one* with respect to which those laws of mechanics hold. Rather, all of the coordinate systems K' with arbitrary spatial orientation that have uniform translational motion with respect to K have the property that relative to them the laws of motion hold. We call the assumption of the equal value of all these coordinate systems K, K', etc. for the formulation of the laws of motion, actually for the general laws of physics, the "relativity principle" (in the narrow sense).

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 3
Gravitation in the Twilight of Classical Physics: Between Mechanics, Field Theory, and Astronomy.

As long as one believed that classical mechanics lies at the foundation of the theoretical representation of all processes, one could not doubt the validity of this relativity principle. But even abstaining from that, it is difficult from an empirical standpoint to doubt the validity of that principle. In fact if it did not hold, then the processes of nature referred to a reference system at rest with respect to the Earth would appear to be influenced by the motion (velocity) of the Earth's yearly orbital motion around the Sun; the terrestrial space of observations would have to behave physically in an anisotropic manner due to the existence of this motion. But despite the most arduous searching, physicists have never observed such an apparent anisotropy.

The relativity principle is hence as old as mechanics, and no one could ever have questioned its validity from an empirical standpoint. That it has been nonetheless doubted, and is again doubted today, is due to the fact that it seemed to be incompatible with Maxwell-Lorentz electrodynamics. Whoever is in a position to judge this theory, in light of its inclusiveness, the small number of its fundamental assumptions, and its successful representation of phenomena in the domain of electrodynamics and optics, will find it difficult to dispel the impression that the main features of this theory are as definitively established as are the equations of mechanics. It has also not been accomplished to set another theory against this one that could even tolerably compete with it. |

[339] It is easy to specify wherein lies the apparent contradiction between Maxwell-Lorentz electrodynamics and the relativity principle. Suppose that the equations of that theory hold relative to the coordinate system K. This means that every light ray propagates in vacuo with a definite velocity c, with respect to K, which is independent of direction and of the state of motion of the light source; this proposition will be called the "principle of the constancy of the speed of light" in the following. Now if one such light ray were to be observed by an observer moving relative to K, then the propagation speed of this light ray, as estimated from the standpoint of this observer, in general seems to be different than c. For example, if the light ray propagates in the direction of the positive x- axis of K with speed c, and our observer moves in the same direction with the temporally constant speed v, then one would believe that one can immediately conclude that the light ray's propagation speed must be $c - v$ according to the moving observer. Relative to the observer, that is, relative to a coordinate system K' moving with the same velocity, the principle of the constancy of the speed of light does not appear to hold. Hence, here is an apparent contradiction with the principle of relativity.

However, an exact analysis of the physical content of our spatial and temporal determinations leads to the well-known result that the implied contradiction is only apparent, since it depends on both of the following arbitrary assumptions:

1. The assertion that whether two events occurring in different places occur simultaneously has content independently of the choice of a reference system.
2. The spatial distance between the places in which two simultaneous events occur is independent of the choice of a reference system.

Given that the Maxwell-Lorentz theory as well as the relativity principle are empirically supported to such a large degree, one must therefore decide to drop both the aforementioned arbitrary assumptions, the apparent evidence for which rests solely on the facts that light gives us information about distant events *apparently instantaneously*, and that the objects | we deal with in daily life have velocities that [340]
are small compared to the velocity of light c.

By abandoning these arbitrary assumptions, one achieves compatibility between the principle of the constancy of the speed of light, which results from Maxwell-Lorentz electrodynamics, and the relativity principle. One can retain the assumption that one and the same light ray propagates with velocity c relative to all reference systems K' in uniform translational motion with respect to a system K, rather than only relative to K. One only has to choose the transformation equations, which exist between the spacetime coordinates (x, y, z, t) with respect to K and those (x', y', z', t') with respect to K', in an appropriate way; one will thereby be led to the system of transformation equations called the "Lorentz transformation." This Lorentz transformation supercedes the corresponding transformations that until the development of relativity theory were regarded as the only conceivable ones, which, however, were based on the assumptions (1) and (2) given above.

The heuristic value of relativity theory consists in the fact that it provides a constraint that all of the systems of equations that express general laws of nature must satisfy. All such systems of equations must be constructed such that with the application of a Lorentz transformation they go into a system of equations of the same form (covariance with respect to the Lorentz transformations). Minkowski presented a simple mathematical schema to which equation systems must be reducible if they are to behave covariantly with respect to Lorentz transformations; thereby he achieved the advantage, that for the accommodation of the system of equations with the constraint mentioned above it is certainly not necessary to in fact carry out a Lorentz transformation on those systems.

From what has been said it clearly follows that relativity theory by no means gives us a tool for deducing previously unknown laws of nature from nothing. It only provides an always applicable criterion that constrains the possibilities; in this respect
| it is comparable to the law of energy conservation or the second law of thermody- [341]
namics.

It follows from a close examination of the most general laws of theoretical physics that Newtonian mechanics must be modified to satisfy the criterion of relativity theory. These altered mechanical equations have proven to be applicable to cathode rays and β-rays (motion of free electrical particles). Moreover, the implementation of relativity theory has lead to neither a logical contradiction nor a conflict with empirical results.

Only one result of relativity theory will be given here in particular, because it is of importance for the following analyses. According to Newtonian mechanics the inertia of a system constituted by a collection of material points (that is, the inertial resistance against acceleration of the system's center of gravity) is independent of the state of the system. By contrast, according to relativity theory the inertia of an isolated system (floating in a vacuum) depends upon the state of the system, such that the inertia increases with the energy content of the system. Thus according to relativity theory, it is ultimately *energy* that inertia can be attributed to. The energy, rather than the inertial mass of a material point, is what we have ascribed indestructibility to; hence the theorem of the conservation of mass is incorporated into the theorem of the conservation of energy.

It was remarked above that it would be a great mistake to regard relativity theory as a universal method that allows one to develop an unequivocally appropriate theory for a domain of phenomena regardless of how little this has been explored empirically. Relativity theory only *reduces* by a significant amount the empirical conclusions necessary for the development of a theory. There is only one domain of fundamental importance where we have such poor empirical knowledge that this knowledge, in combination with relativity theory, is not sufficient, by a wide margin, for a clear determination of the general theory. This is the domain of gravitational phenomena. Here we can only reach our goal by complementing what is empirically
[342] known with physical hypotheses in order to complete the basis | of the theory. The following considerations are firstly to show how one arrives at what are, in my opinion, the most natural such hypotheses.

When we speak of a body's *mass*, we associate with this word two definitions that are logically completely independent. By *mass* we understand, first, a constant inherent to a body, which measures its resistance to acceleration ("inertial mass"), and second, the constant of a body that determines the magnitude of the force that it experiences in a gravitational field ("gravitational mass").

It is in no way self-evident *a priori* that the inertial and gravitational mass of a body must agree; we are simply *accustomed* to assume their agreement. The belief in this agreement stems from the empirical fact that the acceleration that various bodies experience in a gravitational field is independent of their material constitution. Eötvös has shown that, in any case, inertial and gravitational mass agree with very great precision, in that, through his experiments with a torsion balance, he ruled out the existence a relative deviation of the two masses from each other on the order of magnitude of 10^{-8}. [1]

Enormous quantities of energy in the form of heat are discharged into the environment by radioactive processes. According to the result regarding the inertia of energy presented above, the decay products generated by the reaction taken together must have a smaller inertial mass than that of the material existing before the radioactive decay. This change of inertial mass is, for the kind of reaction with known heat effect,

1 Eötvös' experimental method is based on the following. The Earth's gravity and the centrifugal force influence a body found on the Earth's surface. For the first the body's gravitational mass, and for the second the inertial mass, is the determining factor. If the two were not identical, then the direction of the resultant of the two (the apparent weight) would depend on the material of the body. With his torsion balance experiments, Eötvös proved with great precision the non-existence of such a dependence.

of the order 10^{-4}. If the gravitational mass did not change simultaneously with the inertial mass of the system, then the inertial mass would have to differ from the gravitational mass for various elements far more than Eötvös's | experiments would allow. [343] Langevin was the first to call attention to this important point.

From what has been said the identity of the inertial and gravitational mass of closed systems (at rest) follows with great probability; I think that based on the present state of empirical knowledge we should adhere to the assumption of this identity unconditionally. We have thereby attained one of the most important physical demands that, from my point of view, must be imposed on any gravitational theory.

This demand involves a far-reaching constraint on gravitational theories, which one recognizes especially in conjunction with the theorem of the inertia of energy. All energy corresponds to inertial mass, and all inertial mass corresponds to gravitational mass; the gravitational mass of a closed system must therefore be determined by its energy. The energy of its gravitational field also belongs to the energy of a closed system; hence the gravitational field energy itself contributes to the system's gravitational mass rather than only its inertial mass.

Abraham and Mie have proposed gravitational theories. Abraham's theory contradicts the relativity principle, and Mie's theory contradicts the demand of the equality of the inertial and gravitational mass of a closed system. According to the latter theory, were a body to be heated the *inertial* mass *grows* in proportion to the energy gain, but not the gravitational mass; the latter would actually decrease for a gas with rising temperature.[2]

By way of contrast, a gravitational theory recently proposed by Nordström complies with both the relativity principle and the requirement of the gravity of energy of closed systems, with one restriction to be indicated in the following. Abraham's claim to the contrary made in a paper appearing in this journal is not | correct. In fact, I [344] believe that a cogent argument against Nordström's theory cannot be drawn from experience.

According to Nordström's theory the principle of the gravity of energy of closed systems at rest holds as a statistical principle. The gravitational mass of a closed system (with the whole system at rest) is in general an oscillating quantity, whose temporal average is determined by the total energy of the system. As a consequence of the oscillatory character of mass, such a system must emit standing longitudinal gravitational waves. Yet the energy loss expected according to the theory is so small that it must escape our notice.

After a more detailed study of Nordström's theory, everyone will have to admit that this theory, when regarded from an empirical standpoint, is an unobjectionable integration of gravitation into the framework of relativity theory (in the narrower

2 Due to their smallness, these effects are certainly not accessible to experiments. But it seems to me that there is much to be said for taking the connection between inertial and gravitational mass to be warranted in principle, regardless of what forms of energy are taken into account. According to Mie, one can account for the fact that the equality of inertial and gravitational mass holds for radioactive transformations only with assumptions regarding the special nature of energy in the interior of atoms.

sense). Even though I am of the opinion that we cannot be satisfied with this solution, my reasons for this have an epistemological character that will be described in the following.

II. RELATIVITY THEORY IN THE BROADER SENSE

Classical mechanics, as well as relativity theory in the narrow sense briefly described above, suffer from a fundamental defect, which no one can deny, that is accessible to epistemological arguments. The weaknesses of our physical world picture to be discussed below were already uncovered with full clarity by E. Mach in his deeply penetrating investigations of the foundations of Newtonian mechanics, so that what I will assert in this respect can have no claim to novelty. I will explain the essential point with an example, which is chosen to be quite elementary, in order to allow what is essential to stand out.

Two masses float in space at a great distance from all celestial bodies. Suppose that these two are close enough together that they can exert an influence on each other. Now an observer watches the motion of both bodies, such that he continually
[345] looks along the direction of the line | connecting the two masses toward the sphere of fixed stars. He will observe that the line of sight traces out a closed line on the visible sphere of fixed stars, which does not change its position with respect to the visible sphere of fixed stars. If the observer has any natural intelligence, but has learned neither geometry nor mechanics, he would conclude: “My masses carry out a motion, which is at least in part causally determined by the fixed stars. The law by which masses in my surroundings move is co-determined by the fixed stars.” A man who has been schooled in the sciences would smile at the simple-mindedness of our observer and say to him: “The motion of your masses has nothing to do with the heaven of fixed stars; it is rather fully determined by the laws of mechanics entirely independently of the remaining masses. There is a space R in which these laws hold. These laws are such that the masses remain continually in a plane in this space. However, the system of fixed stars cannot rotate in this space, because otherwise it would be disrupted by powerful centrifugal forces. Thus it necessarily must be at rest (at least almost!), if it is to exist permanently; this is the reason that the plane in which your masses move always goes through the same fixed stars.” But our intrepid observer would say: “You may be incomparably learned. But just as I could never be brought to believe in ghosts, so I cannot believe in this gigantic thing that you speak of and call space. I can neither see something like that nor conceive of it. Or should I think of your space R as a subtle net of bodies that the remaining things are all referred to? Then I can imagine a second such net R' in addition to R, that is moving in an arbitrary manner relative to R (for example, rotating). Do your equations also hold at the same time with respect to R'?” The learned man denies this with certainty. In reply to which the ignoramus: “But how do the masses know which “space” R, R', etc. with respect to which they should move according to your equations, how do they recognize the space or spaces they orient themselves with respect to?” Now our

learned man is in a quite embarrassing position. He certainly insists that there must be such privileged spaces, but he knows no reason to give for why such spaces could be distinguished from | other spaces. The ignoramus's reply: "Then I will take, for the [346] time being, your privileged spaces as an idle fabrication, and stay with my conception, that the sphere of fixed stars co-determines the mechanical behavior of my test masses."

I will explain the violation of the most elementary postulate of epistemology of which our physics is guilty in yet another way. One would try in vain to explain what one understands by the simple acceleration of a body. One would only succeed in defining *relative* acceleration of bodies with respect to each other. However, having said that, we base our mechanics on the premise that a force (cause) is necessary for the generation of a body's acceleration, ignoring the fact that we cannot explain what it is that we understand by "acceleration," exactly because only *relative* accelerations can be an object of perception.

The dubious aspect of proceeding in this vein is very nicely illustrated by a comparison, which I owe to my friend Besso. Suppose we think back to an earlier time, when it was assumed that the surface of the Earth must be approximately *flat*. Imagine that the following conception exists among the learned. In the world there is a physically distinguished direction, the vertical. All objects fall in this direction if they are not supported. Because of this the surface of the Earth is essentially perpendicular to this direction, and this is why it tends towards the form of a plane. While in this case, the mistake lies in privileging one direction over all others without good reason (fictitious cause), rather than simply regarding the Earth as the cause of falling, the mistake in our physics lies in the introduction without good reason of privileged reference systems as fictitious causes; both cases are characterized by forgoing the establishment of a sufficient reason.

Since relativity theory in the narrower sense, rather than only classical mechanics, exhibits the fundamental deficiency explained above, I set myself the goal of generalizing relativity theory in such a way that this imperfection will be avoided. First of all, I recognized that gravitation in general | must be assigned a fundamental role in [347] any such theory. Then from what was explained earlier it already follows that every physical process must also produce a gravitational field, because of the quantity of energy corresponding to it. On the other hand, the empirical fact that all bodies fall with the same speed in a gravitational field suggests the idea that physical processes happen in a gravitational field exactly as they do relative to an accelerated reference system (equivalence hypothesis). In taking this idea as a foundation, I came to the conclusion that the velocity of light is not to be regarded as independent of the gravitational potential. Thus the principle of the constancy of the speed of light is incompatible with the equivalence hypothesis; that is why relativity theory in the narrower sense cannot be made consistent with the equivalence hypothesis. This led me to take relativity theory in the narrow sense to be applicable only in regions within which no noticeable differences in the gravitational potential occur. Relativity theory (in the narrower sense) has to be replaced by a general theory, which contains the former as a limiting case.

The path leading to this theory can only be very incompletely described in words.[3] The equations of motion for material points in a gravitational field that follow from the equivalence hypothesis can be easily written in a form such that these laws are completely independent of the choice of variables determining place and time. By leaving the choice of these variables as *a priori* completely arbitrary, and thus not privileging any spacetime system, one averts the epistemological objection discussed above. A quantity

$$ds^2 = \sum_{\mu\nu} g_{\mu\nu} dx_\mu dx_\nu,$$

appears in that law of motion, and it is invariant, i.e., it is a quantity that is independent of the choice of reference system (i.e., of the choice of four spacetime coordinates). The quantities $g_{\mu\nu}$ are functions of $x_1 \ldots x_4$ and represent the gravitational
[348] field. |

With the help of the absolute differential calculus, which has been developed by Ricci and Levi-Civita based on Christoffel's mathematical investigations, one can succeed, based on the existence of the invariant above, in replacing the well-known systems of equations of physics with equivalent systems (when all $g_{\mu\nu}$ are constant), which are valid independent of a choice of the spacetime coordinate system x_ν. All such systems of equations include the quantities $g_{\mu\nu}$, i.e., the quantities that determine the gravitational field. Thus the latter influence all physical processes.

Conversely, physical processes must also determine the gravitational field, i.e., the quantities $g_{\mu\nu}$. One arrives at the differential equations that determine these quantities by means of the hypothesis that the conservation of momentum and energy must hold for material events and the gravitational field taken together. This hypothesis subsequently constrains the choice of spacetime variables x, without thereby evoking again the epistemological doubts analyzed above. Because according to this generalized relativity theory there are no longer privileged spaces with peculiar physical qualities. The quantities $g_{\mu\nu}$ control the course of all processes, which for their part are determined by the physical events in all the rest of the universe.

The principle of the inertia and gravitation of energy is completely satisfied in this theory. Furthermore, the equations of motion for gravitational masses are such that it is, as one must demand based on the considerations above, acceleration with respect to other bodies rather than absolute acceleration (acceleration with respect to "space") that appears as that which is decisive for the appearance of inertial resistance.

Relativity theory in the broader sense signifies a further development of the earlier relativity theory, rather than an abandonment of it, that seems necessary to me for the epistemological reasons I cited.

---

3 Cf. A. Einstein and M. Grossmann, *Zeitschrift f. Math. & Physik* 62 (1914): p. 225.

ALBERT EINSTEIN

ETHER AND THE THEORY OF RELATIVITY

Originally published in 1920 as "Äther und Relativitätstheorie" at Springer, Berlin, on the basis of an address delivered on May 5th, 1920, at the University of Leyden. The English version reproduced here first appeared in 1922 in Albert Einstein: "Sidelights of Relativity" at Methuen, London, pp. 3–24. The page numbers given here refer to the latter edition.

How does it come about that alongside of the idea of ponderable matter, which is [3]
derived by abstraction from everyday life, the physicists set the idea of the existence of another kind of matter, the ether? The explanation is probably to be sought in those phenomena which have given rise to the theory of action at a distance, and in the properties of light which have led to the undulatory theory. Let us devote a little while to the consideration of these two subjects.

Outside of physics we know nothing of action at a distance. When we try to connect cause and effect in the experiences which natural objects afford us, it seems at
first as if there were no other | mutual actions than those of immediate contact, e.g. [4]
the communication of motion by impact, push and pull, heating or inducing combustion by means of a flame, etc. It is true that even in everyday experience weight, which is in a sense action at a distance, plays a very important part. But since in daily experience the weight of bodies meets us as something constant, something not linked to any cause which is variable in time or place, we do not in everyday life speculate as to the cause of gravity, and therefore do not become conscious of its character as action at a distance. It was Newton's theory of gravitation that first assigned a cause for gravity by interpreting it as action at a distance, proceeding from masses. Newton's theory is probably the greatest stride ever made in the effort towards the causal nexus of natural phenomena. And yet this theory evoked a lively sense of discomfort among Newton's contemporaries, because it seemed to be in conflict with the principle springing from the rest of experience, that there can be recip-
rocal | action only through contact, and not through immediate action at a distance. [5]

It is only with reluctance that man's desire for knowledge endures a dualism of this kind. How was unity to be preserved in his comprehension of the forces of nature? Either by trying to look upon contact forces as being themselves distant forces which admittedly are observable only at a very small distance—and this was the road which Newton's followers, who were entirely under the spell of his doctrine,

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 3
Gravitation in the Twilight of Classical Physics: Between Mechanics, Field Theory, and Astronomy.

mostly preferred to take; or by assuming that the Newtonian action at a distance is only apparently immediate action at a distance, but in truth is conveyed by a medium permeating space, whether by movements or by elastic deformation of this medium. Thus the endeavour toward a unified view of the nature of forces leads to the hypothesis of an ether. This hypothesis, to be sure, did not at first bring with it any advance in the theory of gravitation or in physics generally, so that it became customary to
[6] treat Newton's law of force as an axiom | not further reducible. But the ether hypothesis was bound always to play some part in physical science, even if at first only a latent part.

When in the first half of the nineteenth century the far-reaching similarity was revealed which subsists between the properties of light and those of elastic waves in ponderable bodies, the ether hypothesis found fresh support. It appeared beyond question that light must be interpreted as a vibratory process in an elastic, inert medium filling up universal space. It also seemed to be a necessary consequence of the fact that light is capable of polarisation that this medium, the ether, must be of the nature of a solid body, because transverse waves are not possible in a fluid, but only in a solid. Thus the physicists were bound to arrive at the theory of the "quasi-rigid" luminiferous ether, the parts of which can carry out no movements relatively to one another except the small movements of deformation which correspond to light-waves.

[7] This theory—also called the theory of | the stationary luminiferous ether—moreover found a strong support in an experiment which is also of fundamental importance in the special theory of relativity, the experiment of Fizeau, from which one was obliged to infer that the luminiferous ether does not take part in the movements of bodies. The phenomenon of aberration also favoured the theory of the quasi-rigid ether.

The development of the theory of electricity along the path opened up by Maxwell and Lorentz gave the development of our ideas concerning the ether quite a peculiar and unexpected turn. For Maxwell himself the ether indeed still had properties which were purely mechanical, although of a much more complicated kind than the mechanical properties of tangible solid bodies. But neither Maxwell nor his followers succeeded in elaborating a mechanical model for the ether which might furnish a satisfactory mechanical interpretation of Maxwell's laws of the electromagnetic field. The laws were clear and simple, the mechanical
[8] interpretations | clumsy and contradictory. Almost imperceptibly the theoretical physicists adapted themselves to a situation which, from the standpoint of their mechanical programme, was very depressing. They were particularly influenced by the electrodynamical investigations of Heinrich Hertz. For whereas they previously had required of a conclusive theory that it should content itself with the fundamental concepts which belong exclusively to mechanics (e.g. densities, velocities, deformations, stresses) they gradually accustomed themselves to admitting electric and magnetic force as fundamental concepts side by side with those of mechanics, without requiring a mechanical interpretation for them. Thus the purely mechanical view of nature was gradually abandoned. But this change led to a fundamental dualism which in the

long-run was insupportable. A way of escape was now sought in the reverse direction,
by reducing the principles of mechanics to those of electricity, and this especially as
confidence in the strict validity of the equations of Newton's | mechanics was shaken [9]
by the experiments with β- rays and rapid cathode rays.

This dualism still confronts us in unextenuated form in the theory of Hertz, where matter appears not only as the bearer of velocities, kinetic energy, and mechanical pressures, but also as the bearer of electromagnetic fields. Since such fields also occur in vacuo—i.e. in free ether—the ether also appears as bearer of electromagnetic fields. The ether appears indistinguishable in its functions from ordinary matter. Within matter it takes part in the motion of matter and in empty space it has everywhere a velocity; so that the ether has a definitely assigned velocity throughout the whole of space. There is no fundamental difference between Hertz's ether and ponderable matter (which in part subsists in the ether).

The Hertz theory suffered not only from the defect of ascribing to matter and
ether, on the one hand mechanical states, and on the other hand electrical states,
which do not stand in any conceivable | relation to each other; it was also at variance [10]
with the result of Fizeau's important experiment on the velocity of the propagation of
light in moving fluids, and with other established experimental results.

Such was the state of things when H. A. Lorentz entered upon the scene. He brought theory into harmony with experience by means of a wonderful simplification of theoretical principles. He achieved this, the most important advance in the theory of electricity since Maxwell, by taking from ether its mechanical, and from matter its electromagnetic qualities. As in empty space, so too in the interior of material bodies, the ether, and not matter viewed atomistically, was exclusively the seat of electromagnetic fields. According to Lorentz the elementary particles of matter alone are capable of carrying out movements; their electromagnetic activity is entirely confined to the carrying of electric charges. Thus Lorentz succeeded in reducing all electromagnetic happenings to Maxwell's equations for free space.

As to the mechanical nature of the | Lorentzian ether, it may be said of it, in a [11]
somewhat playful spirit, that immobility is the only mechanical property of which it
has not been deprived by H. A. Lorentz. It may be added that the whole change in the
conception of the ether which the special theory of relativity brought about, consisted
in taking away from the ether its last mechanical quality, namely, its immobility. How
this is to be understood will forthwith be expounded.

The spacetime theory and the kinematics of the special theory of relativity were
modelled on the Maxwell-Lorentz theory of the electromagnetic field. This theory
therefore satisfies the conditions of the special theory of relativity, but when viewed
from the latter it acquires a novel aspect. For if K be a system of co-ordinates relatively to which the Lorentzian ether is at rest, the Maxwell-Lorentz equations are
valid primarily with reference to K. But by the special theory of relativity the same
equations without any change of meaning also hold in relation to any new system of
co-ordinates | K' which is moving in uniform translation relatively to K. Now comes [12]
the anxious question:—Why must I in the theory distinguish the K system above all

K' systems, which are physically equivalent to it in all respects, by assuming that the ether is at rest relatively to the K system? For the theoretician such an asymmetry in the theoretical structure, with no corresponding asymmetry in the system of experience, is intolerable. If we assume the ether to be at rest relatively to K, but in motion relatively to K', the physical equivalence of K and K' seems to me from the logical standpoint, not indeed downright incorrect, but nevertheless inacceptable.

The next position which it was possible to take up in face of this state of things appeared to be the following. The ether does not exist at all. The electromagnetic fields are not states of a medium, and are not bound down to any bearer, but they are independent realities which are not reducible to anything else, exactly like the atoms
[13] of ponderable matter. This | conception suggests itself the more readily as, according to Lorentz's theory, electromagnetic radiation, like ponderable matter, brings impulse and energy with it, and as, according to the special theory of relativity, both matter and radiation are but special forms of distributed energy, ponderable mass losing its isolation and appearing as a special form of energy.

More careful reflection teaches us, however, that the special theory of relativity does not compel us to deny ether. We may assume the existence of an ether; only we must give up ascribing a definite state of motion to it, i.e. we must by abstraction take from it the last mechanical characteristic which Lorentz had still left it. We shall see later that this point of view, the conceivability of which I shall at once endeavour to make more intelligible by a somewhat halting comparison, is justified by the results of the general theory of relativity.

Think of waves on the surface of water. Here we can describe two entirely differ-
[14] ent things. Either we may observe how | the undulatory surface forming the boundary between water and air alters in the course of time; or else—with the help of small floats, for instance—we can observe how the position of the separate particles of water alters in the course of time. If the existence of such floats for tracking the motion of the particles of a fluid were a fundamental impossibility in physics—if, in fact, nothing else whatever were observable than the shape of the space occupied by the water as it varies in time, we should have no ground for the assumption that water consists of movable particles. But all the same we could characterize it as a medium.

We have something like this in the electromagnetic field. For we may picture the field to ourselves as consisting of lines of force. If we wish to interpret these lines of force to ourselves as something material in the ordinary sense, we are tempted to interpret the dynamic processes as motions of these lines of force, such that each sep-
[15] arate line of | force is tracked through the course of time. It is well known, however, that this way of regarding the electromagnetic field leads to contradictions.

Generalizing we must say this:—There may be supposed to be extended physical objects to which the idea of motion cannot be applied. They may not be thought of as consisting of particles which allow themselves to be separately tracked through time. In Minkowski's idiom this is expressed as follows:—Not every extended conformation in the four-dimensional world can be regarded as composed of worldthreads. The special theory of relativity forbids us to assume the ether to consist of particles

observable through time, but the hypothesis of ether in itself is not in conflict with the
special theory of relativity. Only we must be on our guard against ascribing a state of
motion to the ether.

Certainly, from the standpoint of the special theory of relativity, the ether hypoth-
esis appears at first to be an empty hypothesis. In the equations of the | electromag- [16]
netic field there occur, in addition to the densities of the electric charge, only the
intensities of the field. The career of electromagnetic processes in vacuo appears to be
completely determined by these equations, uninfluenced by other physical quantities.
The electromagnetic fields appear as ultimate, irreducible realities, and at first it
seems superfluous to postulate a homogeneous, isotropic ether-medium, and to envis-
age electromagnetic fields as states of this medium.

But on the other hand there is a weighty argument to be adduced in favour of the
ether hypothesis. To deny the ether is ultimately to assume that empty space has no
physical qualities whatever. The fundamental facts of mechanics do not harmonize
with this view. For the mechanical behavior of a corporeal system hovering freely in
empty space depends not only on relative positions (distances) and relative velocities,
but also on its state of rotation, which physically may be taken as a characteristic not
appertaining to the system in itself. In order | to be able to look upon the rotation of [17]
the system, at least formally, as something real, Newton objectivizes space. Since he
classes his absolute space together with real things, for him rotation relative to an
absolute space is also something real. Newton might no less well have called his
absolute space "ether"; what is essential is merely that besides observable objects,
another thing, which is not perceptible, must be looked upon as real, to enable accel-
eration or rotation to be looked upon as something real.

It is true that Mach tried to avoid having to accept as real something which is not
observable by endeavouring to substitute in mechanics a mean acceleration with ref-
erence to the totality of the masses in the universe in place of an acceleration with ref-
erence to absolute space. But inertial resistance opposed to relative acceleration of
distant masses presupposes action at a distance; and as the modern physicist does not
believe that he may accept this action at a distance, he comes back once more, if he |
follows Mach, to the ether, which has to serve as medium for the effects of inertia. [18]
But this conception of the ether to which we are led by Mach's way of thinking dif-
fers essentially from the ether as conceived by Newton, by Fresnel, and by Lorentz.
Mach's ether not only *conditions* the behavior of inert masses, but *is also conditioned*
in its state by them.

Mach's idea finds its full development in the ether of the general theory of relativ-
ity. According to this theory the metrical qualities of the continuum of spacetime dif-
fer in the environment of different points of spacetime, and are partly conditioned by
the matter existing outside of the territory under consideration. This spacetime vari-
ability of the reciprocal relations of the standards of space and time, or, perhaps, the
recognition of the fact that empty space in its physical relation is neither homoge-
neous nor isotropic, compelling us to describe its state by ten functions (the gravita-
tion potentials $g_{\mu\nu}$), has, I think, finally disposed of the view that | space is [19]

physically empty. But therewith the conception of the ether has again acquired an intelligible content, although this content differs widely from that of the ether of the mechanical undulatory theory of light. The ether of the general theory of relativity is a medium which is itself devoid of *all* mechanical and kinematical qualities, but helps to determine mechanical (and electromagnetic) events.

What is fundamentally new in the ether of the general theory of relativity as opposed to the ether of Lorentz consists in this, that the state of the former is at every place determined by connections with the matter and the state of the ether in neighboring places, which are amenable to law in the form of differential equations; whereas the state of the Lorentzian ether in the absence of electromagnetic fields is conditioned by nothing outside itself, and is everywhere the same. The ether of the general theory of relativity is transmuted conceptually into the ether of Lorentz if we
[20] substitute constants for the functions of space which describe the | former, disregarding the causes which condition its state. Thus we may also say, I think, that the ether of the general theory of relativity is the outcome of the Lorentzian ether, through relativation.

As to the part which the new ether is to play in the physics of the future we are not yet clear. We know that it determines the metrical relations in the spacetime continuum, e.g. the configurative possibilities of solid bodies as well as the gravitational fields; but we do not know whether it has an essential share in the structure of the electrical elementary particles constituting matter. Nor do we know whether it is only in the proximity of ponderable masses that its structure differs essentially from that of the Lorentzian ether; whether the geometry of spaces of cosmic extent is approximately Euclidean. But we can assert by reason of the relativistic equations of gravitation that there must be a departure from Euclidean relations, with spaces of cosmic order of magnitude, if there exists a positive mean density, no matter how small, of
[21] the matter in the universe. | In this case the universe must of necessity be spatially unbounded and of finite magnitude, its magnitude being determined by the value of that mean density.

If we consider the gravitational field and the electromagnetic field from the standpoint of the ether hypothesis, we find a remarkable difference between the two. There can be no space nor any part of space without gravitational potentials; for these confer upon space its metrical qualities, without which it cannot be imagined at all. The existence of the gravitational field is inseparably bound up with the existence of space. On the other hand a part of space may very well be imagined without an electromagnetic field; thus in contrast with the gravitational field, the electromagnetic field seems to be only secondarily linked to the ether, the formal nature of the electromagnetic field being as yet in no way determined by that of gravitational ether. From the present state of theory it looks as if the electromagnetic field, as opposed to the
[22] gravitational field, rests upon an entirely new formal *motif*, as though | nature might just as well have endowed the gravitational ether with fields of quite another type, for example, with fields of a scalar potential, instead of fields of the electromagnetic type.

Since according to our present conceptions the elementary particles of matter are also, in their essence, nothing else than condensations of the electromagnetic field, our present view of the universe presents two realities which are completely separated from each other conceptually, although connected causally, namely, gravitational ether and electromagnetic field, or—as they might also be called—space and matter.

Of course it would be a great advance if we could succeed in comprehending the gravitational field and the electromagnetic field together as one unified conformation. Then for the first time the epoch of theoretical physics founded by Faraday and Maxwell would reach a satisfactory conclusion. The contrast between ether and matter would fade away, and, through the general theory of relativity, the whole of | physics [23]
would become a complete system of thought, like geometry, kinematics, and the theory of gravitation. An exceedingly ingenious attempt in this direction has been made by the mathematician H. Weyl; but I do not believe that his theory will hold its ground in relation to reality. Further, in contemplating the immediate future of theoretical physics we ought not unconditionally to reject the possibility that the facts comprised in the quantum theory may set bounds to the field theory beyond which it cannot pass.

Recapitulating, we may say that according to the general theory of relativity space is endowed with physical qualities; in this sense, therefore, there exists an ether. According to the general theory of relativity space without ether is unthinkable; for in such space there not only would be no propagation of light, but also no possibility of existence for standards of space and time (measuring-rods and clocks), nor therefore any spacetime intervals in the physical sense. But this ether may not be thought of as endowed with the quality characteristic | of ponderable media, as consisting of parts [24]
which may be tracked through time. The idea of motion may not be applied to it.

FROM AN ELECTROMAGNETIC THEORY OF MATTER TO A NEW THEORY OF GRAVITATION

CHRISTOPHER SMEENK AND CHRISTOPHER MARTIN

MIE'S THEORIES OF MATTER AND GRAVITATION

Unifying physics by describing a variety of interactions—or even all interactions—within a common framework has long been an alluring goal for physicists. One of the most ambitious attempts at unification was made in the 1910s by Gustav Mie. Mie aimed to derive electromagnetism, gravitation, and aspects of the emerging quantum theory from a single variational principle and a well-chosen world function (Hamiltonian). Mie's main innovation was to consider nonlinear field equations to allow for stable particle-like solutions (now called solitons); furthermore he clarified the use of variational principles in the context of special relativity. The following brief introduction to Mie's work has three main objectives.[1] The first is to explain how Mie's project fit into the contemporary development of the electromagnetic worldview. Part of Mie's project was to develop a relativistic theory of gravitation as a consequence of his generalized electromagnetic theory, and our second goal is to briefly assess this work, which reflects the conceptual resources available for developing a new account of gravitation by analogy with electromagnetism. Finally, Mie was a vocal critic of other approaches to the problem of gravitation. Mie's criticisms of Einstein, in particular, bring out the subtlety and novelty of the ideas that Einstein used to guide his development of general relativity.

In September 1913 Einstein presented a lecture on the current status of the problem of gravitation at the 85th *Naturforscherversammlung* in Vienna. Einstein's lecture and the ensuing heated discussion, both published later that year in the *Physikalische Zeitschrift*, reflect the options available for those who took on the task of developing a new theory of gravitation. The conflict between Newtonian gravitational theory and special relativity provided a strong motivation for developing a new gravitational theory, but it was not clear whether a fairly straightforward modification of Newton's theory based on classical field theory would lead to a successful replacement. Einstein clearly aimed to convince his audience that success would require the more radical step of extending the principle of relativity. For Einstein the development of a new gravitational theory was intricately connected with foundational prob-

1 There are several recent, more comprehensive discussions of Mie's work, which we draw on here: (Kohl 2002; Vizgin 1994; 26–38; Corry 1999, 2004, chaps. 6 and 7). Born (1914) gives an insightful, influential reformulation of Mie's framework, and (Pauli 1921, §64, 188–192 in the English translation) and (Weyl 1918, §25, 206–217 (§26) in the English translation of the fourth edition) both give clear contemporary reviews.

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 4
Gravitation in the Twilight of Classical Physics: The Promise of Mathematics.

lems in classical mechanics, and in the Vienna lecture he motivated the need to extend the principle of relativity with an appeal to Mach's analysis of inertia. According to Einstein Mach had accurately identified an "epistemological defect" in classical mechanics, namely the introduction of a distinction between inertial and non-inertial reference frames without an appropriate observational basis.[2] The special theory of relativity had replaced Galilean transformations between reference frames with Lorentz transformations, but the principle of relativity still did not apply to accelerated motion. Extending the principle of relativity to accelerated motion depended on an idea Einstein later called "the most fortunate thought of my life," the principle of equivalence. This idea received many different formulations over the years, but in 1913 Einstein gave one version of this principle as a postulate: his second postulate requires the exact equality of inertial and gravitational mass. He further argued that this equality undermines the ability to observationally distinguish between a state of uniform acceleration and the presence of a gravitational field. The principle of equivalence gave Einstein a valuable link between acceleration and gravitation, tying together the problem of gravitation and the problem of extending the principle of relativity. At the time of the Vienna lecture Einstein was in the midst of an ongoing struggle to clarify the connections among Mach's insight, a generalized principle of relativity, and the formal requirement of general covariance, a struggle that would continue for several more years. Although he also drew heavily on classical field theory in his work, he was convinced that this cluster of ideas would provide the key to a new theory of gravitation.

Gustav Mie's approach to the problem of gravitation stands in sharp contrast to Einstein's. In the discussion following the Vienna lecture, Mie pointedly criticized Einstein's requirement of general covariance and complained that Einstein had overlooked other approaches to gravitation, including his own work and that of Max Abraham.[3] Mie commented that Einstein might have missed his theory of gravitation since it was "tucked away in a work on a comprehensive theory of matter" (CPAE 5, Doc. 18, 1262). This remark aptly characterizes where Mie placed the problem of gravitation conceptually; in Mie's approach the problem of gravitation would be solved as a by-product of an extension of classical field theory. The problem of gravitation was one of the issues, among many, that a "comprehensive theory of matter" would resolve. The pressing issue for Mie was to develop a unified field theory that would succeed where earlier attempts at a reduction of mechanics to electromagnetic theory had failed. By way of contrast with Einstein, Mie's project did not lead out of special

---

2 Einstein discusses these issues in §4 and §9 of the Vienna lecture (Einstein 1913), as well as in part II of (Einstein 1914), both included in this volume. For a thorough discussion of the role of Machian ideas in Einstein's discovery of general relativity, see "The Third Way to General Relativity" (in vol. 3 of this series).

3 In the published version of the lecture Einstein does briefly mention Abraham's theory only to remark that it fails to satisfy his third postulate, namely the requirement of Lorentz covariance. Mie later noted (Mie 1914, note 13, 175) that the reference to Abraham was only added in the published version of the lecture.

relativity, and Mie was not convinced by Einstein's attempt to link issues in the foundations of mechanics to the problem of gravitation. In Vienna, Einstein justified his sin of omission by pointing out that Mie's theory violated one of his starting assumptions, namely the principle of equivalence. But this clearly did not sway Mie, who expressed doubts that the principle could serve as the basis for a theory and whether it even held in Einstein's own *Entwurf* theory.[4] Mie was also a forceful critic of Einstein's search for a generalized principle of relativity. In the discussion following the Vienna talk and in subsequent articles (Mie 1914, 1915), Mie argued that Einstein had failed to establish a clear link between a principle of general relativity and accelerated motion and questioned the physical content of the principle. Mie had put his finger on the ambiguity of Einstein's guiding principles and the slippage between these ideas and the formal requirement of general covariance. More generally, Mie's criticisms illustrate that Einstein's idiosyncratic path to developing a new gravitational theory seemed to lead into the wilderness in 1913, and that Einstein had not provided entirely convincing reasons to abandon a more conservative path toward a new theory.

Mie's comprehensive theory of matter was presented in a series of three ambitious papers in 1912–13. Mie was eleven years older than Einstein and had held a position as a theoretical physicist in Greifswald since 1902. He was well known for work in applied optics and electromagnetism, including an insightful treatment of the scattering of electromagnetic radiation by spherical particles (Mie 1908) and a widely used textbook (Mie 1910). Mie's textbook endorsed the electromagnetic worldview prominently advocated in the previous decade by Wilhelm Wien and Max Abraham. This worldview amounted to the claim that electromagnetic theory had replaced mechanics as the foundation of physical theory, and Mie characterized electromagnetic theory as "aether physics." Mie emphasized the appeal of reducing physics to a simple set of equations governing the state of the aether and its dynamical evolution, and conceiving of elementary particles as stable "knots" in the aether rather than independent entities (Mie 1912a, 512–13). The aim of the trilogy on matter theory was to develop a unified theory able to account for the existence and properties of electrons (as well as atoms or molecules), explain recent observations of atomic spectra, and yield field equations for gravitation. Although Mie ultimately failed to achieve these grand goals, the approach and formalism he developed influenced later work in unified field theory.

Mie's program differed in important ways from electron theories from the previous decade.[5] The main obstacle to earlier attempts to realize the electromagnetic worldview was the difficulty of explaining the nature and structure of the electron itself in purely electromagnetic terms. Electron theory was an active research area in the first decade of the twentieth century, drawing the attention of many of the best

4 In the discussion, Mie announced that he would soon publish a proof that equality does not hold in the *Entwurf* theory, which appeared in §3 of (Mie 1914).

5 Here we draw primarily on the insightful analysis of the transition from electron theory to relativistic electrodynamics in (Janssen and Mecklenburg 2005); see also the essays collected in (Buchwald and Warwick 2001)

physicists of that generation, such as Lorentz, Abraham, and Sommerfeld. By the time of Mie's work the aim of determining the internal structure of the electron, treated as an extended particle with a definite shape and charge distribution, had been largely abandoned and interest in electron theory had begun to wane. With the advent of special relativity came the realization that the velocity dependence of the electron's mass, a quantity that had been touted as a sensitive experimental test of the internal structure of the electron, was instead a direct consequence of the principle of relativity (Pauli 1921, 185; Pais 1972).[6] Developments in electron theory also threatened the goal of replacing Newtonian mechanics with electromagnetism. Poincaré (1906) proved that an electron treated as a distribution of charge over a spherical shell is not a stable configuration if only the electromagnetic forces are included—the repulsive Coulomb forces would cause it to break apart. Thus it was necessary to introduce the so-called "Poincaré stress," an attractive force needed to maintain the stability of the electron.

One way of responding to these results was to temper the reductive ambitions of the electromagnetic worldview, and to follow Lorentz in admitting charged particles or non-electromagnetic forces as basic elements of the theory. Mie took a different route, and chose instead to alter the field equations of electromagnetism so that there are solutions corresponding to stable particles. A successful theory along these lines would describe the fundamental particles as stable solutions to a set of field equations (with laws of motion derived directly from the field equations) without introducing particles as independent entities, and in this sense reduce mechanics to (generalized) electrodynamics. In effect, Mie treated Maxwell's equations as a weak-field limit of more general field equations. In order to allow for stable charge configurations such as an electron Mie considered non-linear field equations. The fundamental desideratum for the theory was to find generalized field equations that admitted stable solutions representing elementary particles and also reduced to Maxwell's equations in an appropriate limit for regions far from the particles.[7] Mie further aimed to show that gravitation would naturally emerge as a consequence of the generalized field equations.

The key to Mie's theory was the "world function" (Hamiltonian), which he used to derive the field equations via Hamilton's principle.[8] Maxwell's field equations in empty space follow from a Lagrangian $\Phi_{EM} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu}$, where $F_{\mu\nu}$ is the Maxwell tensor and the repeated indices (with $\mu, \nu = 1...4$) are summed over. Mie's pro-

6 Lorentz put the point as follows in 1922, "the formula for momentum is a general consequence of the principle of relativity, and a verification of that formula is a verification of the principle and tells us nothing about the nature of mass or of the structure of the electron," quoted in (Janssen and Mecklenburg 2005).

7 Mie was not the first to consider this way of extending classical electromagnetism. Prior to Mie's work Einstein considered replacing Maxwell's field equations with non-linear, inhomogeneous, and/or higher order equations, as reflected in correspondence with Lorentz and Besso in 1908–1910 (see (McCormach 1970) and (Vizgin 1994), 19–26). Einstein, however, was much more keenly aware than Mie of the deep challenges posed by the quantum structure of radiation.

8 Although Mie formulated his theory within a generalized Hamiltonian framework, in the following we focus on the Lagrangian for his field theory (following Born 1914) for ease of exposition.

gram was to find the terms added to Φ_{EM} that would yield the desired generalized field equations. Mie introduced two fundamental assumptions regarding Φ at the outset of the "Grundlagen" (Mie 1912a). First, electrons and other charged particles should be regarded as "states of the aether" rather than independent entities. Mie insisted that the states of the aether should suffice for a complete physical description of matter, although he admitted that failure of his program might force one to enlarge the allowed fundamental variables. Mie distinguished two different types of fundamental variables, the "intensive quantities" and "extensive quantities," treating the latter as analogous to conjugate momenta in Hamiltonian mechanics (see Mie 1915, 254). To enforce the first assumption Mie required that the world function depends only on the field variables (including the electric charge density, the convection current, the magnetic field strength, and the electric displacement). As Born emphasized (1914, 32), this ruled out treating charged particles with trajectories given by independent equations of motion as the source of the field, since including a coupling to a background current in the Lagrangian (i.e., adding a term proportional to $J^{\mu}\phi_{\mu}$) would explicitly introduce dependence on spacetime coordinates. The second assumption was the validity of special relativity, with the consequence that Φ must be Lorentz covariant. The Lagrangian could only include functions of Lorentz invariant terms constructed from $F_{\mu\nu}$ and ϕ_{μ}, the four-vector potential.[9] Mie argued that functions of only two of these invariants, namely $F_{\mu\nu}F^{\mu\nu}$ and $\phi_{\nu}\phi^{\nu}$ should appear in Φ. For the general field equations to reduce to Maxwell's equations, the $\phi_{\nu}\phi^{\nu}$ term could have non-zero values only in regions occupied by particles.

Mie failed to flesh out his formalism with a specific world function satisfying these constraints that led to a reasonable physical theory. Instead he was limited to illustrating his approach with simple examples, such as a Lagrangian $\Phi = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \alpha(\phi_{\nu}\phi^{\nu})^{3}$, where α is an arbitrary constant. Solutions to the field equations that follow from this Lagrangian could be taken to represent elementary particles, and Mie calculated the charge and mass of the particles. However, these solutions had a number of undesirable features. The arbitrary coefficient appearing in the Lagrangian implied that these solutions placed no constraints on the charge and mass of the "particles," rather than leading to the distinctive values of charge and mass for known particles such as the electron. Mie was further forced to admit (1912b, 38) that his simple world function did not lead to reasonable solutions for interacting charged particles; instead the solutions described a world that eventually separated into two lumps of opposite charge moving away from each other. The simple world

9 Since $F_{\mu\nu} = \frac{\partial\phi_{\nu}}{\partial x_{\mu}} - \frac{\partial\phi_{\mu}}{\partial x_{\nu}}$, the Lagrangian depends on ϕ_{ν} and its first derivatives. The list of invariants included the following quantities:

$$\frac{1}{2}F_{\mu\nu};\ \ \phi_{\nu}\phi^{\nu};\ \ F_{\mu\nu}\phi^{\nu}F^{\mu\rho}\phi_{\rho};\ \ (F_{\mu\nu}\phi_{\rho} + F_{\nu\rho}\phi_{\mu} + F_{\rho\mu}\phi_{\nu})^{2}$$

One invariant was missing from Mie's original list, as Pauli (1921) noted: the quantity $\frac{1}{4}F^{*}_{\mu\nu}\ F^{\mu\nu}$, where $F^{*}_{\mu\nu}$ is the dual of $F_{\mu\nu}$, is an invariant of the restricted Lorentz group, and its square is an invariant of the full Lorentz group.

functions considered by Mie were not viable candidates for a comprehensive description of matter, but he clearly hoped that these problems could be blamed on his lack of ingenuity rather than on his formal framework. However, Pauli (1921, 192) highlighted a problem that went deeper than the failure to find a suitable world function. Mie's world function and the resulting equations of motion both include functions of ϕ_ν. As a result, a stable solution with some value of ϕ is in general not also a solution for $\phi + \text{constant}$, and the world function also fails to be gauge invariant.[10]

Mie hoped that the appropriate world function (supposing one could be found) would incorporate gravity without needing to put it in by hand. At the outset of the *Grundlagen,* Mie announced his goal of deriving gravity from his matter theory without introducing new dynamical variables and sketched a fanciful picture according to which gravity was a consequence of a cohesive shell or atmosphere binding particles together within an atom (Mie 1912a, 512–514). Mie's description of his project may have raised hopes that the third paper would introduce a truly novel approach to gravitation based on non-linear electrodynamics. But like his other grand goals, this one also eluded Mie's grasp.

Mie's gravitational theory has a great deal in common with competing theories due to Abraham and Nordström. Like Nordström, Mie retained an invariant speed of light and upheld the strict validity of the principle of relativity. This sets his approach apart from Abraham's work; Abraham renounced the constancy of the speed of light and retained the validity of the principle of relativity, restricted to infinitesimal spacetime regions, in his first theory, and renounced the principle of relativity all together in his second theory.[11] But like Abraham and Nordström, Mie treated both the source of the gravitational field and the gravitational potential as four-dimensional (Lorentz) scalars, and these were introduced as independent quantities in the world function with no connection to the electromagnetic field. The source of the gravitational field, h, the density of gravitational mass in Mie's theory, is identical to the Hamiltonian density. It is then a short step to derive field equations for the gravitational field appealing to Hamilton's principle. As Mie emphasized, the resulting field equations would be identical to those given by (Abraham 1912) except for the introduction of another variable in the world function.[12] By analogy with his matter theory, Mie introduced an extensive quantity (the excitation of the gravitational field, analogous to electric displacement) conjugate to the gravitational field strength, and argued that the two are identical in an "ideal vacuum" but have an unspecified functional relation in regions occupied by matter.

---

10 Mie recognized this problem and argued that the resulting dependence on the absolute value of the potential would not lead to conflicts with experimental results (Mie 1912b, 24; Mie 1913, 62). Born and Infeld (1934) revived Mie's idea of using a more general Lagrangian, but they excluded additional terms that depended on ϕ_ν to preserve gauge invariance.

11 For discussions of Abraham's and Nordström's theories, see "The Summit Almost Scaled …" and "Einstein, Nordström, and the Early Demise of Scalar, Lorentz Covariant Theories of Gravitation," (both in this volume).

12 See (Mie 1913, 28–29) and the discussion following the Vienna lecture (CPAE 5, Doc. 18).

Mie's failure to achieve a substantial unification illustrates the obstacles to treating gravitation by analogy with electromagnetism. Mie (1915) clearly explained the necessity of introducing the gravitational potential as a dynamical variable in order to resolve the negative energy problem, the most important disanalogy. This problem arises if energy is attributed to the gravitational field itself (as with the electromagnetic field), since the gravitational field strength of, for example, two gravitating masses increases as two masses approach each other, releasing energy in the form of work extracted from the system. One way to save energy conservation in light of this feature of gravitation was to attribute negative energy to the gravitational field, as is suggested by treating Newtonian gravitation in close formal analogy to electrostatics. However, a field with negative energy cannot maintain a stable equilibrium since any small perturbation of the field would in general grow without limit. Following Abraham, Mie argued that the way out of this dilemma was instead to include the gravitational potential in the world function. With this, the internal energy of two approaching masses can be shown to decrease with the decrease of the gravitational potential, thereby compensating for the increase in the field energy.

Including the gravitational potential as a dynamical variable has the consequence that, unlike in electromagnetism, the equations governing physical phenomena depend upon the absolute value of the potential rather than on just potential differences. However, no such dependence had been empirically detected. It remains, moreover, to specify exactly how the field energy depends on the gravitational potential. As (Mie 1915) noted, different choices for this dependence correspond to different gravitational theories. Given the lack of empirical guidance to settle the issue, Mie argued in favor of introducing a principle that would dictate this dependence rather than making what he regarded as arbitrary assumptions. Mie hoped to reconcile his theory's explicit dependence on the absolute value of the gravitational potential with the failure to experimentally detect any such dependence via the theorem (later called a principle) of the relativity of the gravitational potential. The principle plays a central role in the development of Mie's theory, and in elucidating this idea Mie drew a sharp contrast between his approach and Einstein's insistence on generalizing the principle of relativity.

Mie (1915) formulated the principle of the relativity of the gravitational potential as follows:

> In two regions of different gravitational potential exactly the same processes can run according to exactly the same laws if one only thinks of the units of measurement as changing in a suitable way with the value of the gravitational potential. (Mie 1915, 257)

In order to understand the content of this principle, it is perhaps helpful to consider that the principle is equivalent to the requirement that the world function be a homogenous function of the dynamical variables, including the gravitational potential (Mie 1915, 258). From this it immediately follows that in regions of constant gravitational potential, one can transform the potential away, or into any other constant potential, through a rescaling of the remaining dynamical variables and, in general, the space-time coordinates. Thus, for an observer using correspondingly rescaled measuring units to measure the dynamical variables, the gravitational potential will be undetectable. Thinking of Mie's principle of relativity of the gravitational potential along these lines as an invariance of the theory under rescaling, we see the gravitational potential and rescalings in Mie's theory as analogous, respectively, to the metric tensor and general linear transformations in Einstein's tensor theory.[13] Simply put, Mie introduces the gravitational potential into the world function to solve the negative energy problem, and introduces an invariance principle, the principle of the relativity of the gravitational potential, to remove any dependence of physical laws on the potential.

By contrast with Einstein, Mie's introduction of this invariance principle for the gravitational field had no connection with foundational problems in mechanics or with extending the principle of relativity. Mie was clearly quite skeptical of the heuristic value of Einstein's guiding ideas. In the discussion following the Vienna lecture, Mie pointedly criticized the idea of extending the principle of relativity to arbitrary states of motion. Mie pressed Einstein to clarify what would be gained by treating a complicated non-uniform motion, such as a bumpy train ride, as physically equivalent to the gravitational field produced by some array of fictitious planets (CPAE 5, Doc. 18). The underlying problem stemmed from Einstein's failure to distinguish between two claims. In the familiar cases of relativity of uniform motion, the two systems in relative motion are entirely physically equivalent. But Einstein's extension of relativity to non-uniform motion involves a very different claim; as he would later clarify, what is relative in the case of non-uniform motion is how the metric field is split into inertial and gravitational components. This does not, however, imply that two observers in non-uniform motion with respect to each other are physically equivalent. In 1913 Einstein did not answer Mie by drawing this distinction; instead, he replied that his theory did not satisfy an entirely general principle of relativity due to a restriction on allowed coordinate transformations (needed, Einstein thought, to insure energy-momentum conservation). Mie (1914) further argued that since the *Entwurf* theory admits only general linear transformations, it does not realize a general principle of relativity, but in fact satisfies precisely Mie's principle of the relativity of the gravitational potential.

Einstein's equivalence principle was also a target of Mie's criticisms. This is not surprising, since Mie's commitment to retaining the framework of special relativity implied that in his theory inertial and gravitational mass would not be exactly equal. Mie (1915, §§5, 6) calculated the effect of the thermal motions of the constituents of bodies on the relation between inertial and gravitational mass, and argued that departures from exact equality would be well within experimental bounds. Exact equivalence could be had at the price of various auxiliary assumptions, according to Mie, but he did not see the need for such extra assumptions, given that his theory fit experimental constraints. He further claimed that Einstein's theory can only guarantee

13 Our understanding here was guided by (CPAE 8, Doc. 346, fn. 3). Note that in his earlier work, Mie (1912, 61) refers to this as the *theorem* of the relativity of the gravitational potential. Even at this early juncture, though, Mie is quick to elevate this theorem, immediately dubbing it a principle.

exact equivalence by making inconsistent assumptions (Mie 1914, 176). This attitude toward the equivalence principle marks another contrast with Einstein, who took the "unity of essence" of inertia and gravitation to be one of the central foundational insights to be respected by his new theory.

In summary, Mie's work illustrates the potential and limitations of approaching the problem of gravitation within the framework of relativistic field theory. Mie's main innovation in the *Grundlagen* was to consider nonlinear field equations, which opened up the possibility of reducing physics to an electromagnetic matter theory. The appeal of this idea has to be balanced against the theory's glaring deficiency, namely the failure to find a particular world function describing even a simple physical system such as two interacting particles. To paraphrase Einstein, although Mie's theory provided a fine formal framework, it was not clear how to fill it with physical content.[14] Even those sympathetic to Mie's program had to admit doubts that this innovation would lead to a successful matter theory, especially given the recent discoveries of quantum phenomena. But whatever the prospects for matter theory based on generalized electrodynamics, Mie's innovations in the *Grundlagen* turned out to provide few insights for developing a gravitational theory. His own gravitational theory shared the insights and limitations of other Lorentz-covariant theories of gravitation.

In terms of the further development of gravitational theories, Mie's influence on David Hilbert is more significant than his own theory. This influence was mediated by Born's clear reformulation of Mie's theory (Born 1914), which showed how Mie's theory fit into the more general framework of (four-dimensional) Lagrangian continuum mechanics as a special case. Mie's project of unification and his mathematical framework, as refined by Born, shaped Hilbert's distinctive path to a new gravitational theory.[15] But this influence depended on Mie's matter theory and not his gravitational theory, which Born and Hilbert both set aside. Furthermore, Hilbert differed from Mie sharply with regard to the status of special relativity. Mie was a persistent critic of Einstein's move to a metric theory of gravitation and saw no reason to leave the framework of special relativity. Hilbert, on the other hand, took Einstein's *Entwurf* theory as one of his starting points, and his synthesis of Mie's matter theory with Einstein's gravitational theory involved replacing the fixed Minkowski metric of special relativity with Einstein's metric tensor. The fertility of Mie's matter theory for Hilbert depended upon setting aside Mie's own gravitational theory as well as his criticisms of Einstein's extension of special relativity.

---

14 In a 1922 letter to Weyl regarding Eddington's later attempt at a unified field theory, quoted in (Vizgin 1994, 37), Einstein commented that "I find the Eddington argument to have this in common with Mie's theory: it is a fine frame, but one cannot see how it can be filled"; see also his negative assessments of Mie's theory (directly or as it was used by Hilbert) in letters to Freundlich (CPAE 5, Doc. 468), Ehrenfest (CPAE 8, Doc. 220) and Weyl (CPAE 8, Doc. 278). Weyl gave a similar assessment of Mie's theory; see §25 of (Weyl 1918); pp. 214–216 of the English translation.

15 This is explored in great detail in "Hilbert's Foundation of Physics …" (in this volume). See also (Sauer 1999) and (Corry 1999, 2004) for assessments of Hilbert's project and the influence of Mie's theory.

REFERENCES

Abraham, Max. 1912. “Relativität und Gravitation. Erwiderung auf eine Bemerkung des Herrn A. Einstein.” *Annalen der Physik* (38) 1056–1058.

Born, Max. 1914. “Der Impuls-Energie-Satz in der Elektrodynamik von Gustav Mie.” *Königliche Gesellschaft der Wissenschaften zu Göttingen*. Nachrichten (1914): 23–36. (English translation in this volume.)

Buchwald, Jed Z, and Andrew Warwick (eds.). 2001. *Histories of the Electron. The Birth of Microphysics*. Cambridge: The MIT Press.

CPAE 4: Martin J. Klein, A. J. Kox, Jürgen Renn, and Robert Schulmann (eds.). 1995. *The Collected Papers of Albert Einstein. Vol. 4. The Swiss Years: Writings, 1912–1914*. Princeton: Princeton University Press.

CPAE 6: A. J. Kox, Martin J. Klein, and Robert Schulmann (eds.). 1996. *The Collected Papers of Albert Einstein. Vol. 6. The Berlin Years: Writings, 1914–1917*. Princeton: Princeton University Press.

CPAE 6E: *The Collected Papers of Albert Einstein. Vol. 6. The Berlin Years: Writings, 1914–1917*. English edition translated by Alfred Engel, consultant Engelbert Schucking. Princeton: Princeton University Press, 1996.

CPAE 8: Robert Schulmann, A. J. Kox, Michel Janssen, and József Illy (eds.). 1998. *The Collected Papers of Albert Einstein. Vol. 8. The Berlin Years: Correspondence, 1914–1918*. Princeton: Princeton University Press.

CPAE 8E: T*he Collected Papers of Albert Einstein. Vol. 8. The Berlin Years: Correspondence, 1914–1918*. English edition translated by Ann M. Hentschel, consultant Klaus Hentschel. Princeton: Princeton University Press, 1998.

Corry, Leo. 1999. “From Mie’s Electromagnetic Theory of Matter to Hilbert’s Unified Foundations of Physics.” S*tudies in History and Philosophy of Modern Physics* 30 B (2):159–183.

———. 2004. *David Hilbert and the Axiomatization of Physics (1898–1918): From Grundlagen der Geometrie to Grundlagen der Physik*. Dordrecht: Kluwer Academic Publishers.

Einstein, Albert. 1913. “Zum gegenwärtigen Stande des Gravitationsproblems.” *Physikalische Zeitschrift* 14: 1249–1262, (CPAE 4, Doc. 17).

_____.1914. “Zum Relativitätsproblem.” *Scientia* 15: 337–348, CPAE 4, Doc. 31.

Janssen, Michel and Mecklenburg, Matthew. Forthcoming. “The Transition from Classical to Relativistic Mechanics: Electromagnetic Models of the Electron.” To appear in a volume edited by Jesper Lützen based on the proceedings of the conference, *The Interaction between Mathematics, Physics and Philosophy from 1850 to 1940*, Copenhagen, September 26–28, 2002. Earlier version available as Max Planck Institute for the History of Science Preprint 277.

Kohl, Gunter. 2002. “Relativität in der Schwebe: Die Rolle von Gustav Mie”’ Max Planck Institute for the History of Science Preprint 209.

McCormach, Russell. 1970. “Einstein, Lorentz, and the electron theory.” *Historical Studies in the Physical and Biological Sciences*, 2: 41–87.

Mie, Gustav. 1908. “Beiträge zur Optik trüber Medien, speziell kolloidaler Metalllösungen,” *Annalen der Physik* 25, 378–445.

———. 1910. *Lehrbuch der Elektrizität und des Magnetismus*. Stuttgart: F. Enke.

———. 1914. “Bemerkungen zu der Einsteinschen Gravitationstheorie. I und II.” *Physikalische Zeitschrift* 14: 115–122, 169–176.” (English translation in this volume.)

———. 1915. “Das Prinzip von der Realität des Gravitationspotentials.” In *Arbeiten aus den Gebieten der Physik, Mathematik, Chemie. Festschrift Julius Elster und Hans Geitel zum sechzigsten Geburtstag*. Braunschweig: Friedr. Vieweg & Sohn, 251–268. (English translation in this volume.)

Pais, Abraham. 1972. “The Early History of the Theory of the Electron: 1897–1947.” In A. Salam and E. P. Wigner (eds.), *Aspects of Quantum Theory.* Cambridge: Cambridge University Press, 79–93.

Pauli, Wolfgang. 1921. “Relativitätstheorie.” In A. Sommerfeld (ed.), *Encyklopädie der mathematischen Wissenschaften, mit Einschluss ihrer Anwendungen*. Leipzig: B. G. Teubner, 539–775.

———. 1958. *Theory of Relativity.* (Translated by G. Field.) London: Pergamon.

Poincaré, Henri. 1906. “Sur la dynamique de l’électron.” *Rendiconti del Circolo Matematico di Palermo* 21: 129–175. Reprinted in (Poincaré 1934–54), Vol. 9, 494–550. (English translation of excerpt in vol. 3 of this series.)

Sauer, Tilman. 1999. “The Relativity of Discovery: Hilbert’s First Note on the Foundations of Physics.” *Archive for History of Exact Science*s 53: 529–575.

Vizgin, Vladimir. 1994. *Unified Field Theories in the First Third of the 20th Century.* (Translated by Barbour, Julian B, edited by E. Hiebert and H. Wussing.) *Science Networks, Historical Studies,* Vol. 13. Basel, Boston, Berlin: Birkhäuser.

Weyl, Hermann. 1918. *Raum-Zeit-Materie. Vorlesungen über allgemeine Relativitätstheorie*. (First edition.) Berlin: Julius Springer.

GUSTAV MIE

FOUNDATIONS OF A THEORY OF MATTER (EXCERPTS)

Originally published as "Grundlagen einer Theorie der Materie" in Annalen der Physik, first communication, 37, 1912, pp. 511–534 and third communication, 40, 1913, pp. 1–65. Greifswald, Physikalisches Institut, 6 January and 31 October 1912. Chapters 2 and 4 are omitted in the translation.

INTRODUCTION

1. The significance of the recently acquired empirical facts about the nature of the atoms ultimately amounts to something essentially only negative, namely that in the atoms' interior the laws of mechanics and Maxwell's equations cannot be valid. But regarding what should replace these equations in order to encompass from a single standpoint the profusion of remarkable facts associated with the notion of quantum of action, and in addition the laws of atomic spectra and so forth, the experimental evidence is silent. In fact, I believe that one must not expect anything like that from experiment alone. Experiment and theory must work hand in hand, and that is not possible as long as the theory has no foundation on which it can be based.

Thus it seems to me absolutely necessary for further progress of our understanding to supply a new foundation for the theory of matter. With this work, I have tried in the following to make a start, but in view of the difficulty of the matter one should not right away expect results accessible to experiment. The immediate goals that I set myself are: to explain the existence of the indivisible electron and: to view the actuality of gravitation as in a necessary connection with the existence of matter. I believe one must start with this, for electric and gravitational effects are surely the most direct expression of those forces upon which rests the very existence of matter. It would be senseless to imagine matter whose | smallest parts did not possess electric [512]
charges, equally senseless however matter without gravitation. Only when the two goals I mentioned are reached will we be able to consider making the connection between the theory and the complex phenomena mentioned above. But achieving both of these goals is still a long way off, and below I can publish only preliminary work, which will perhaps help us to find the way.

The basic assumption of my theory is *that electric and magnetic fields occur also in the interior of electrons*. According to this view, electrons and accordingly the

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 4
Gravitation in the Twilight of Classical Physics: The Promise of Mathematics.

smallest particles of matter in general are not different in nature from the world aether; they are not foreign bodies in the aether, as was thought maybe 20 years ago, but *they are only locations where the aether has taken on a particular state, which we designate by the term electric charge*. However, the enormous intensity of the field- and charge-states at the location itself that we designate as the electron implies that here the usual Maxwell equations are no longer valid. The behavior of the electromagnetic field inside the electron presumably will be very strange when compared to the laws of the "pure aether." But if we can speak at all of an electromagnetic field in the interior of the electron, then it would not be reasonable that there should not be a continuous transition between the behavior of "pure" aether and that of aether in the electron's interior. Therefore, in my theory the electron is not a particle in the aether with a sharp boundary, but consists of a nucleus with a continuous transition into an atmosphere of electric charge that extends to infinity, but which becomes so extraordinarily dilute already quite close to the nucleus that it cannot be experimentally detected in any way. An atom is an agglomeration of a larger number of electrons glued together by a relatively dilute charge of opposite sign. Atoms are probably surrounded by more substantial atmospheres, which however are still so dilute that they
[513] do not cause noticeable | electric fields, but which presumably are asserted in gravitational effects.

It may seem that there is not much to be gained from the basic assumption just formulated. Still, it leads to a general form for the basic equations of the physics of the aether when combined with two further assumptions. The first is *that the principle of relativity shall be valid generally* and the second *that the presently known states of the aether (that is electric field, magnetic field, electric charge, charge current) suffice completely to describe all phenomena of the material world*. The justification of the first assumption should be beyond doubt. Whether the second holds cannot be said *a priori*. An attempt has to be made. If it can produce a theory that mirrors the material world correctly, then it is thereby vindicated. In the opposite case one will have to ask how the system of fundamental quantities is to be enlarged.

In the following I will present in some detail the reasoning that led me from the three assumptions to a general form of the equations of the aether, in order perhaps to stimulate discussion whether the form I assume is the only possible one, or whether there may not also be other basic equations of aether physics consistent with the three assumptions. I admit that I did not succeed in finding other possibilities. That I presuppose the principle of energy conservation as correct, and assume energy to be a localizable quantity, should go without saying.

FIRST CHAPTER: THE FIELD EQUATIONS

General Form of the Field Equations

2. If one considers Maxwell's equations, preferably in the form given to them by Minkowski, one sees immediately that the four-dimensional six-vector "electromag-

netic field strength" in itself does | not suffice to describe the phenomena in space and [514] time completely. For in addition Maxwell's equations contain an independent four-vector, the "four-current"; which at least therefore has to be included to make the description complete.

In our view, the time component of the four-current, the charge density ρ, represents a peculiar condition of the world aether, which it assumes to a noticeable extent only at isolated places, and which entail that at these places the lines of the electric field $\mathfrak{d}$ simply die out, so that div $\mathfrak{d}$ differs from zero. Therefore we can take the value of div $\mathfrak{d}$ as a measure of this new state of the aether:

$$\rho = \operatorname{div} \mathfrak{d}.$$

Similarly the space component of the four-current, the current of charge $\mathfrak{v}$, describes a peculiar behavior of the aether, which gains noticeable strength only at particular locations; it entails locations of non-vanishing curl of the magnetic field $\mathfrak{h}$ that are not compensated by a time rate of change of the electric field $\mathfrak{d}$. Therefore we can use the difference $\operatorname{curl} \mathfrak{h} - \dot{\mathfrak{d}}$ as a measure of the new state of the aether:

$$\operatorname{curl} \mathfrak{h} - \dot{\mathfrak{d}} = \mathfrak{v}.$$

3. Now we use the basic assumptions mentioned in 1. If all of the material world's processes are to be described by the "electromagnetic field" and the "four-current" together, then by the principle of causality there must be ten differential equations for the ten components of the variables of state $\mathfrak{d}$, $\mathfrak{h}$, ρ, $\mathfrak{v}$ whose left side is always a first-order time derivative of one of the ten quantities, or of a function of them, while on the right-hand side there is a function of the quantities and of their space derivatives. Only such a system of equations can determine from the distribution of the states of the aether at *one* moment the distribution that occurs in the next moment, after passage of an infinitesimal | time dt thus satisfying the principle of causality. [515]

Further, if the principle of relativity is to be valid, then it must be possible to write the derivatives in these equations as vectorial differential operators of four-dimensional quantities. This reduces the number of possibilities enormously. For example, one sees immediately that also with respect to the coordinates only first derivatives can occur; that all derivatives enter only in the first power, etc.

Finally, one must also demand that the equations for the "pure" aether go over into Maxwell's equations, since a continuous transition is assumed between pure aether and matter. Also, the existence of true magnetic charges must be excluded, therefore it must be possible to use a quantity $\mathfrak{b}$ to characterize the magnetic field that everywhere has the property: div $\mathfrak{b} = 0$. Thus we arrive at the equations:

$$\frac{\partial \mathfrak{d}}{\partial t} = \operatorname{curl} \mathfrak{h} - \mathfrak{v}, \tag{1}$$

$$\frac{\partial \mathfrak{b}}{\partial t} = -\operatorname{curl} \mathfrak{e} \tag{2}$$

and here in pure aether $\mathfrak{b}$ must become identical to $\mathfrak{h}$ and $\mathfrak{e}$ identical to $\mathfrak{d}$ whereas in the interior of matter $\mathfrak{e}$ and $\mathfrak{b}$ can be complicated functions of $\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v}$. Equations (1) and (2) then have only a very superficial resemblance to Maxwell's equations. Since at least half of them are no longer linear, the laws of the field in the interior of atoms are quite different from those in pure aether; for example, electromagnetic waves whose existence presupposes linear equations are excluded there, and further such differences.

Thus in the following we will strictly differentiate between the two "intensive quantities": electric field strength $\mathfrak{e}$, and magnetic induction $\mathfrak{b}$, and the "extensive quantities": electric displacement $\mathfrak{d}$, and magnetic field strength $\mathfrak{h}$. Only in pure aether does the principle of superposition of electromagnetic fields hold, which we will express through the equations $\mathfrak{e} = \mathfrak{d}, \mathfrak{b} = \mathfrak{h}$. |

[516] In terms of symbols of the four-dimensional vector analysis[1] the two equations (1) and (2) take the following form:

$$\Delta t v(\mathfrak{h}, -i\mathfrak{d}) = (\mathfrak{v}, i\rho), \tag{1a}$$

$$\Delta t v(\mathfrak{e}, i\mathfrak{b}) = 0. \tag{2a}$$

Now the four equations corresponding to the four-vector $(\mathfrak{v}, i\rho)$ must still be dealt with. For a four-vector there are two kinds of four-dimensional first-order differential operator, namely the operators Div and $\mathfrak{Curl}$.[2] In the first operator the time component, and in the second the three space components, are differentiated with respect to t. So we have to use these two operators to obtain the four differential equations that are still missing. The operator Div occurs in the well-known equation:

$$\frac{\partial\rho}{\partial t} + \operatorname{div} \mathfrak{v} = 0 \tag{3}$$

for this equation becomes in four-dimensional notation:

$$\operatorname{Div}(\mathfrak{v}, i\rho) = 0. \tag{3a}$$

The missing equations must be contained in a formula:

$$\mathfrak{Curl}\,(\mathfrak{f}, i\varphi) = \mathfrak{F}$$

where $(\mathfrak{f}, i\varphi)$ is a four-vector related to $(\mathfrak{v}, i\rho)$ in a similar way as the six-vector $(\mathfrak{b}, -i\mathfrak{e})$ is related to $(\mathfrak{h} - i\mathfrak{d})$. Initially we know only this, that $\mathfrak{f}$ and φ are some functions of all state variables, which taken together form a four-vector. The right side of the equation $\mathfrak{F}$ is some six-vector, also a function of the state variables, of which we know only this, that it must satisfy the condition[3]

1 M. Laue, *Das Relativitätsprinzip*, p. 70. Friedr. Vieweg & Sohn, 1911.
2 M. Laue, loc. cit. p. 70.
3 M. Laue, loc. cit. p. 71.

$$\Delta tv\mathfrak{F}^* = 0,$$

because otherwise it could not be obtained from a four-vector by the $\mathfrak{Curl}$ operator. But now, this | condition must further be identical with equation (2a) unless we assume $\mathfrak{F} = \text{const.}$, in which case it would admittedly be an identity. For, if this were not the case, then we would have three supernumerary equations, besides the ten differential equations that are necessary for the ten state variables by the principle of causality. The time development of the processes in the aether would then be over-determined, which is of course impossible. Therefore we must necessarily have: either $\mathfrak{F} = \text{const.}$ or: $\mathfrak{F} = C \cdot (\mathfrak{b}, -i\mathfrak{e})$, where C is an arbitrary constant factor. We can bring this factor to the other side of our equation $\mathfrak{Curl}(\mathfrak{f}, i\varphi) = \mathfrak{F}$ and absorb it in $\mathfrak{f}, i\varphi$, by simply putting $\mathfrak{F} = (\mathfrak{b}, -i\mathfrak{e})$. The three equations containing a time derivative therefore have this general form: [517]

$$-\frac{\partial \mathfrak{f}}{\partial t} = \nabla\varphi + C \cdot \mathfrak{e} + \mathfrak{c},$$

where C is either zero or one, and $\mathfrak{c}$ denotes a vector that is constant in the entire spacetime region. In a region of pure aether, where $\mathfrak{f} = 0$ as well as $\mathfrak{e} = 0$, it would follow that: $\nabla\varphi = -\mathfrak{c}$. Although all state variables are constant and zero here, φ, which is to be a function of the state variables, would have a non-vanishing gradient, so it would not be constant. This is impossible, therefore we must have: $\mathfrak{c} = 0$. On the other hand it is easy to show that C must be different from zero. If all states of the aether are in equilibrium in the neighborhood of an electron moving with constant velocity, then all time derivatives must be zero. The equation then reads:

$$\nabla\varphi + C \cdot \mathfrak{e} = 0.$$

Now if $C = 0$, then we would also have $\nabla\varphi = 0$, hence $\varphi = \text{const.}$ Then the quantity φ would not depend on the field quantities at all, the same conclusion would hold by the principle of relativity also for $\mathfrak{f}$, and the equation we found would then reduce to an identity. Therefore it must be that $C = 1$. Accordingly, the last three equations of aether dynamics are:

$$-\frac{\partial \mathfrak{f}}{\partial t} = \nabla\varphi + \mathfrak{e}, \tag{4}$$

| or, written in four-dimensional symbols: [518]

$$\mathfrak{Curl}(\mathfrak{f}, i\varphi) = (\mathfrak{b}, -i\mathfrak{e}). \tag{4a}$$

The expression (4a) contains also the following three equations, which contain no time derivative:

$$\mathfrak{curl}\ \mathfrak{f} = \mathfrak{b}. \tag{4b}$$

It is easily seen that the equations (4b) can be derived from (4) with the aid of (2), so they contain nothing new.

If everything is in equilibrium in the vicinity of an electron at rest or in uniform motion, then equation (4) becomes:

$$\nabla\varphi + \mathfrak{e} = 0.$$

We may denote this as the *equilibrium condition* for the field in the vicinity of the electron. It can be interpreted intuitively as saying that the two forces $\mathfrak{e}$ and $\nabla\varphi$ shall be equal and opposite to each other. The electric field strength $\mathfrak{e}$ tends to pull the electron's charge outward and to spread it over as large a region as possible, so it represents the *force of expansion* inherent in matter. It is balanced by the *force of cohesion* $\nabla\varphi$, computed as the gradient of a pressure of cohesion φ peculiar to the electric charge in itself.[4] Forces of expansion and cohesion are the two effects upon which any existence of matter is based, so they must occur in every possible theory of matter.

Equation (4) can be characterized as the equation of motion of the charge current. The vector $\mathfrak{f}$ is the *momentum* [*Bewegungsgröße*] related to the charge current $\mathfrak{v}$. In the usual mechanics we know the momentum to be mass times velocity and measure it by the impact necessary to produce that velocity. Since momentum and pressure are to be characterized as "intensive quantities", that is, quantities to be measured
[519] through | action of forces, we will also contrast φ and $\mathfrak{f}$ as "intensive quantities" with their corresponding "extensive quantities" ρ and $\mathfrak{v}$.

Thus we can describe the state of the aether either by ten extensive quantities $(\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v})$ or by ten intensive quantities $(\mathfrak{e}, \mathfrak{b}, \varphi, \mathfrak{f})$.

4. The six differential equations (4) and (4b), which are combined in (4a) into one formula, are exactly the same as the differential equations for the so-called four-potential, which is composed of the scalar potential φ and the vector potential $\mathfrak{f}$. Therefore it could be said with some justification that the theory developed here consists simply in attributing to the two potentials φ and $\mathfrak{f}$ the meaning of physical states of the world aether, namely as cohesion pressure and momentum.

Here we must add an important remark. One knows that the solution of equations (4a) for a given six-vector $(\mathfrak{b}, -i\mathfrak{e})$ is undetermined unless one makes a further assumption about $\mathrm{Div}(\mathfrak{f}, i\varphi)$. In the theory of electricity one defines the two potentials by simply putting $\mathrm{Div}(\mathfrak{f}, i\varphi) = 0$. But this equation does not hold for the states of the aether assumed in our theory, and therefore they are in general not identical with the potentials as usually calculated. Namely, the equation of electricity written above is replaced in our aether dynamics by equation (3): $\mathrm{Div}(\mathfrak{v}, i\rho) = 0$. This cannot coexist with the other equation because then the temporal evolution of the aether processes would be governed by eleven equations, and would hence be overdetermined, which is impossible. Therefore in general we have $\mathrm{Div}(\mathfrak{f}, i\varphi) \neq 0$. In a later section (p. 651 [p. 534 in the original]) we will find a simple interpretation for the quantity $\mathrm{Div}(\mathfrak{f}, i\varphi)$.

4 As is well known, such a pressure was first assumed by H. Poincaré; (*Compt. rend.* 140, p. 1504, 1905. Cf. also H. Th. Wolff, *Ann. d. Phys.* 36, p. 1066, 1911.

In the case of rest $(\mathfrak{v} = 0,\ \mathfrak{h} = 0)$ the quantity φ is indeed identical with the electrostatic potential, because we then have the equation:

$$\mathfrak{e} + \nabla\varphi = 0.$$

| 5. When associating φ with pressure and ρ with a density one could easily [520]
believe that it is advantageous always to associate positive values with these quantities, similar to the way it is done in the physics of gases.

We would then ascribe a constant positive value ρ_0, the normal density, to the pure aether when entirely free from fields; for an arbitrary choice of spacetime coordinate system it would of course have to be a four-vector $(\mathfrak{v}_0, i\rho_0)$ that would be constant in the entire spacetime region. Electric and magnetic fields would appear only where ρ and $\mathfrak{v}$ take on values different from ρ_0 and $\mathfrak{v}_0$, and equations (1) and (3) would therefore have to be:

$$\Delta tv(\mathfrak{h}, -i\mathfrak{d}) = ((\mathfrak{v} - \mathfrak{v}_0),\ \ i \cdot (\rho - \rho_0)),$$
$$\mathrm{Div}((\mathfrak{v} - \mathfrak{v}_0),\ \ i \cdot (\rho - \rho_0)) = 0.$$

Now, one could of course choose ρ_0 so that the quantity ρ occurring in these equations, the "density of aether", would always be positive. But one cannot see what advantage this would bring. In the following I will therefore always again write simply $\mathfrak{v} - \mathfrak{v}_0$ and $\rho - \rho_0$ instead of $\mathfrak{v}$ and ρ; that is, I will calculate with positive and negative densities by putting the density of pure aether equal to zero.

The same applies to the cohesive pressure φ. Since φ and $\mathfrak{f}$ occur only differentiated with respect to time or space in the fundamental equations of aether physics, one can add to them a quite arbitrary time- and space-independent quantity $\varphi_0, \mathfrak{f}_0$ without changing the description of the processes in any essential way. For example, one could choose a large enough value φ_0, so that $\varphi_0 - \varphi$ always remains positive. The equilibrium condition would then be:

$$\mathfrak{e} - \nabla(\varphi_0 - \varphi) = 0.$$

In pure aether we would now have the large positive pressure φ_0, in an electron we would have the smaller pressure $(\varphi_0 - \varphi)$, and $\mathfrak{e}$ would keep the pressure gradient $-\nabla(\varphi_0 - \varphi)$, which the aether exerts on the electron, in balance. Indeed H. Poincaré (loc. cit.) speaks of a pressure exerted on the electron from the outside. But I believe that for a description it is easier if | we put the zero point of the pressure in the pure [521]
aether, and therefore in my calculations I will always put φ equal to zero at infinite distance from an electron.

Similarly for energy, which we know can always be augmented by an arbitrary additive constant, let us fix the zero point in such a way that the energy density in pure, field-free aether equals zero. Similar to ρ and φ the energy density W will then admittedly be allowed to assume negative as well as positive values; but after all there is not the slightest reason that would force us to always make W positive.

With these standardizations, ρ, φ, W are now to be regarded as completely determined quantities without any additive arbitrariness.

The Energy

6. I presuppose that not only the principle of energy conservation, but also the principle of energy localization and energy transfer[5] is valid. That is: if we denote the energy density by W and the energy flux by $\mathfrak{s}$, then from the field equations (1) to (4), the following relation must follow:

$$\frac{\partial W}{\partial t} = -\mathrm{div}\ \mathfrak{s},$$

where the scalar W as well as the vector $\mathfrak{s}$ are universal functions of the state prevailing at the spacetime point concerned. From the field equations one can arrive at this energy equation in only *one* way: one must determine factors $\mathfrak{k}, \mathfrak{l}, m, \mathfrak{n}$, which are universal functions of the state variables, multiply the equations (1) to (4) by them, and then add the equations. So it must then be possible to pick the factors $\mathfrak{k}, \mathfrak{l}, m, \mathfrak{n}$ in such a way that the left side then becomes a complete time derivative, and the right
[522] side becomes a divergence. | Let us now examine the conditions under which this is possible.

$$\mathfrak{k} \cdot \frac{\partial \mathfrak{d}}{\partial t} + \mathfrak{l} \cdot \frac{\partial \mathfrak{b}}{\partial t} + m \cdot \frac{\partial \rho}{\partial t} + \mathfrak{n} \cdot \frac{\partial \mathfrak{f}}{\partial t}$$
$$= \mathfrak{k} \cdot \mathrm{curl}\ \mathfrak{h} - \mathfrak{k} \cdot \mathfrak{v} - \mathfrak{l} \cdot \mathrm{curl}\ \mathfrak{e} - m \cdot \mathrm{div}\ \mathfrak{v} - \mathfrak{n} \cdot \nabla\varphi - \mathfrak{n} \cdot \mathfrak{e}.$$

First we see that the two terms $-\mathfrak{k} \cdot \mathfrak{v}$ and $-\mathfrak{n} \cdot \mathfrak{e}$, which are pure universal functions of the state variables, must cancel, because div $\mathfrak{s}$ can consist only of terms containing derivatives with respect to the coordinates. Therefore we must have:

$$\mathfrak{k} = u \cdot \mathfrak{e},$$
$$\mathfrak{n} = -u \cdot \mathfrak{v},$$

where u is again a universal function of the state variables. A small manipulation yields for the right side of the equation:

$$\mathrm{div}(u \cdot [\mathfrak{h} \cdot \mathfrak{e}]) + \mathrm{div}(u \cdot \varphi \cdot \mathfrak{v}) + (u \cdot \mathfrak{h} - \mathfrak{l}) \cdot \mathrm{curl}\ \mathfrak{e} - \mathfrak{h} \cdot [\mathfrak{e} \cdot \nabla u]$$
$$-(m + u \cdot \varphi) \cdot \mathrm{div}\ \mathfrak{v} - \varphi \cdot (\mathfrak{v} \cdot \nabla u).$$

This expression can in general be a divergence only if the last four terms are annulled, that is if:

5 G. Mie, *Wiener Sitzungsber.* 107, sec. IIa, p. 1117 and 1126, 1898.

$$\begin{aligned} \nabla u &= 0, \\ u \cdot \mathfrak{h} - \mathfrak{l} &= 0, \\ m + u \cdot \varphi &= 0. \end{aligned}$$

The first of these equations implies $u = \text{const.}$, and specifically the value of this constant is determined by the demand that in pure aether the energy flux should become the well-known Poynting expression $[\mathfrak{d} \cdot \mathfrak{h}] = [\mathfrak{e} \cdot \mathfrak{h}]$. From this it follows that:

$$u = 1, \quad \mathfrak{k} = \mathfrak{e}, \quad \mathfrak{l} = \mathfrak{h}, \quad m = -\varphi, \quad \mathfrak{n} = -\mathfrak{v}.$$

Thus we have found for the energy equation:

$$\mathfrak{e} \cdot \frac{\partial \mathfrak{d}}{\partial t} + \mathfrak{h} \cdot \frac{\partial \mathfrak{b}}{\partial t} - \varphi \cdot \frac{\partial \rho}{\partial t} - \mathfrak{v} \cdot \frac{\partial \mathfrak{f}}{\partial t} = -\operatorname{div}([\mathfrak{e} \cdot \mathfrak{h}] - \varphi \cdot \mathfrak{v}).$$

Accordingly the expression for the *energy flux* in the general aether dynamics is:

$$\mathfrak{s} = [\mathfrak{e} \cdot \mathfrak{h}] - \varphi \cdot \mathfrak{v}. \tag{5}$$

| 7. The energy principle further demands that the expression on the left side of the energy equation be a complete differential. So we must formulate the condition that: [523]

$$\mathfrak{e} \cdot d\mathfrak{d} + \mathfrak{h} \cdot d\mathfrak{b} - \varphi \cdot d\rho - \mathfrak{v} \cdot d\mathfrak{f} = dW \tag{6}$$

be a complete differential, so that W can be determined as a function of $(\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v})$. Just as well as W we can examine a quantity H, determined by the following equation:

$$W = H + \mathfrak{h} \cdot \mathfrak{b} - \mathfrak{v} \cdot \mathfrak{f}. \tag{7}$$

If W is a function of $(\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v})$, then so is H, and vice versa. From (6) and (7) we obtain the following expression for the differential of H:

$$dH = \mathfrak{e} \cdot d\mathfrak{d} - \mathfrak{b} \cdot d\mathfrak{h} - \varphi \cdot d\rho + \mathfrak{f} \cdot d\mathfrak{v}, \tag{8}$$

where $\mathfrak{e}, \mathfrak{b}, \varphi, \mathfrak{f}$ are functions of $(\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v})$. For brevity a vector whose components are

$$\frac{\partial H}{\partial \mathfrak{d}_x}, \quad \frac{\partial H}{\partial \mathfrak{d}_y}, \quad \frac{\partial H}{\partial \mathfrak{d}_z}$$

will be called simply $\partial H / \partial \mathfrak{d}$, and analogously in similar cases. Then it follows directly from (8) that:

$$\mathfrak{e} = \frac{\partial H}{\partial \mathfrak{d}}, \quad \mathfrak{b} = -\frac{\partial H}{\partial \mathfrak{h}}, \quad \varphi = -\frac{\partial H}{\partial \rho}, \quad \mathfrak{f} = \frac{\partial H}{\partial \mathfrak{v}}. \tag{9}$$

The condition that the energy principle be valid is that all intensive quantities $\mathfrak{e}, \mathfrak{b}, \varphi, \mathfrak{f}$ *can be calculated by means of a single function of the extensive quantities*

$H(\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v})$, *which we will call the Hamiltonian function. Specifically, every intensive quantity is to be obtained as the derivative of H with respect to the corresponding extensive quantity, in two cases (*$\mathfrak{b}$ *and* φ *) with a negative sign.*

The energy density W can now also be found from the Hamiltonian function alone. For (7) together with (9) result in:

$$W = H - \frac{\partial H}{\partial \mathfrak{h}} \cdot \mathfrak{h} - \frac{\partial H}{\partial \mathfrak{v}} \cdot \mathfrak{v}. \tag{10}$$

The form of the basic equations (1) to (4) of aether dynamics, with equations (9) taken into account, leads immediately to the following theorem: |

[524] *The relativity principle is valid for all physical processes, provided the Hamiltonian function* $H(\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v})$ *is invariant under Lorentz transformations.*

Now we would have the complete formulation of the equations of aether dynamics, if we only knew the form of the universal function H. To find this form is an extremely difficult task indeed.

The problem of a theory of matter is reduced to the problem of finding the universal function $H(\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v})$.

So far we know only one thing about H: in pure aether the superposition principle for electromagnetic fields holds with great precision; so if one takes an additive term $(\mathfrak{d}^2 - \mathfrak{h}^2)/2$ out of H:

$$H = \frac{1}{2}(\mathfrak{d}^2 - \mathfrak{h}^2) + H_1,$$

then the remainder H_1 must be quite vanishingly small compared to the first term at places where ρ is very small. However, in contrast, in the interior of atoms, where ρ is large, H_1 will dominate by far, so that here the laws of the field are quite different than those in pure aether.

8. For calculations it is in general much more convenient to choose the intensive variables $(\mathfrak{e}, \mathfrak{b}, \varphi, \mathfrak{f})$ as the independent variables that describe the state of the aether, and to view the extensive quantities $(\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v})$ as functions of the former.

Let us now form the following function Φ:

$$\Phi(\mathfrak{e}, \mathfrak{b}, \varphi, \mathfrak{f}) = H - (\mathfrak{e} \cdot \mathfrak{d} - \mathfrak{b} \cdot \mathfrak{h}) + (\varphi \cdot \rho - \mathfrak{f} \cdot \mathfrak{v}), \tag{11}$$

by first solving equations (9) for $\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v}$ as functions of $\mathfrak{e}, \mathfrak{b}, \varphi, \mathfrak{f}$ and then substituting the expressions so found on the right side of equation (11). Using (8) we get for the differential of Φ the following expression:

$$d\Phi = -\mathfrak{d} \cdot d\mathfrak{e} + \mathfrak{h} \cdot d\mathfrak{b} + \rho \cdot d\varphi - \mathfrak{v} \cdot d\mathfrak{f}. \tag{12}$$

From this follows:

$$\mathfrak{d} = -\frac{\partial \Phi}{\partial \mathfrak{e}}, \quad \mathfrak{h} = \frac{\partial \Phi}{\partial \mathfrak{b}}, \quad \rho = \frac{\partial \Phi}{\partial \varphi}, \quad \mathfrak{v} = -\frac{\partial \Phi}{\partial \mathfrak{f}}. \tag{13}$$

[525] | *The extensive quantities* $\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v}$ *can all be calculated with the aid of a single function of the intensive quantities* $\Phi(\mathfrak{e}, \mathfrak{b}, \varphi, \mathfrak{f})$ *by differentiating the latter with*

respect to the corresponding intensive quantities. In two cases ($\mathfrak{d}$ and $\mathfrak{v}$) one has to give the derivative a negative sign.

The energy density W results from Φ as follows:

$$W = \Phi + \mathfrak{e}\cdot\mathfrak{d} - \varphi\cdot\rho = \Phi - \frac{\partial\Phi}{\partial\mathfrak{e}}\cdot\mathfrak{e} - \frac{\partial\Phi}{\partial\varphi}\cdot\varphi. \tag{14}$$

The Hamiltonian function H is calculated according to (11):

$$H = \Phi - \frac{\partial\Phi}{\partial\mathfrak{e}}\cdot\mathfrak{e} - \frac{\partial\Phi}{\partial\mathfrak{b}}\cdot\mathfrak{b} - \frac{\partial\Phi}{\partial\varphi}\cdot\varphi - \frac{\partial\Phi}{\partial\mathfrak{f}}\cdot\mathfrak{f}. \tag{15}$$

Instead of looking for the universal function $H(\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v})$ one can as well search for the universal function $\Phi(\mathfrak{e}, \mathfrak{b}, \varphi, \mathfrak{f})$.

I will often designate Φ as *the world function* for short.

Under Lorentz transformations Φ as well as H must be an invariant. Similar to H, one can divide Φ in two parts:

$$\Phi = \frac{1}{2}(\mathfrak{b}^2 - \mathfrak{e}^2) + \Phi_1,$$

the first of which dominates in pure aether, and the second in the interior of atoms.

9. With the aid of the world function one can form a four by four matrix[6] that contains the energy flux and Maxwell's aether stresses for our general aether dynamics:

$$S = \left|\begin{array}{c}
\Phi - \mathfrak{b}\mathfrak{h} + \mathfrak{e}_x\mathfrak{d}_x + \mathfrak{h}_x\mathfrak{b}_x + \mathfrak{f}_x\mathfrak{v}_x, \quad \mathfrak{e}_x\mathfrak{d}_y + \mathfrak{h}_x\mathfrak{b}_y + \mathfrak{f}_x\mathfrak{v}_y, \\
\mathfrak{e}_x\mathfrak{d}_z + \mathfrak{h}_x\mathfrak{b}_z + \mathfrak{f}_x\mathfrak{v}_z, \quad -i\cdot(\mathfrak{d}_y\mathfrak{b}_z - \mathfrak{d}_z\mathfrak{b}_y - \rho\mathfrak{f}_x), \\
\mathfrak{e}_y\mathfrak{d}_x + \mathfrak{h}_y\mathfrak{b}_x + \mathfrak{f}_y\cdot\mathfrak{v}_x, \quad \Phi - \mathfrak{b}\mathfrak{h} + \mathfrak{e}_y\mathfrak{d}_y + \mathfrak{h}_y\mathfrak{b}_y + \mathfrak{f}_y\mathfrak{v}_y, \\
\mathfrak{e}_y\mathfrak{d}_z + \mathfrak{h}_y\mathfrak{b}_z + \mathfrak{f}_y\mathfrak{v}_z, \quad -i\cdot(\mathfrak{d}_z\mathfrak{b}_x - \mathfrak{d}_x\mathfrak{b}_z - \rho\mathfrak{f}_y), \\
\mathfrak{e}_z\mathfrak{d}_x + \mathfrak{h}_z\mathfrak{b}_x + \mathfrak{f}_z\mathfrak{v}_x, \quad \mathfrak{e}_z\mathfrak{d}_y + \mathfrak{h}_z\mathfrak{b}_y + \mathfrak{f}_z\mathfrak{v}_y, \\
\Phi - \mathfrak{b}\mathfrak{h} + \mathfrak{e}_z\mathfrak{d}_z + \mathfrak{h}_z\mathfrak{b}_z + \mathfrak{f}_z\mathfrak{v}_z, \quad -i\cdot(\mathfrak{d}_x\mathfrak{b}_y - \mathfrak{d}_y\mathfrak{b}_x - \rho\cdot\mathfrak{f}_z), \\
-i\cdot(\mathfrak{e}_y\mathfrak{h}_z - \mathfrak{e}_z\mathfrak{h}_y - \varphi\cdot\mathfrak{v}_x), \quad -i(\mathfrak{e}_z\cdot\mathfrak{h}_x - \mathfrak{e}_x\cdot\mathfrak{h}_z - \varphi\cdot\mathfrak{v}_y), \\
-i\cdot(\mathfrak{e}_x\mathfrak{h}_y - \mathfrak{e}_y\mathfrak{h}_x - \varphi\cdot\mathfrak{v}_z), \quad \Phi + \mathfrak{e}\mathfrak{d} - \varphi\cdot\rho
\end{array}\right| \tag{16}$$

| If the operation [526]

$$\Delta tv = \frac{\partial}{\partial x} + \frac{\partial}{\partial y} + \frac{\partial}{\partial z} + \frac{\partial}{i\cdot\partial t}$$

is applied to the lowest row of the matrix one obtains the energy equation by putting the expression so obtained equal to zero:

6 H. Minkowski, *Zwei Abhandlungen*. B. G. Teubner 1910, p. 36.

$$\operatorname{div}([\mathfrak{e}\cdot\mathfrak{h}]-\varphi\cdot\mathfrak{v})+\frac{\partial}{\partial t}(\Phi+\mathfrak{e}\cdot\mathfrak{d}-\varphi\cdot\rho) \;=\; 0,$$

because according to (14) we have $\Phi+\mathfrak{e}\cdot\mathfrak{d}-\varphi\cdot\rho \;=\; W$. From the principle of relativity it then follows directly that:

$$\Delta tvS \;=\; 0. \tag{17}$$

Incidentally, it is also not much effort to obtain the first three rows of S directly from the field equations (1) to (4).

Whether the matrix (16) is symmetric about its diagonal is a question to which we will return again later (p. 650) [p. 533 in original].

Hamilton's Principle

10. Whether the above is an unobjectionable proof that only the form of the field equations as formulated by me is possible, may still be open to discussion. Therefore it seems to me to be valuable to show that the field equations can be obtained by quite simple mathematical operations, assuming the validity of Hamilton's principle.

So I make only the following two assumptions: First, the state of the aether is completely characterized by the quantities $\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v}$, where the last two are defined by the constraints:

$$\rho \;=\; \operatorname{div}\mathfrak{d}, \quad \mathfrak{v} \;=\; \mathfrak{curl}\,\mathfrak{h}-\dot{\mathfrak{d}};$$

second, the aether processes satisfy Hamilton's principle formulated as follows.

Hamilton's Principle. There exists a function $H(\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v})$, *whose integral over*
[527] *any spacetime region with determined boundary is an extremum | for all actual processes if the state variables are varied at all points in the interior of the region, but not on the boundary of the region*:

$$\int_G \delta H(\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v})\cdot dx\cdot dy\cdot dz\cdot dt \;=\; 0. \tag{18}$$

On the boundary of the region G we have:

$$\delta\mathfrak{d} \;=\; \delta\mathfrak{h} \;=\; \delta\rho \;=\; \delta\mathfrak{v} \;=\; 0.$$

It can be shown that the principle of relativity is valid if H is invariant under Lorentz transformations. Let us assume that this is the case and replace the quantities $\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v}$ by the well-known expressions in terms of $\mathfrak{d}', \mathfrak{h}', \rho', \mathfrak{v}'$, that are to take their place upon a transformation of the coordinate system (x, y, z, t) to another (x', y', z', t'); then we must obtain a function $H'(\mathfrak{d}', \mathfrak{h}', \rho', \mathfrak{v}')$, that is built out of the new variables $(\mathfrak{d}', \mathfrak{h}', \rho', \mathfrak{v}')$ in exactly the same way as H is built out of the old variables $(\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v})$. We express this by setting: $H' \;=\; H$. G' be the region in the new coordinate system (x', y', z', t') into which G transforms. We then have:

$$\int_{G'} H(\mathfrak{d}', \mathfrak{h}', \rho', \mathfrak{v}') \cdot dx' \cdot dy' \cdot dz' \cdot dt' = \int_{G} H(\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v}) \cdot dx \cdot dy \cdot dz \cdot dt.$$

It follows from this equation that if Hamilton's principle holds for the coordinate system (x, y, z, t), then it also holds for every other arbitrary system (x', y', z', t'). And here the Hamiltonian function is in all coordinate system the same function H.

Accordingly the laws of nature, that is the differential equations resulting from Hamilton's principle, are the same in all coordinate systems that can be obtained by Lorentz transformations. That is the principle of relativity.

Now we want to derive the field equations from Hamilton's principle. To this end we form the variation

$$\delta H = \frac{\partial H}{\partial \mathfrak{d}} \cdot \delta\mathfrak{d} + \frac{\partial H}{\partial \mathfrak{h}} \cdot \delta\mathfrak{h} + \frac{\partial H}{\partial \rho} \cdot \delta\rho + \frac{\partial H}{\partial \mathfrak{v}} \cdot \delta\mathfrak{v}.$$

Now we want to introduce the following abbreviations:

$$\frac{\partial H}{\partial \mathfrak{d}} = \mathfrak{e}, \quad \frac{\partial H}{\partial \mathfrak{h}} = -\mathfrak{b}, \quad \frac{\partial H}{\partial \rho} = -\varphi, \quad \frac{\partial H}{\partial \mathfrak{v}} = \mathfrak{f}. \tag{19}$$

| The variation of H is then: [528]

$$\delta H = \mathfrak{e} \cdot \delta\mathfrak{d} - \mathfrak{b} \cdot \delta\mathfrak{h} - \varphi \cdot \delta\rho + \mathfrak{f} \cdot \delta\mathfrak{v}. \tag{20}$$

To transform this expression further we use a formula from four-dimensional vector calculus, which we want briefly to derive: The product of the four-vector $\mathsf{P} = (\mathfrak{f}, i\varphi)$ and the six-vector $\mathfrak{F} = (\mathfrak{h}, -i\mathfrak{d})$ is taken to be the following four-vector:[7]

$$[\mathsf{P} \cdot \mathfrak{F}] = ([\mathfrak{f} \cdot \mathfrak{h}] + \varphi \cdot \mathfrak{d}, i \cdot (\mathfrak{f} \cdot \mathfrak{d})).$$

We form the Div of this vector:

$$\mathrm{Div}[\mathsf{P} \cdot \mathfrak{F}] = \mathrm{div}\{[\mathfrak{f} \cdot \mathfrak{h}] + \varphi \cdot \mathfrak{d}\} + \frac{\partial(\mathfrak{f} \cdot \mathfrak{d})}{\partial t}.$$

But we have:

$$\begin{aligned} \mathrm{div}[\mathfrak{f} \cdot \mathfrak{h}] &= \mathfrak{h} \cdot \mathrm{curl}\, \mathfrak{f} - \mathfrak{f} \cdot \mathrm{curl}\, \mathfrak{h}, \\ \mathrm{div}(\varphi \cdot \mathfrak{d}) &= \mathfrak{d} \cdot \nabla\varphi + \varphi \cdot \mathrm{div}\, \mathfrak{d}, \\ \frac{\partial(\mathfrak{f} \cdot \mathfrak{d})}{\partial t} &= \mathfrak{d} \cdot \frac{\partial \mathfrak{f}}{\partial t} + \mathfrak{f} \cdot \frac{\partial \mathfrak{d}}{\partial t}. \end{aligned}$$

and therefore:

7 M. Laue, *Das Relativitätsprinzip*, p. 67.

$$\operatorname{div}\{[\mathfrak{f}\cdot\mathfrak{h}]+\varphi\cdot\mathfrak{d}\}+\frac{\partial(\mathfrak{f}\cdot\mathfrak{d})}{\partial t} = \mathfrak{h}\cdot\operatorname{\mathfrak{curl}}\mathfrak{f}+\mathfrak{d}\cdot\left(\nabla\varphi+\frac{\partial\mathfrak{f}}{\partial t}\right) -\mathfrak{f}\cdot\left(\operatorname{\mathfrak{curl}}\mathfrak{h}-\frac{\partial\mathfrak{d}}{\partial t}\right)+\varphi\cdot\operatorname{div}\mathfrak{d}. \tag{21}$$

In four-dimensional symbols this formula becomes:

$$\operatorname{Div}[\mathsf{P}\cdot\mathfrak{F}] = -(\mathfrak{F}\cdot\mathfrak{Curl}\,\mathsf{P})-(\mathsf{P}\cdot\Delta tv\mathfrak{F}). \tag{22}$$

Now we want to apply this formula to our problem, noting that:

$$\operatorname{\mathfrak{curl}}\mathfrak{h}-\frac{\partial\mathfrak{d}}{\partial t} = \mathfrak{v},\quad \operatorname{div}\mathfrak{d} = \rho.$$

Therefore we have:

$$\operatorname{Div}[\mathsf{P}\cdot\mathfrak{F}] = \mathfrak{h}\cdot\operatorname{\mathfrak{curl}}\mathfrak{f}+\mathfrak{d}\cdot\left(\nabla\varphi+\frac{\partial\mathfrak{f}}{\partial t}\right)-\mathfrak{f}\cdot\mathfrak{v}+\varphi\cdot\rho.$$

Replacing $\mathfrak{d}$ and $\mathfrak{h}$ by the variation $\delta\mathfrak{d}$ and $\delta\mathfrak{h}$ we find:

$$\mathfrak{f}\cdot\delta\mathfrak{v}-\varphi\cdot\delta\rho = \operatorname{\mathfrak{curl}}\mathfrak{f}\cdot\delta\mathfrak{h}+\left(\nabla\varphi+\frac{\partial\mathfrak{f}}{\partial t}\right)\cdot\delta\mathfrak{d}-\operatorname{Div}[\mathsf{P}\cdot\delta\mathfrak{F}].$$

[529] | Now, the integral:

$$\int_G \operatorname{Div}[\mathsf{P}\cdot\delta\mathfrak{F}]\cdot dx\cdot dy\cdot dz\cdot dt,$$

can, just like the volume integral over a three dimensional divergence, be converted into an integral over the boundary of G. But since Hamilton's principle prescribes that the variations of all state variables, including $\delta\mathfrak{F}$, vanish on the boundary, we have:

$$\int_G \operatorname{Div}[\mathsf{P}\cdot\delta\mathfrak{F}]\cdot dx\cdot dy\cdot dz\cdot dt = 0.$$

Consequently, use of formula (20) for δH results in:

$$\int_G \delta H\cdot dx\cdot dy\cdot dz\cdot dt = \int_G\left(\left(\mathfrak{e}+\nabla\varphi+\frac{\partial\mathfrak{f}}{\partial t}\right)\cdot\delta\mathfrak{d}+(\operatorname{\mathfrak{curl}}\mathfrak{f}-\mathfrak{b})\cdot\delta\mathfrak{h}\right)\cdot dx\ dy\ dz\ dt.$$

Since there are no longer any constraints between $\mathfrak{d}$ and $\mathfrak{h}$, so that $\delta\mathfrak{h}$ and $\delta\mathfrak{d}$ are quite independent of each other, Hamilton's principle can be satisfied only if the following two differential equations hold:

$$\mathfrak{e} + \nabla\varphi + \frac{\partial \mathfrak{f}}{\partial t} = 0,$$
$$\mathrm{curl}\,\mathfrak{f} - \mathfrak{b} = 0.$$

These two equations furthermore lead to:

$$\frac{\partial \mathfrak{b}}{\partial t} + \mathrm{curl}\,\mathfrak{e} = 0.$$

Since the equations of definition (19) for $\mathfrak{e}, \mathfrak{b}, \varphi, \mathfrak{f}$ agree completely with equations (9), these equations are identical with the field equations (2) and (4); and equations (1) and (3) we assumed *a priori* as equations of definition.

Hereby it has been proved that the form of the field equations I assumed is the only form in accordance with Hamilton's principle. |

Finally let us remark that equation (21) can be given yet another interesting form [530]
by noting that:

$$\mathrm{curl}\,\mathfrak{f} = \mathfrak{b}, \quad \nabla\varphi + \frac{\partial \mathfrak{f}}{\partial t} = -\mathfrak{e}, \quad \mathrm{curl}\,\mathfrak{h} - \frac{\partial \mathfrak{d}}{\partial t} = \mathfrak{v}, \quad \mathrm{div}\,\mathfrak{d} = \rho.$$

Taking into account equation (11) we thus find:

$$\frac{\partial(\mathfrak{f} \cdot \mathfrak{d})}{\partial t} + \mathrm{div}\{[\mathfrak{f} \cdot \mathfrak{h}] + \varphi \cdot \mathfrak{d}\} = \Phi - H. \tag{23}$$

The Invariants

11. In order that the function $H(\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v})$ be invariant under Lorentz transformations, i.e. be a four-dimensional scalar, it must be a function of nothing but four-dimensional scalars that can be formed from $\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v}$. There are four such quantities that are independent of each other.

1. The absolute value of the four-vector $\mathsf{P} = (\mathfrak{v}, i\rho)$. It is:

$$\sigma = \sqrt{\rho^2 - \mathfrak{v}^2} = \rho \cdot \sqrt{1 - \beta^2}, \quad \beta = \frac{\mathfrak{v}}{\rho}.$$

2. The absolute value of the six-vector $\mathfrak{F} = (\mathfrak{h}, -i\mathfrak{d})$. We will take its square:

$$p = \mathfrak{d}^2 - \mathfrak{h}^2.$$

3. The scalar product of the six-vector $\mathfrak{F} = (\mathfrak{h}, -i\mathfrak{d})$ and its dual vector $\mathfrak{F}^* = (-i\mathfrak{d}, \mathfrak{h})$. We will multiply this product by $i/2$ to obtain the quantity

$$q = (\mathfrak{h} \cdot \mathfrak{d}).$$

4. By multiplying the four-vector P by the six-vector $\mathfrak{F}$ and its dual $\mathfrak{F}^*$ one finds two new four-vectors

$$\begin{aligned} \mathsf{A} &= \mathsf{P} \cdot \mathfrak{F} = ((\rho \cdot \mathfrak{d} + [\mathfrak{v} \cdot \mathfrak{h}]), \quad -i \cdot (\mathfrak{v} \cdot \mathfrak{d})), \\ \mathsf{B} &= \mathsf{P} \cdot \mathfrak{F}^* = (i(\rho \cdot \mathfrak{h} - [\mathfrak{v} \cdot \mathfrak{d}]), \quad (\mathfrak{v} \cdot \mathfrak{h})). \end{aligned}$$

The square of their absolute values are:

$$\begin{aligned} \mathsf{A}^2 &= (\rho \cdot \mathfrak{d} + [\mathfrak{v} \cdot \mathfrak{h}])^2 - (\mathfrak{v} \cdot \mathfrak{d})^2, \\ \mathsf{B}^2 &= -(\rho \cdot \mathfrak{h} - [\mathfrak{v} \cdot \mathfrak{d}])^2 + (\mathfrak{v} \cdot \mathfrak{h})^2. \end{aligned}$$

[531] | These two quantities are no longer independent of each other, for we can easily see that:

$$\mathsf{A}^2 + \mathsf{B}^2 = (\mathfrak{h}^2 - \mathfrak{d}^2) \cdot (\mathfrak{v}^2 - \rho^2) = \sigma^2 \cdot p.$$

In the same way the scalar product of the two yields nothing new:

$$\begin{aligned} (\mathsf{A} \cdot \mathsf{B}) &= i((\rho \cdot \mathfrak{d} + [\mathfrak{v} \cdot \mathfrak{h}])(\rho \cdot \mathfrak{h} - [\mathfrak{v} \cdot \mathfrak{d}]) - (\mathfrak{v} \cdot \mathfrak{d}) \cdot (\mathfrak{v} \cdot \mathfrak{h})) \\ &= -i \cdot (\mathfrak{h} \cdot \mathfrak{d}) \cdot (\mathfrak{v}^2 - \rho^2) = i \cdot \sigma^2 \cdot q. \end{aligned}$$

So we get only *one* fourth scalar, for which we will choose the quantity $s = -\mathsf{B}^2$:

$$s = (\rho \cdot \mathfrak{h} - [\mathfrak{v} \cdot \mathfrak{d}])^2 - (\mathfrak{v} \cdot \mathfrak{h})^2.$$

From the theory of four-dimensional vectors one can prove that there can be no further independent scalars, but I will omit the proof here.

Accordingly we have found as possibilities four independent variables,

$$\begin{cases} \sigma = \sqrt{\rho^2 - \mathfrak{v}^2} = \rho \cdot \sqrt{1 - \dfrac{\mathfrak{v}^2}{\rho^2}}, \\ p = \mathfrak{d}^2 - \mathfrak{h}^2, \\ q = (\mathfrak{d} \cdot \mathfrak{h}), \\ s = (\rho \cdot \mathfrak{h} - [\mathfrak{v} \cdot \mathfrak{d}])^2 - (\mathfrak{v} \cdot \mathfrak{h})^2. \end{cases} \tag{24}$$

12. The intensive quantities $\mathfrak{e}, \varphi, \mathfrak{b}, \mathfrak{f}$ are calculated as follows:

$$\left\{\begin{aligned}
\mathfrak{e} &= 2\cdot\frac{\partial H}{\partial p}\cdot\mathfrak{d}+\frac{\partial H}{\partial q}\cdot\mathfrak{h}+2\cdot\frac{\partial H}{\partial s}\cdot[\mathfrak{v}\cdot(\rho\mathfrak{h}-[\mathfrak{v}\cdot\mathfrak{d}])],\\
\varphi &= -\frac{\partial H}{\partial\sigma}\cdot\frac{\rho}{\sigma}-2\cdot\frac{\partial H}{\partial s}\cdot(\rho\cdot\mathfrak{h}-[\mathfrak{v}\cdot\mathfrak{d}])\cdot\mathfrak{h},\\
\mathfrak{b} &= 2\cdot\frac{\partial H}{\partial p}\cdot\mathfrak{h}-\frac{\partial H}{\partial q}\cdot\mathfrak{d}-2\cdot\frac{\partial H}{\partial s}\cdot(\rho\cdot(\rho\cdot\mathfrak{h}-[\mathfrak{v}\cdot\mathfrak{d}])-\mathfrak{v}\cdot(\mathfrak{v}\cdot\mathfrak{h})),\\
\mathfrak{f} &= \frac{\partial H}{\partial\sigma}\cdot\frac{\mathfrak{v}}{\sigma}-2\cdot\frac{\partial H}{\partial s}\cdot([\mathfrak{d}\cdot(\rho\mathfrak{h}-[\mathfrak{v}\cdot\mathfrak{d}])]+\mathfrak{h}\cdot(\mathfrak{v}\cdot\mathfrak{h})).
\end{aligned}\right. \tag{25}$$

By noting that:

$$(\mathfrak{v}\cdot\mathfrak{h}) = \frac{1}{\rho}\cdot(\mathfrak{v}\cdot(\rho\mathfrak{h}-[\mathfrak{v}\cdot\mathfrak{d}])),$$

| we recognize immediately that the factor $\partial H/\partial s$ vanishes in the four expressions [532]
(25) if:

$$\rho\mathfrak{h}-[\mathfrak{v}\cdot\mathfrak{d}] = 0.$$

If we assume further that $\mathfrak{b} = 0$ in the field of an electron at rest, then $\partial H/\partial q$ must contain either the factor q or the factor s, because otherwise it would not vanish for $\mathfrak{v} = 0$, $\mathfrak{h} = 0$; but now

$$q = (\mathfrak{d}\cdot\mathfrak{h}) = \frac{1}{\rho}\cdot(\mathfrak{d}\cdot(\rho\mathfrak{h}-[\mathfrak{v}\mathfrak{d}])).$$

Accordingly $\partial H/\partial q$ vanishes under the same conditions as the factor $\partial H/\partial s$, namely if:

$$\rho\cdot\mathfrak{h}-[\mathfrak{v}\cdot\mathfrak{d}] = 0.$$

But now one can obtain the quantity $\rho'\cdot\mathfrak{h}' = \rho\cdot\mathfrak{h}-[\mathfrak{v}\cdot\mathfrak{d}]$ by applying a Lorentz transformation to the states of the aether for which one of the coordinate systems moves with respect to the other with a velocity $\mathfrak{q} = \mathfrak{v}/\rho$. If $\mathfrak{q}$ is constant in space and time one can transform to rest, so that $\mathfrak{h}' = 0$, that is: for a stationary motion the condition just written down is satisfied.

If we assume that in the field of an electron at rest not only $\mathfrak{v}$ *and* $\mathfrak{h}$, *but also* $\mathfrak{b}$ *and* $\mathfrak{f}$ *are everywhere zero, then for stationary motion all terms due to the invariants* q *and* s *drop out of the intensive quantities.*

Since all experiences with electrons and matter in general to date refer only to quasistationary motions, and there is no point in burdening the investigations by keeping quantities that presumably will have no influence on the results, we will in the following make the simplifying assumption, that q and s do not occur in H at all.

13. *Hypothesis. The Hamiltonian function* H *depends only on the two invariants* σ and p.

Then we have the following very simple expressions for the intensive quantities:

$$\begin{cases} \mathfrak{e} \;=\; 2 \cdot \dfrac{\partial H}{\partial p} \cdot \mathfrak{d}, \quad \mathfrak{b} = 2 \cdot \dfrac{\partial H}{\partial p} \cdot \mathfrak{h}, \\ \varphi \;=\; -\dfrac{\partial H}{\partial \sigma} \cdot \dfrac{\rho}{\sigma}, \qquad \mathfrak{f} = -\dfrac{\partial H}{\partial \sigma} \cdot \dfrac{\mathfrak{v}}{\sigma}. \end{cases} \tag{26}$$

[533] | *Each of the intensive vectors* $\mathfrak{e}, \mathfrak{b}, \mathfrak{f}$ *is parallel to its corresponding extensive vector* $\mathfrak{d}, \mathfrak{h}, \mathfrak{v}$, *and in addition they are related by the two proportions*:

$$\mathfrak{f} : \mathfrak{v} \;=\; \varphi : \rho, \quad \mathfrak{b} : \mathfrak{h} = \mathfrak{e} : \mathfrak{d}.$$

From this follows directly the theorem: *The world matrix (16) is symmetric about its diagonal*.

Like H, so also Φ of course depends only on two variables; for these we will take the following two quantities:

$$\begin{cases} \chi \;=\; \sqrt{\varphi^2 - \mathfrak{f}^2}, \\ \eta \;=\; \sqrt{\mathfrak{e}^2 - \mathfrak{b}^2}. \end{cases} \tag{27}$$

If we put

$$\frac{\mathfrak{v}}{\rho} \;=\; \frac{\mathfrak{f}}{\varphi} \;=\; \mathfrak{q},$$

we can also write:

$$\chi \;=\; \varphi \cdot \sqrt{1 - \mathfrak{q}^2}. \tag{27a}$$

Finally let us remark that one can find an interesting interpretation for the quantity:

$$\mathrm{Div}(\mathfrak{f}, i\varphi) \;=\; \mathrm{div}\, \mathfrak{f} + \frac{\partial \varphi}{\partial t}.$$

I will use the abbreviation:

$$-\frac{1}{\sigma} \cdot \frac{\partial H}{\partial \sigma} \;=\; \psi.$$

Then we have:

$$\varphi \;=\; \psi \cdot \rho, \quad \mathfrak{f} \;=\; \psi \cdot \mathfrak{v},$$

therefore:

$$\mathrm{div}\, \mathfrak{f} + \frac{\partial \varphi}{\partial t} \;=\; \psi \cdot \left(\mathrm{div}\, \mathfrak{v} + \frac{\partial \rho}{\partial t}\right) + (\mathfrak{v} \cdot \nabla \psi) + \rho \cdot \frac{\partial \psi}{\partial t}.$$

Now,

$$\mathrm{div}\, \mathfrak{v} + \frac{\partial \rho}{\partial t} \;=\; 0$$

and further we can put: $\mathfrak{v} = \rho \cdot \mathfrak{q}$, where we can interpret $\mathfrak{q}$ as the *velocity* with which the charge is being displaced at that place and time. Then we have:

$$(\mathfrak{v} \cdot \nabla\psi) + \rho \cdot \frac{\partial\psi}{\partial t} = \rho \cdot \left(\frac{\partial\psi}{\partial t} + \frac{\partial\psi}{\partial x} \cdot \mathfrak{q}_x + \frac{\partial\psi}{\partial y} \cdot \mathfrak{q}_y + \frac{\partial\psi}{\partial z} \cdot \mathfrak{q}_z\right).$$

Let us think of the several volume elements having charges as individualized, similar to the way we are used to do it with material volume elements, | and consider ψ as a [534] property of the moving element of charge. Then the time rate of change of ψ is:

$$\frac{D\psi}{Dt} = \frac{\partial\psi}{\partial t} + \frac{\partial\psi}{\partial x} \cdot \mathfrak{q}_x + \frac{\partial\psi}{\partial y} \cdot \mathfrak{q}_y + \frac{\partial\psi}{\partial z} \cdot \mathfrak{q}_z.$$

So we arrive at the equation:

$$\operatorname{div} \mathfrak{f} + \frac{\partial\varphi}{\partial t} = \rho \cdot \frac{D\psi}{Dt}. \tag{28}$$

This last equation is of particular interest in view of a theory of gravitation[8] published recently by Abraham. Namely, in a region where the electric field vanishes the quantities that I denote by $\mathfrak{f}_x, \mathfrak{f}_y, \mathfrak{f}_z, i\varphi$ obey the same equations as the quantities called $\mathfrak{F}_x, \mathfrak{F}_y, \mathfrak{F}_z, \mathfrak{F}_u$ by Abraham, with the only difference that Abraham puts:

$$\operatorname{Div} \mathfrak{F} = -4\pi\gamma \cdot \nu,$$

where γ denotes the gravitational constant, ν the mass density, whereas my vector satisfies the equation just derived:

$$\operatorname{Div}(\mathfrak{f}, i\varphi) = \rho \cdot \frac{D\psi}{Dt}.$$

Thus my ansatz would lead to Abraham's theory of gravitation if one wanted to make the assumption that wherever there is material mass, there is a constant increase of the quantity ψ in time. The flux $\mathfrak{f}$ that therefore streams out of the mass particle would be the gravitational field. But since such an assumption is physically absurd, it is excluded to arrive at a gravitational theory in such a simple way from my ansatz. How this probably has to happen has been indicated in the introduction (pp. 633 and 634 [pp. 512 and 513 in the original]).

In the next chapter I will first need to examine whether the existence of indivisible electrons is compatible with my ansatz.

[…]

8 M. Abraham, *Physik. Zeitschr.* 13, p. 1, 1912.

[1] ## THIRD CHAPTER: FORCE AND INERTIAL MASS[9]

Calculation of the Force Acting on a Mass Particle

25. To calculate the force we use the world matrix (16), written down in I. p. 525. In doing so we presuppose no restrictions concerning the invariants that enter into the world function, but assume quite generally that all four of the variables (24) enumerated in I. p. 531 occur in H. An easy calculation shows that the theorem established on p. 533 under a restrictive assumption is valid quite generally:

The world matrix is symmetric about its diagonal.

Namely, by applying the multiplication rule

$$[[\mathfrak{a}\cdot\mathfrak{b}]\cdot\mathfrak{c}] = (\mathfrak{a}\cdot\mathfrak{c})\cdot\mathfrak{b}-(\mathfrak{b}\cdot\mathfrak{c})\cdot\mathfrak{a}$$

and the formula that results therefrom

$$[[\mathfrak{a}\cdot\mathfrak{b}]\cdot\mathfrak{c}]+[[\mathfrak{b}\cdot\mathfrak{c}]\cdot\mathfrak{a}]+[[\mathfrak{c}\cdot\mathfrak{a}]\cdot\mathfrak{b}] = 0,$$

one easily finds from the general formula (25) in I. p. 531 the following two equations:

$$[\mathfrak{e}\cdot\mathfrak{d}]+[\mathfrak{h}\cdot\mathfrak{b}]+[\mathfrak{f}\cdot\mathfrak{v}] = 0, \tag{54}$$

$$[\mathfrak{e}\cdot\mathfrak{h}]+[\mathfrak{b}\cdot\mathfrak{d}]+(\rho\cdot\mathfrak{f}-\varphi\cdot\mathfrak{v}) = 0, \tag{55}$$

[2] | and hence, by writing out the components of these expressions,

$$\begin{aligned}\mathfrak{e}_x\cdot\mathfrak{d}_y+\mathfrak{h}_x\cdot\mathfrak{b}_y+\mathfrak{f}_x\cdot\mathfrak{v}_y &= \mathfrak{d}_x\cdot\mathfrak{e}_y+\mathfrak{b}_x\cdot\mathfrak{h}_y+\mathfrak{v}_x\cdot\mathfrak{f}_y \quad \text{etc.},\\ \mathfrak{d}_y\cdot\mathfrak{b}_z-\mathfrak{d}_z\cdot\mathfrak{b}_y-\rho\cdot\mathfrak{f}_x &= \mathfrak{e}_y\cdot\mathfrak{h}_z-\mathfrak{e}_z\cdot\mathfrak{h}_y-\varphi\cdot\mathfrak{v}_x \quad \text{etc.}\end{aligned}$$

The theorem is thereby proved.

26. Let us now imagine a material particle, that is either a location of an electric node or a more complicated structure composed of similar singularities, which moves in an electromagnetic field of large extent. Let $\mathfrak{s}$ denote the energy flux that is connected with the progressive motion of the states of the aether, as in I. (5) p. 522. Then we have

$$\begin{cases}\mathfrak{s}_x = \mathfrak{e}_y\cdot\mathfrak{h}_z-\mathfrak{e}_z\cdot\mathfrak{h}_y-\varphi\cdot\mathfrak{v}_x = \mathfrak{d}_y\cdot\mathfrak{b}_z-\mathfrak{d}_z\cdot\mathfrak{b}_y-\rho\cdot\mathfrak{f}_x,\\ \mathfrak{s}_y = \mathfrak{e}_z\cdot\mathfrak{h}_x-\mathfrak{e}_x\cdot\mathfrak{h}_z-\varphi\cdot\mathfrak{v}_y = \mathfrak{d}_z\cdot\mathfrak{b}_x-\mathfrak{d}_x\cdot\mathfrak{b}_z-\rho\cdot\mathfrak{f}_y,\\ \mathfrak{s}_z = \mathfrak{e}_x\cdot\mathfrak{h}_y-\mathfrak{e}_y\cdot\mathfrak{h}_x-\varphi\cdot\mathfrak{v}_z = \mathfrak{d}_x\cdot\mathfrak{b}_y-\mathfrak{d}_y\cdot\mathfrak{b}_x-\rho\cdot\mathfrak{f}_z.\end{cases} \tag{56}$$

---

9 Continuation of the two articles: *Ann. d. Phys.* 37, p. 511, is quoted as I.; *Ann. d. Phys.* 39, p. 1, is cited as II.

We will further define the three-dimensional vectors $\mathfrak{p}_1, \mathfrak{p}_2, \mathfrak{p}_3$ by the following equations:

$$\begin{cases} \Phi - \mathfrak{b}\cdot\mathfrak{h} + \mathfrak{e}_x\cdot\mathfrak{d}_x + \mathfrak{h}_x\cdot\mathfrak{b}_x + \mathfrak{f}_x\cdot\mathfrak{v}_x = \mathfrak{p}_{1x}, \\ \mathfrak{e}_y\cdot\mathfrak{d}_x + \mathfrak{h}_y\cdot\mathfrak{b}_x + \mathfrak{f}_y\cdot\mathfrak{v}_x = \mathfrak{p}_{1y}, \\ \mathfrak{e}_z\cdot\mathfrak{d}_x + \mathfrak{h}_z\cdot\mathfrak{b}_x + \mathfrak{f}_z\cdot\mathfrak{v}_x = \mathfrak{p}_{1z}, \\ \mathfrak{e}_x\cdot\mathfrak{d}_y + \mathfrak{h}_x\cdot\mathfrak{b}_y + \mathfrak{f}_x\cdot\mathfrak{v}_y = \mathfrak{p}_{2x}, \\ \Phi - \mathfrak{b}\cdot\mathfrak{h} + \mathfrak{e}_y\cdot\mathfrak{d}_y + \mathfrak{h}_y\cdot\mathfrak{b}_y + \mathfrak{f}_y\cdot\mathfrak{v}_y = \mathfrak{p}_{2y}, \\ \mathfrak{e}_z\cdot\mathfrak{d}_y + \mathfrak{h}_z\cdot\mathfrak{b}_y + \mathfrak{f}_z\cdot\mathfrak{v}_y = \mathfrak{p}_{2z}, \\ \mathfrak{e}_x\cdot\mathfrak{d}_z + \mathfrak{h}_x\cdot\mathfrak{b}_z + \mathfrak{f}_x\cdot\mathfrak{v}_z = \mathfrak{p}_{3x}, \\ \mathfrak{e}_y\cdot\mathfrak{d}_z + \mathfrak{h}_y\cdot\mathfrak{b}_z + \mathfrak{f}_y\cdot\mathfrak{v}_z = \mathfrak{p}_{3y}, \\ \Phi - \mathfrak{b}\cdot\mathfrak{h} + \mathfrak{e}_z\cdot\mathfrak{d}_z + \mathfrak{h}_z\cdot\mathfrak{b}_z + \mathfrak{f}_z\cdot\mathfrak{v}_z = \mathfrak{p}_{3z}. \end{cases} \tag{57}$$

As we saw in I. on p. 526 eq. (17), the first three rows of the world matrix provide three differential equations, which in consideration of (56) and (57) are to be written as follows:

$$\begin{cases} \dfrac{\partial\mathfrak{s}_x}{\partial t} = \dfrac{\partial\mathfrak{p}_{1x}}{\partial x} + \dfrac{\partial\mathfrak{p}_{2x}}{\partial y} + \dfrac{\partial\mathfrak{p}_{3x}}{\partial z}, \\ \dfrac{\partial\mathfrak{s}_y}{\partial t} = \dfrac{\partial\mathfrak{p}_{1y}}{\partial x} + \dfrac{\partial\mathfrak{p}_{2y}}{\partial y} + \dfrac{\partial\mathfrak{p}_{3y}}{\partial z}, \\ \dfrac{\partial\mathfrak{s}_z}{\partial t} = \dfrac{\partial\mathfrak{p}_{1z}}{\partial x} + \dfrac{\partial\mathfrak{p}_{2z}}{\partial y} + \dfrac{\partial\mathfrak{p}_{3z}}{\partial z}. \end{cases} \tag{58}$$

| Let us now imagine the energy as a fluid, flowing with a certain speed $\mathfrak{q}$. If W is [3]
the density of energy, then $\mathfrak{q}$ is determined by the definition

$$\mathfrak{s} = W\cdot\mathfrak{q}. \tag{59}$$

If we further denote by dM the amount of energy occupying at some moment the volume element $dx\cdot dy\cdot dz = dV$, then $dM = W\cdot dV$ and we can as well write equations (58) as follows:

$$\frac{\partial}{\partial t}(dM\cdot\mathfrak{q}_x) = \left(\frac{\partial\mathfrak{p}_{1x}}{\partial x} + \frac{\partial\mathfrak{p}_{2x}}{\partial y} + \frac{\partial\mathfrak{p}_{3x}}{\partial z}\right)\cdot dV,$$

$$\frac{\partial}{\partial t}(dM\cdot\mathfrak{q}_y) = \left(\frac{\partial\mathfrak{p}_{1y}}{\partial x} + \frac{\partial\mathfrak{p}_{2y}}{\partial y} + \frac{\partial\mathfrak{p}_{3y}}{\partial z}\right)\cdot dV,$$

$$\frac{\partial}{\partial t}(dM\cdot\mathfrak{q}_z) = \left(\frac{\partial\mathfrak{p}_{1z}}{\partial x} + \frac{\partial\mathfrak{p}_{2z}}{\partial y} + \frac{\partial\mathfrak{p}_{3z}}{\partial z}\right)\cdot dV.$$

Let us integrate these equations over a volume V. Let

$$M = \int_V dM$$

be the total energy contained in the volume V at the moment we are considering, let $\bar{\mathfrak{q}}$ be the velocity of the "center of mass" in V, defined by the equation:

$$M \cdot \bar{\mathfrak{q}} = \int_V \mathfrak{q} \cdot dM \tag{60}$$

further, let the surface enclosing the volume V be denoted by S, and let N be the outward pointing normal at a point in S; and finally let $\mathfrak{p}_N$ be a three dimensional vector defined by the equation:

$$\mathfrak{p}_N = \mathfrak{p}_1 \cdot \cos(N, x) + \mathfrak{p}_2 \cdot \cos(N, y) + \mathfrak{p}_3 \cdot \cos(N, z)\ldots \tag{61}$$

So the components of $\mathfrak{p}_N$ are computed as follows:

$$\mathfrak{p}_{Nx} = \mathfrak{p}_{1x} \cdot \cos(N, x) + \mathfrak{p}_{2x} \cdot \cos(N, y) + \mathfrak{p}_{3x} \cdot \cos(N, z) \text{ etc.}$$

Integration over V then yields the following result:

$$\frac{\partial(M \cdot \bar{\mathfrak{q}})}{\partial t} = \int_S \mathfrak{p}_N \cdot dS. \tag{62}$$

[4] | Now let us choose the volume V so that it is infinitely small within the extended field in which the material particle moves, but infinitely large in comparison to the enclosed particle. The latter condition is meant to convey, first that the energy of the singularities that constitute the material particle is as good as completely contained in the volume, so that only a quite vanishingly small fraction of the total particle energy resides outside the surface S; and second that on the surface S the vacuum laws already hold as good as exactly, so that ρ and $\mathfrak{v}$ can be taken to be zero, and $\mathfrak{e} = \mathfrak{d}$, $\mathfrak{b} = \mathfrak{h}$. For this choice of the volume V is $M \cdot \bar{\mathfrak{q}}$ the momentum of the particle, its inertial mass is identical with its energy M, and the right side of equations (62) yields the accelerating force [*bewegende Kraft*] acting on the particle. In view of the second condition, and except for vanishingly small correction terms, $\mathfrak{p}_N$ is identical with the component of the Maxwell stress tensor on the corresponding surface element of S; therefore the result for the accelerating force is a value that is independent of the choice of the volume V, provided that the two conditions mentioned above are satisfied; and the value is perfectly identical with what electron theory would yield for a material particle that would be surrounded by the same electric and magnetic field as the particle under consideration. Exactly as in electron theory the accelerating force does not depend on the specific arrangement of the electric charges and the

electric and magnetic dipoles in the interior of the material particle, as long as the particle's own exterior field is the same; in addition it does not depend on the laws of the cohesive forces that hold the particle together, nor on the laws for the electromagnetic field that take the place of Maxwell's equations in the interior of the particle. Exactly the same theorem, which we here have first encountered for the linear motion $\bar{\mathfrak{q}}$ of the particle, can also be shown directly for its rotational motion. Here the moments of inertia are calculated as in the | usual mechanics, by always putting the [5]
energy in place of the inertial mass. It is essential for the proof that we have, according to (54), $\mathfrak{p}_{1y} = \mathfrak{p}_{2x}, \mathfrak{p}_{1z} = \mathfrak{p}_{3x}, \mathfrak{p}_{2z} = \mathfrak{p}_{3y}$.

The ponderomotive forces that cause linear or rotational motion in a material particle as a whole are calculated from the electric and magnetic field in which the particle is located according to exactly the same rules as in the usual theory of electricity. The existence of a special four-vector $(\mathfrak{v}, i\rho)$ *in the interior of the particle, and the deviation of the laws of the electromagnetic field from Maxwell's equations in the interior of the particle have no perceptible influence on the exterior ponderomotive forces.*

For example, an electron of total charge e, moving with velocity $\mathfrak{q}$ in an electromagnetic field feels the force, according to our theory:

$$\mathfrak{P} = e \cdot (\mathfrak{e} + [\mathfrak{q} \cdot \mathfrak{h}]). \tag{63}$$

This expression agrees exactly with that taken as the basis of electron theory.

By contrast, the effects of forces in the interior of the elementary particles of matter, which may cause delicate changes in the structure of these particles themselves, are something entirely different than the ponderomotive forces of the ordinary theory of relativity. But they cannot be calculated without knowing the world function.

Among the exterior forces acting on the material particle there is also gravity. The theorem just proved implies *that the basic equations of aether dynamics I. (1) to (4), on which we based the theory so far, do not suffice to explain gravity.* Thus the expectation that I expressed at the beginning of my work (I, top of p. 513) has not been fulfilled. In a later chapter we will examine how the basic equations have to be enlarged in order to include gravity as well.

The Inertial Mass of a Material Particle

27. We understand a material particle quite generally to be a small region in the aether where the state | variables take on enormously large values. In the following [6]
we will frequently have to evaluate integrals of some state variables over the whole volume of the particle. This is to be understood as a volume whose exterior boundary is sufficiently distant from the center of the particle that the state variables may be treated as infinitely small. Thus, if the outer boundary of the volume is chosen arbitrarily, only such that the particle is "completely" contained by it, as defined here, then this choice cannot have any appreciable effect on the value of the integral.

When saying that a particle is at rest and unchanging we will mean either that all state variables in the volume occupied by the particle are constant, or that the average value of every state variable at every point in the volume is constant when averaged over a time that is infinitely small as far as the experiment is concerned.

For example, let K be the value of a state variable at a point (x, y, z) of the particle. Further let τ be a time that is infinitely small for the experiment. Then the average value we are talking about is

$$\overline{K} = \frac{1}{\tau} \cdot \int_0^\tau K \cdot dt.$$

It is well known[10] that the equations

$$\overline{\frac{\partial K}{\partial t}} = \frac{\partial \overline{K}}{\partial t}, \quad \overline{\frac{\partial K}{\partial x}} = \frac{\partial \overline{K}}{\partial x}, \text{etc.}$$

are valid here.

Therefore the conditions that the particle be unchanging and at rest are:

$$\frac{\partial \bar{\mathfrak{d}}}{\partial t} = 0, \quad \frac{\partial \bar{\mathfrak{h}}}{\partial t} = 0, \quad \frac{\partial \bar{\rho}}{\partial t} = 0, \quad \frac{\partial \bar{\mathfrak{v}}}{\partial t} = 0, \text{etc.}$$

Now the basic equations I (1) to (4) entail the two relations:

$$\mathfrak{e} \cdot \mathfrak{d} - \varphi \cdot \rho = -\operatorname{div}(\varphi \cdot \mathfrak{d}) - \mathfrak{d} \cdot \frac{\partial \mathfrak{f}}{\partial t},$$

$$\mathfrak{b} \cdot \mathfrak{h} - \mathfrak{f} \cdot \mathfrak{v} = -\operatorname{div}(\mathfrak{h} \cdot \mathfrak{f}) - \mathfrak{f} \cdot \frac{\partial \mathfrak{d}}{\partial t}.$$

[7] | Inside a particle at rest we therefore have:

$$\overline{\mathfrak{e} \cdot \mathfrak{d}} - \overline{\varphi \cdot \rho} = -\operatorname{div}(\overline{\varphi \cdot \mathfrak{d}}),$$

$$\overline{\mathfrak{b} \cdot \mathfrak{h}} - \overline{\mathfrak{f} \cdot \mathfrak{v}} = -\operatorname{div}(\overline{\mathfrak{h} \cdot \mathfrak{f}}).$$

If we now integrate over a volume that completely encloses the particle, and note that we may set $\overline{\varphi \cdot \mathfrak{d}}$ and $\overline{\mathfrak{h} \cdot \mathfrak{f}}$ equal to zero on the surface of the volume, we find *for a material particle at rest*:

$$\int \overline{\mathfrak{e} \cdot \mathfrak{d}} \cdot dV = \int \overline{\rho \cdot \varphi} \cdot dV, \tag{64}$$

10 H. A. Lorentz, *Versuch einer Theorie der elektrischen und optischen Erscheinungen in bewegte Körpern*, p.13

$$\int\overline{\mathfrak{b}\cdot\mathfrak{h}}\cdot dV = \int\overline{\mathfrak{f}\cdot\mathfrak{v}}\cdot dV. \tag{65}$$

According to I (7) and (14) on p. 523 and p. 525 the energy density is calculated to be:

$$W = H + \mathfrak{b}\cdot\mathfrak{h} - \mathfrak{f}\cdot\mathfrak{v} = \Phi + \mathfrak{e}\cdot\mathfrak{d} - \varphi\cdot\rho.$$

So the result for *the energy* E_0 *of a material particle at rest* is according to (64) and (65):

$$E_0 = \int\overline{H}\cdot dV = \int\overline{\Phi}\cdot dV. \tag{66}$$

Let S be some unbounded[11] surface that cuts through to particle, and let N be the surface normal at some point. Since on either side of the surface there must occur no permanent energy changes *provided the particle is at rest*, we must have

$$\int_S \bar{\mathfrak{s}}_N\cdot dS = 0, \tag{67}$$

where $\bar{\mathfrak{s}}_N$ is the average value of the component of the vector $\mathfrak{s}$ normal to S. According to (56) this vector is given by

$$\mathfrak{s} = [\mathfrak{e}\cdot\mathfrak{h}] - \varphi\cdot\mathfrak{v} = [\mathfrak{d}\cdot\mathfrak{b}] - \rho\cdot\mathfrak{f}.$$

Laue[12] has shown that as long as equation (67) is valid—and this is the case for any material particle—the following theorem is also valid:

Theorem of Laue. *The integral of each component of the world matrix over the volume of a static material | particle is zero, except for only the component with the index 4,4, which yields the energy of the particle.* [8]

In general we here have to take the average of each component over a short time, as in equation (67).

As M. Laue has shown, this theorem can be used to calculate the energy of a moving particle. I will carry out this calculation for the theory being advanced here. Let all the field quantities at a point x_0, y_0, z_0 of the static particle be characterized by the index 0. From these, according to the theory of relativity, one can find the values at a point x, y, z of a moving particle, having speed q in the direction of the z-axis, and if this point (x, y, z) has at time t the position given by the following equations:

$$x = x_0, \quad y = y_0, \quad \frac{z - q\cdot t}{\sqrt{1-q^2}} = z_0$$

according to the following transformation formulas:

11 i.e. either closed or extending to infinity

12 M. Laue, *Das Relativitätsprinzip*, p. 168 ff.

$$\begin{aligned}
\mathfrak{d}_x &= \frac{\mathfrak{d}_{x0} + q \cdot \mathfrak{h}_{y0}}{\sqrt{1-q^2}}, \quad \mathfrak{d}_y = \frac{\mathfrak{d}_{y0} - q \cdot \mathfrak{h}_{x0}}{\sqrt{1-q^2}}, \quad \mathfrak{d}_z = \mathfrak{d}_{z0}, \\
\mathfrak{h}_x &= \frac{\mathfrak{h}_{x0} - q \cdot \mathfrak{d}_{y0}}{\sqrt{1-q^2}}, \quad \mathfrak{h}_y = \frac{\mathfrak{h}_{y0} + q \cdot \mathfrak{d}_{x0}}{\sqrt{1-q^2}}, \quad \mathfrak{h}_z = \mathfrak{h}_{z0}, \\
\rho &= \frac{\rho_0 + q \cdot \mathfrak{v}_{z0}}{\sqrt{1-q^2}}, \quad \mathfrak{v}_x = \mathfrak{v}_{x0}, \quad \mathfrak{v}_y = \mathfrak{v}_{y0}, \quad \mathfrak{v}_z = \frac{\mathfrak{v}_{z0} + q \cdot \rho_0}{\sqrt{1-q^2}}.
\end{aligned}$$

Exactly the same relations as those between $(\mathfrak{d}, \mathfrak{h})$ and $(\mathfrak{d}_0, \mathfrak{h}_0)$ also hold between $(\mathfrak{e}, \mathfrak{b})$ and $(\mathfrak{e}_0, \mathfrak{b}_0)$, and the same as those between $(\rho, \mathfrak{v})$ and $(\rho_0, \mathfrak{v}_0)$ also hold between $(\varphi, \mathfrak{f})$ and $(\varphi_0, \mathfrak{f}_0)$.

The application of these formulas leads through some quite elementary calculations to the following equation:

$$\begin{aligned}
\mathfrak{b} \cdot \mathfrak{h} - \mathfrak{f} \cdot \mathfrak{v} = {} & \mathfrak{b}_0 \cdot \mathfrak{h}_0 - \mathfrak{f}_0 \cdot \mathfrak{v}_0 + \frac{q^2}{1-q^2} \cdot (\mathfrak{e}_0 \mathfrak{d}_0 - \varphi_0 \cdot \rho_0) \\
& - \frac{q^2}{1-q^2} \cdot (\mathfrak{e}_{z0} \cdot \mathfrak{d}_{z0} - \mathfrak{b}_{x0} \cdot \mathfrak{h}_{x0} - \mathfrak{b}_{y0} \mathfrak{h}_{y0} + \mathfrak{f}_{z0} \cdot \mathfrak{v}_{z0}) \\
& - \frac{q^2}{1-q^2} \cdot ([\mathfrak{e}_0 \cdot \mathfrak{h}_0]_z - \varphi_0 \cdot \mathfrak{v}_{z0} + [\mathfrak{d}_0 \cdot \mathfrak{b}_0]_z - \rho_0 \cdot \mathfrak{f}_{z0}).
\end{aligned}$$

Now we form the time average and integrate over the volume occupied by the
[9] material particle. By applying the equations (64), (65), (67) | and noting the relation, in consequence of the definition of the point x, y, z:

$$dx \cdot dy \cdot dz = \sqrt{1-q^2} \cdot dx_0 \cdot dy_0 \cdot dz_0$$

or

$$dV = \sqrt{1-q^2} \cdot dV_0,$$

we reach the result:

$$\left\{ \begin{aligned}
\int (\overline{\mathfrak{b} \cdot \mathfrak{h}} - \overline{\mathfrak{f} \cdot \mathfrak{v}}) \cdot dV = {} & \frac{q^2}{1-q^2} \\
& \cdot \int (\overline{\mathfrak{b}_0 \cdot \mathfrak{h}_0} - \overline{\mathfrak{e}_{z0} \cdot \mathfrak{d}_{z0}} - \overline{\mathfrak{b}_{z0} \cdot \mathfrak{h}_{z0}} - \overline{\mathfrak{f}_{z0} \cdot \mathfrak{v}_{z0}}) \cdot dV.
\end{aligned} \right. \tag{68}$$

If we denote the value of the quantity H at the point x_0, y_0, z_0 of the static particle by H_0, we can regard

$$H_0 = F(x_0, y_0, z_0)$$

as a function of (x_0, y_0, z_0). Further, let (x, y, z) be the point of the moving particle that is obtained at time t by the Lorentz transformation from (x_0, y_0, z_0). Since H is an invariant under Lorentz transformations, its value at the point (x, y, z) of the moving particle at time t is to be calculated as:

$$H = F\left(x, y, \frac{z - qt}{\sqrt{1 - q^2}}\right),$$

where F denotes exactly the same function as above. From this it follows that:

$$\int H \cdot dV = \sqrt{1 - q^2} \cdot \int H_0 \cdot dV_0 = \sqrt{1 - q^2} \cdot E_0. \tag{69}$$

Now the energy E of the moving particle results from adding (68) and (69):

$$E = \int(\overline{H} + \overline{\mathfrak{b} \cdot \mathfrak{h}} - \overline{\mathfrak{f} \cdot \mathfrak{v}}) \cdot dV,$$

$$E = \sqrt{1 - q^2} \cdot \int \overline{H}_0 \cdot dV_0 + \frac{q^2}{\sqrt{1 - q^2}}$$

$$\cdot \int(\overline{\mathfrak{b}_0 \cdot \mathfrak{h}_0} - \overline{\mathfrak{e}_{z0} \cdot \mathfrak{d}_{z0}} - \overline{\mathfrak{v}_{z0} \cdot \mathfrak{h}_{z0}} - \overline{\mathfrak{f}_{z0} \cdot \mathfrak{v}_{z0}}) \cdot dV_0.$$

This result can be simplified further with the aid of Laue's theorem. Namely, if we apply this theorem | to the term of the world matrix (16) with index 3,3 we get: [10]

$$\int(\overline{\Phi}_0 - \overline{\mathfrak{b}_0 \cdot \mathfrak{h}_0} + \overline{\mathfrak{e}_{z0} \cdot \mathfrak{d}_{z0}} + \overline{\mathfrak{h}_{z0} \cdot \mathfrak{b}_{z0}} + \overline{\mathfrak{f}_{z0} \cdot \mathfrak{v}_{z0}}) \cdot dV_0 = 0.$$

Thus, since according to (66):

$$\int \overline{\Phi}_0 \cdot dV_0 = \int \overline{H}_0 \cdot dV_0 = E_0,$$

$$\int(\overline{\mathfrak{b}_0 \cdot \mathfrak{h}_0} - \overline{\mathfrak{e}_{z0} \cdot \mathfrak{d}_{z0}} - \overline{\mathfrak{h}_{z0} \cdot \mathfrak{b}_{z0}} - \overline{\mathfrak{f}_{z0} \cdot \mathfrak{v}_{z0}}) \cdot dV_0 = \int \overline{\Phi}_0 \cdot dV_0 = E_0.$$

The result is what M. Laue has already shown in general (*Das Relativitätsprinzip* p. 170):

$$E = \frac{E_0}{\sqrt{1 - q^2}}. \tag{70}$$

28. Another interesting consequence can be derived from Laue's theorem. By applying it to the three terms of the diagonal of the world matrix with the indices 1,1, as well as 2,2 and 3,3 one obtains:

$$\int(\overline{\mathfrak{b}_0 \cdot \mathfrak{h}_0} - \overline{\mathfrak{b}_{x0} \cdot \mathfrak{h}_{x0}} - \overline{\mathfrak{e}_{x0} \cdot \mathfrak{d}_{x0}} - \overline{\mathfrak{f}_{x0} \cdot \mathfrak{v}_{x0}}) \cdot dV_0 = E_0,$$

$$\int(\overline{\mathfrak{b}_0 \cdot \mathfrak{h}_0} - \overline{\mathfrak{b}_{y0} \cdot \mathfrak{h}_{y0}} - \overline{\mathfrak{e}_{y0} \cdot \mathfrak{d}_{y0}} - \overline{\mathfrak{f}_{y0} \cdot \mathfrak{v}_{y0}}) \cdot dV_0 = E_0,$$

$$\int(\overline{\mathfrak{b}_0\cdot\mathfrak{h}_0}-\overline{\mathfrak{b}_{z0}\cdot\mathfrak{h}_{z0}}-\overline{\mathfrak{e}_{z0}\cdot\mathfrak{d}_{z0}}-\overline{\mathfrak{f}_{z0}\cdot\mathfrak{v}_{z0}})\cdot dV_0 \;=\; E_0.$$

By addition of these equations one gets:

$$\int(2\cdot\overline{\mathfrak{b}_0\mathfrak{h}_0}-\overline{\mathfrak{e}_0\cdot\mathfrak{d}_0}-\overline{\mathfrak{f}_0\cdot\mathfrak{v}_0})\cdot dV_0 \;=\; 3\cdot E_0,$$

or, taking into account (64) and (65):

$$\begin{cases} E_0 \;=\; -\frac{1}{3}\cdot\int(\overline{\mathfrak{e}_0\cdot\mathfrak{d}_0}-\overline{\mathfrak{b}_0\cdot\mathfrak{h}_0})\cdot dV_0, \\ \quad\;\; =\; -\frac{1}{3}\cdot\int(\overline{\rho_0\cdot\varphi_0}-\overline{\mathfrak{f}_0\cdot\mathfrak{v}_0})\cdot dV_0. \end{cases} \tag{71}$$

In addition it is immediately seen from the three equations just written down that:

$$\begin{cases} \quad\int(\overline{\mathfrak{e}_{x0}\cdot\mathfrak{d}_{x0}}+\overline{\mathfrak{b}_{x0}\cdot\mathfrak{h}_{x0}}+\overline{\mathfrak{f}_{x0}\cdot\mathfrak{v}_{x0}})\cdot dV_0 \\ =\int(\overline{\mathfrak{e}_{y0}\cdot\mathfrak{d}_{y0}}+\overline{\mathfrak{b}_{y0}\cdot\mathfrak{h}_{y0}}+\overline{\mathfrak{f}_{y0}\cdot\mathfrak{v}_{y0}})\cdot dV_0 \\ =\int(\overline{\mathfrak{e}_{z0}\cdot\mathfrak{d}_{z0}}+\overline{\mathfrak{b}_{z0}\cdot\mathfrak{h}_{z0}}+\overline{\mathfrak{f}_{z0}\cdot\mathfrak{v}_{z0}})\cdot dV_0 \\ =\frac{1}{3}\int(\overline{\mathfrak{e}_0\cdot\mathfrak{d}_0}+\overline{\mathfrak{b}_0\cdot\mathfrak{h}_0}+\overline{\mathfrak{f}_0\cdot\mathfrak{v}_0})\cdot dV_0. \end{cases} \tag{72}$$

[11] | These equations become particularly interesting when $\mathfrak{h} = 0, \mathfrak{v} = 0,$ as is the case for an electron.

In the field of an electron we have:

$$E_0 \;=\; -\frac{1}{3}\cdot\int\mathfrak{e}_0\cdot\mathfrak{d}_0\cdot dV \;=\; -\frac{1}{3}\cdot\int\varphi_0\cdot\rho_0\cdot dV_0 \tag{73}$$

and besides:

$$\begin{cases} \int\mathfrak{e}_{x0}\cdot\mathfrak{d}_{x0}\cdot dV_0 \;=\int\mathfrak{e}_{y0}\cdot\mathfrak{d}_{y0}\cdot dV_0 \;=\int\mathfrak{e}_{z0}\cdot\mathfrak{d}_{z0}\cdot dV_0 \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad =\;\frac{1}{3}\int\mathfrak{e}_0\cdot\mathfrak{d}_0\cdot dV_0. \end{cases} \tag{74}$$

29. For the special case discussed thoroughly in II. on pp. 18 ff. the relation (73) can easily be verified. When we substitute into the world function

$$\Phi \;=\; -\frac{1}{2}\eta^2-\frac{1}{6}a\cdot\chi^6$$

the values for the static field, $\eta = \mathfrak{e}_0, \chi = \varphi_0,$ the result is:

$$E_0 = \int \Phi_0 \cdot dV_0 = -\frac{1}{2} \cdot \int \mathfrak{e}_0^2 \cdot dV_0 + \frac{1}{6} a \cdot \int \varphi_0^6 \cdot dV_0 .$$

But furthermore:

$$\mathfrak{d}_0 = -\frac{\partial \Phi_0}{\partial \mathfrak{e}_0} = \mathfrak{e}_0, \quad \rho_0 = \frac{\partial \Phi_0}{\partial \varphi_0} = a \cdot \varphi_0^5$$

and so we may write:

$$E_0 = -\frac{1}{2} \cdot \int \mathfrak{e}_0 \cdot \mathfrak{d}_0 \cdot dV + \frac{1}{6} \cdot \int \varphi_0 \cdot \rho_0 \cdot dV .$$

When (64) is applied to this, the result is (73).

Had we given the wave function the more general form:

$$\Phi = -\frac{1}{2}\eta^2 + \frac{1}{\nu} \cdot a \cdot \chi^\nu ,$$

quite an analogous calculation would yield:

$$E_0 = -\frac{1}{2} \cdot \int \mathfrak{e}_0 \cdot \mathfrak{d}_0 \cdot dV + \frac{1}{\nu} \cdot \int \varphi_0 \cdot \rho_0 \cdot dV ,$$

so that relation (73) could not possibly be satisfied except when $\nu = 6$. This implies:

For all wave functions of the form:

$$\Phi = -\frac{1}{2}\eta^2 + \frac{1}{\nu} \cdot a \cdot \chi^\nu$$

| *only the case* $\nu = 6$ *can lead to isolated nodes of electric charge.* [12]

If one takes some different value for ν, then all integrals of equation (34) in II. p. 15 must have essential singularities, either a singularity at the origin, or at infinity, or both. Then there is no single integral that could represent an electron.

From this one sees that equation (73) can be used on occasion as a criterion whether or not a particular form of the wave function is consistent with the existence of isolated nodes (electrons).

30. From formula (73) it follows that the energy of a node is *negative* in the example discussed in II. So in this case the negative energy attributed to the cohesive effect of the charges exceeds the positive energy of the electric field. Since in the Hamiltonian function:

$$H(\mathfrak{d}_0, 0, \rho_0, 0) = \Phi_0 + \mathfrak{e}_0 \mathfrak{d}_0 - \varphi_0 \cdot \rho_0 = W$$

$$= \frac{1}{2}\mathfrak{d}^{02} - \frac{5}{6}\sqrt[5]{\frac{\rho_0^6}{a}}$$

$\mathfrak{d}_0$ and ρ_0 occur quite separate from each other, the two amounts of energy can also be calculated separately. For the energy of the electric field one gets:

$$\frac{1}{2} \cdot \int \mathfrak{d}_0^2 \cdot dV_0 = \frac{1}{2} \cdot \int \mathfrak{e}_0 \cdot \mathfrak{d}_0 dV_0,$$

and for the energy of the cohesive forces:

$$-\frac{5}{6} \cdot \int \sqrt[5]{\frac{\rho_0^6}{a}} \cdot dV_0 = -\frac{5}{6} \cdot \int \rho_0 \cdot \varphi_0 \cdot dV_0 = -\frac{5}{6} \cdot \int \mathfrak{e}_0 \cdot \mathfrak{d}_0 \cdot dV_0.$$

But if now the energy of a particle is negative, the same must be true for its inertial mass. The nodes mentioned in II. on p. 37 thus have a negative inertial mass; in fields of force they must accordingly assume accelerations that are exactly opposite to the accelerating forces. This explains the behavior that at first seems absurd, to which we were led in II. p. 38 by general reasoning, namely that equal nodes tend to
[13] congregate, and opposite | nodes tend to separate, although the ponderomotive forces of the electric fields act in precisely the opposite direction.

Another general conclusion can be drawn from (73):

The necessary and sufficient condition that the inertial mass of an electron be positive is:

$$\int \mathfrak{e}_0 \cdot \mathfrak{d}_0 \cdot dV < 0$$

or equally well:

$$\int \varphi_0 \cdot \rho_0 \cdot dV < 0.$$

At a large distance from the electron we have $\mathfrak{e} = \mathfrak{d}$, so that $\mathfrak{e}_0 \cdot \mathfrak{d}_0$ is certainly positive. This implies:

In the interior of the electron the two vectors $\mathfrak{e}$ and $\mathfrak{d}$ must have opposite sign.

It is seen from this that it is quite impossible that Maxwell's equations continue to be valid in the interior of an electron.

Similarly φ being equal to the electric potential, has the same sign as ρ in the outer spheres of the electron. In particular, φ reaches its maximum at the place where $\mathfrak{e}$ crosses zero as it assumes the opposite direction in the interior of the electron. Farther in the interior φ must then decrease sufficiently also eventually to change its sign and make $\varphi_0 \cdot \rho_0$ so large and negative that the volume integral of $\varphi_0 \cdot \rho_0$ must be negative.

In the very interior of the electron φ must attain the opposite sign to ρ.

[…]

FIFTH CHAPTER: GRAVITATION [25]

The Extended Basic Equation of the Dynamics of the Aether

37. We saw on p. 655 [p. 5 in the original] that the assumed cohesive pressure of electric charges together with the electromagnetic field still is not sufficient to explain all actions of force in the world of matter. Gravitation is missing, and we are now forced to enlarge the system of fundamental quantities, into which at first we accepted as few quantities as at all possible (I, p. 634 [p. 513 in the original]), namely only the six-vector $(\mathfrak{h}, -i \cdot \mathfrak{d})$ and the four-vector $(\mathfrak{v}, i\rho)$.

It would be most straightforward to conceive of gravity as a cohesive action that resides in the energy itself. But if we want to maintain the validity of the principle of relativity, we cannot allow energy by itself to enter into the extended basic equations, for in relativity theory the energy density is the last entry of the world matrix (cf. I, p. 643, equation (16) [p. 525 in the original]), so the whole matrix as such would have to appear in the equations. One runs into insuperable difficulties if one tries to connect this matrix with some other four-dimensional quantity by means of four-dimensional differential operators, and thus to obtain equations that obey both the causality principle (I. p. 635 [p. 514 in the original]) and the energy principle (I. p. 640 [p. 521 in the original). | I have struggled for a long time with such attempts, [26]
which always led to quite cumbersome systems of equations, and I am convinced that it is quite impossible to attain in this way a theory of gravitation that obeys both the relativity principle and the energy principle.

By contrast it is extraordinarily easy and simple to reach the goal if the cohesive tendency is ascribed not to the quantity W, but to the quantity H, which is defined as $H = W - \mathfrak{b} \cdot \mathfrak{h} + \mathfrak{v} \cdot \mathfrak{f}$ by the equation (7) in I. on p. 641 [p. 523 in the original]. As long as the velocities of the material elementary particles are small compared to the speed of light, it will be experimentally undecidable whether W or H control the gravitational effects. To wit, according to equations (69) and (70) we have for a moving massive particle:

$$\int H dV = \sqrt{1-q^2} E_0,$$
$$\int W dV = \frac{1}{\sqrt{1-q^2}} E_0,$$

where the integrals are to be extended over the volume occupied by the particle, and where E_0 denotes the energy of the particle when at rest, and q the ratio of its velocity to the speed of light. So we see that practically there is no appreciable difference between the two integrals.

But the quantity H is a four-dimensional scalar, and to it the differential operator can be applied in only a single way; this produces a four-vector, the gradient of the scalar. Conversely the four-vector can also be associated with a scalar by applying to it the “divergence” operation. By contrast a six-vector cannot be related to a scalar through a four-dimensional differential operation of first order. This implies:

The gravitational field must necessarily be represented by a four-vector, not by a six-vector.

This theorem is however based on the supposition that the gravitating mass is to
[27] be numerically represented by a four-dimensional scalar, | namely the quantity H. It would be different if the density of the gravitating mass were the fourth component of a four-vector, such as the density of electric charge. Then the gravitational field would require a six-vector, similar to the electromagnetic field. But as far as I can see it is impossible to find a four-vector whose fourth component would approximately equal the energy density, and the theories of gravitation that treat the gravitational field in the same way as the electromagnetic field, such as those of O. Heaviside,[13] H. A. Lorentz,[14] R. Gans,[15] therefore either cannot be in accord with the principle of relativity, or the gravitational mass cannot be equal to the inertial mass in these theories.

To establish the equations of the gravitational field we proceed in the same way as we did when setting up the basic electromagnetic equation in I, sections 2. to 5. We assume that for a complete description of the material world we need, in addition to the six-vector $(\mathfrak{h}, -i\mathfrak{d})$ and the four-vector $(\mathfrak{v}, i\rho)$, yet another four-vector $(\mathfrak{g}, iu)$ and a scalar ω. This system of quantities is paralleled by a second, which is completely determined if all the quantities of the first system are given. Of the second system we already know the six-vector $(\mathfrak{b}, -i\mathfrak{e})$ and the four-vector $(\mathfrak{f}, i\varphi)$, which however now depend not only on $(\mathfrak{h}, -i\mathfrak{d})$ and $(\mathfrak{v}, i\rho)$ but also on $(\mathfrak{g}, iu)$ and ω. To this we must further add a four-vector $(\mathfrak{k}, iw)$ and a scalar H, which correspond to $(\mathfrak{g}, iu)$ and ω. The scalar H shall be essentially identical to the quantity defined in I, p. 523. However, like the energy density W, it depends not only on $(\mathfrak{h}, -i\mathfrak{d})$ and $(\mathfrak{v}, i\rho)$ but also on the quantities of the gravitational field, that is $(\mathfrak{g}, iu)$ and ω, and the relation (7) will accordingly have to be subjected to a minor alteration. Now we apply one of the two possible four-dimensional vector operations to $(\mathfrak{g}, iu)$ and ω, and we apply the other operation to $(\mathfrak{k}, iw)$ and H. In this way we obtain the only
[28] possible form for the | laws of gravitation that is in accord with the principle of relativity:

---

13 O. Heaviside, *Electromagnetic Theory* 1, p. 455, 1894.
14 H. A. Lorentz, *Versl. Kon. Ak. Wet.* Amsterdam 8. p. 603, 1900.
15 R. Gans, *Physik. Zeitschr.* 6. p. 803, 1905.

$$\left.\begin{aligned} \mathfrak{g}_x &= \frac{\partial\omega}{\partial x}, \\ \mathfrak{g}_y &= \frac{\partial\omega}{\partial y}, \\ \mathfrak{g}_z &= \frac{\partial\omega}{\partial z}, \\ u &= -\frac{\partial\omega}{\partial t}, \end{aligned}\right\} \tag{85}$$

$$\frac{\partial\mathfrak{k}_x}{\partial x} + \frac{\partial\mathfrak{k}_y}{\partial y} + \frac{\partial\mathfrak{k}_z}{\partial z} + \frac{\partial w}{\partial t} = -\gamma H. \tag{86}$$

Here γ shall denote a universal constant. The equations (85) are equivalent to the following:

$$\left.\begin{aligned} \frac{\partial\mathfrak{g}_x}{\partial t} + \frac{\partial u}{\partial x} &= 0, \\ \frac{\partial\mathfrak{g}_y}{\partial t} + \frac{\partial u}{\partial y} &= 0, \\ \frac{\partial\mathfrak{g}_z}{\partial t} + \frac{\partial u}{\partial z} &= 0, \end{aligned}\right\} \tag{87}$$

$$\frac{\partial\omega}{\partial t} = -u. \tag{88}$$

Equations (86), (87), and (88) together form a system of five mutually independent equations, each containing a first derivative with respect to time of one of the five new state variables. Consequently the causality principle is satisfied.

The complete system of basic equations of the physics of the aether, including the effects of gravitation [Gravitationswirkungen], is given by the equations: (1), (2), (3), (4), (86), (87), (88).

In terms of the symbols of four-dimensional vector analysis the equations (85) to (88) can also be written as follows:

$$\begin{aligned} (\mathfrak{g}, iu) &= \Gamma\rho\alpha\delta\ \omega, \\ \mathrm{Div}(\mathfrak{k}, iw) &= -\gamma \cdot H, \\ \mathfrak{Curl}(\mathfrak{g}, iu) &= 0. \end{aligned}$$

| The system of equations (85) and (86) would formally agree with that upon [29]
which M. Abraham[16] bases his theory of gravitation, if one would put the two vectors $(\mathfrak{g}, iu)$ and $(\mathfrak{k}, iw)$ equal to each other. M. Abraham proceeds in his theory from the

16 M. Abraham, *Physik. Zeitschr.* 13. p. 1. 1912.

presupposition that the density of gravitating mass (which he calls ν) is a four-dimensional scalar, and since moreover he makes use of the theory of relativity in the cited paper, he had to arrive at this system of equations, the only one that the theory of relativity can produce.

38. The primary question is now whether the energy principle is still valid after including equations (86), (87), (88). So we will multiply equation (87) by the components of a three dimensional vector, say $\mathfrak{a}$, similarly equation (86) by a three-dimensional scalar s, and add both equations. The terms containing derivatives with respect to the coordinates are then:

$$\mathfrak{a}_x \cdot \frac{\partial u}{\partial x} + \mathfrak{a}_y \cdot \frac{\partial u}{\partial y} + \mathfrak{a}_z \cdot \frac{\partial u}{\partial z} + s \cdot \left(\frac{\partial \mathfrak{k}_x}{\partial x} + \frac{\partial \mathfrak{k}_y}{\partial y} + \frac{\partial \mathfrak{k}_z}{\partial z} \right).$$

For this expression to represent a divergence, $\mathfrak{a} = \mathfrak{k}$, $s = u$ must hold. Thus we have found for the last part of the energy equation:

$$\operatorname{div}(u\mathfrak{k}) + \mathfrak{k} \cdot \frac{\partial \mathfrak{g}}{\partial t} + u \cdot \frac{\partial w}{\partial t} - \gamma \cdot H \cdot \frac{\partial \omega}{\partial t} = 0,$$

where in the last term $\partial w / \partial t$ was substituted for u, according to equation (88).

So including the effects of gravitation results in the *total energy current* (instead of I, equation (5) on p. 641 [p. 522 in the original]):

$$\mathfrak{s} = [\mathfrak{e} \cdot \mathfrak{h}] - \varphi \cdot \mathfrak{v} + u \cdot \mathfrak{k} \tag{89}$$

and the *total change of the energy density*:

$$dW = \mathfrak{e} \cdot d\mathfrak{d} + \mathfrak{h} \cdot d\mathfrak{b} - \varphi d\rho - \mathfrak{v} \cdot d\mathfrak{f} + \mathfrak{k} \cdot d\mathfrak{g} + u\,dw - \gamma H d\omega. \tag{90}$$

The function H must now be defined by the following equation, instead of equation (7) of I on p. 641, [p. 523 in the original]:

$$W = H + \mathfrak{h} \cdot \mathfrak{b} - \mathfrak{v} \cdot \mathfrak{f} + uw. \tag{91}$$

[30] | It then follows from (90):

$$dH = \mathfrak{e} \cdot d\mathfrak{d} - \mathfrak{b} \cdot d\mathfrak{h} - \varphi d\rho + \mathfrak{f} \cdot d\mathfrak{v} + \mathfrak{k} \cdot d\mathfrak{g} - w\,du - \gamma H d\omega. \tag{92}$$

Because H is a function of the following variables: $(\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v}, \mathfrak{g}, u, \omega)$, we have:

$$\mathfrak{e} = \frac{\partial H}{\partial \mathfrak{d}},\quad \mathfrak{b} = -\frac{\partial H}{\partial \mathfrak{h}},\quad \varphi = -\frac{\partial H}{\partial \rho},\quad \mathfrak{f} = \frac{\partial H}{\partial \mathfrak{v}},\quad \mathfrak{k} = \frac{\partial H}{\partial \mathfrak{g}},\quad w = -\frac{\partial H}{\partial u},\quad \frac{\partial H}{\partial \omega} = -\gamma H. \tag{93}$$

From the last equation of (93) it follows that:

$$H = e^{-\gamma\omega} H'(\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v}, \mathfrak{g}, u). \tag{94}$$

If we now define:

$$\mathfrak{e}' = \frac{\partial H'}{\partial \mathfrak{d}}, \quad \mathfrak{b}' = -\frac{\partial H'}{\partial \mathfrak{h}}, \quad \varphi' = -\frac{\partial H'}{\partial \rho}, \quad \mathfrak{f}' = \frac{\partial H'}{\partial \mathfrak{v}}, \quad \mathfrak{k}' = \frac{\partial H'}{\partial \mathfrak{g}}, \quad w' = -\frac{\partial H'}{\partial u}, \qquad (95)$$

where all the primed quantities depend only on $(\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v}, \mathfrak{g}, u)$ but not on ω, then we have:

$$\mathfrak{e} = e^{-\gamma\omega}\mathfrak{e}', \quad \mathfrak{b} = e^{-\gamma\omega}\mathfrak{b}', \quad \varphi = e^{-\gamma\omega}\varphi', \quad \mathfrak{f} = e^{-\gamma\omega}\mathfrak{f}', \quad \mathfrak{k} = e^{-\gamma\omega}\mathfrak{k}', \quad w = e^{-\gamma\omega}w'. \qquad (96)$$

If equations (93) are satisfied, then the energy principle is valid also for the extended basic equations, and if all variables occur in H only in combinations that are invariant under Lorentz transformations, then the principle of relativity is also valid.

Thus we succeeded in devising a theory of gravitation in which both the energy principle and the principle of relativity are valid.

I want to stress particularly the last, because in the theory of matter here proposed an ansatz contradicting the principle of relativity should be rejected outright. In his papers on gravitation M. Abraham advocates the view[17] that gravitation and | relativ- [31]
ity theory are not compatible with each other. If this were the case one would have to conclude that gravitation is so to speak a purely external force, which plays no part in the existence of matter itself. For if it belonged, as I assume here, to the forces that determine in an essential way the form of the material elementary particles and the whole internal structure of the atoms, and if it did not obey the principle of relativity, then it would be unthinkable that the elementary particles of matter and the action of forces that bind them into atoms, molecules, and tangible bodies should, when the matter moves through space, quite generally be subjected to precisely those changes that lead to the contraction of matter, which was proved by Michelson's experiment. On the other hand, however, I also believe that one would encounter great difficulties if one wanted to treat gravitation as an action that did not play any appreciable role in the internal processes of atoms, and hence I believe that one must abandon M. Abraham's point of view as soon as one treats the theory of gravitation not detached from the theory of matter. Therefore it seems to me very important that gravitation and relativity theory can be joined together in such a simple way as we have just done.

Let me add the remark that in the dynamics of the aether when extended by equations (86), (87), (88), *Hamilton's principle* is valid in the form that we encountered in I, section 10. The proof offers no difficulties whatsoever.

The Invariants

39. The number of invariants is considerably increased by including the gravitational quantities. Besides the gravitational potential ω, four further quantities join the four quantities found in (24) of I. on p. 648 [p. 531 in the original], so the function H' in

17 M. Abraham, *Ann. d. Phys.* 38. p. 1056. 1912.

(94) can possibly depend on *eight* independent variables. These can be taken as the
[32] following combinations of the variables of state: |

$$\left.\begin{aligned}
p &= \mathfrak{d}^2 - \mathfrak{h}^2, \\
q &= (\mathfrak{d}\cdot\mathfrak{h}), \\
\sigma &= \sqrt{\rho^2 - \mathfrak{v}^2}, \\
s &= (\rho\mathfrak{h} - [\mathfrak{v}\cdot\mathfrak{d}])^2 - (\mathfrak{v}\cdot\mathfrak{h})^2, \\
\kappa &= \sqrt{\mathfrak{g}^2 - u^2}, \\
\mathfrak{k} &= (u\mathfrak{h} - [\mathfrak{g}\cdot\mathfrak{d}])^2 - (\mathfrak{g}\cdot\mathfrak{h})^2, \\
h &= (\mathfrak{g}\cdot\mathfrak{v}) - u\rho, \\
\mathfrak{b} &= (\rho\mathfrak{h} - [\mathfrak{v}\cdot\mathfrak{d}])\cdot(u\mathfrak{h} - [\mathfrak{g}\cdot\mathfrak{d}]) - (\mathfrak{v}\cdot\mathfrak{h})\cdot(\mathfrak{g}\cdot\mathfrak{h}).
\end{aligned}\right\} \tag{97}$$

It can be proved by means of four-dimensional vector analysis that all other invariants can be computed from these eight quantities. But I do not want to reproduce that proof here.

Similarly I do not wish to write down here the formulas that now lead to the calculation of the quantities $\mathfrak{e}', \mathfrak{b}', \varphi', \mathfrak{f}', w', \mathfrak{k}'$ from the function H', analogous to the formulae (25) in I, p. 649 [p. 531 in the original], since they can be derived quite easily.

The Differential Equation of the Electron

40. The following quantities are of course also invariants under Lorentz transformations:

$$\begin{aligned}
\mathfrak{e}\cdot\mathfrak{d} - \mathfrak{b}\cdot\mathfrak{h} &= e^{-\gamma\omega}(\mathfrak{e}'\cdot\mathfrak{d} - \mathfrak{b}'\cdot\mathfrak{h}), \\
\varphi\rho - \mathfrak{f}\cdot\mathfrak{v} &= e^{-\gamma\omega}(\varphi'\rho - \mathfrak{f}'\cdot\mathfrak{v}), \\
\mathfrak{k}\cdot\mathfrak{g} - wu &= e^{-\gamma\omega}(\mathfrak{k}'\cdot\mathfrak{g} - w'u).
\end{aligned}$$

For many purposes it is more convenient to use other functions instead of H, which differ from the latter only by an additional term formed from the quantities just written down. Let us define exactly as in I, p. 642 [p. 524 in the original]:

$$\Phi = H - (\mathfrak{e}\cdot\mathfrak{d} - \mathfrak{b}\cdot\mathfrak{h}) + (\varphi\rho - \mathfrak{f}\cdot\mathfrak{v}). \tag{98}$$

We can also set:

$$\left.\begin{aligned}
\Phi &= e^{-\gamma\omega}\Phi', \\
\Phi' &= H' - (\mathfrak{e}'\cdot\mathfrak{d} - \mathfrak{b}'\cdot\mathfrak{h}) + (\varphi'\rho - \mathfrak{f}'\cdot\mathfrak{v}),
\end{aligned}\right\} \tag{99}$$

where accordingly Φ' is a quantity depending only on the variables $\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v}, \mathfrak{g}, u$
[33] and not on ω. Since $(\mathfrak{e}', \mathfrak{b}', \varphi', \mathfrak{f}')$ | can be calculated from the variables

$(\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v}, \mathfrak{g}, u)$, one can also, conversely, calculate $(\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v})$ from $(\mathfrak{e}', \mathfrak{b}', \varphi', \mathfrak{f}', \mathfrak{g}, u)$, and so one may consider Φ' as a function of this new system of variables:

$$\Phi = e^{-\gamma\omega}\Phi'(\mathfrak{e}', \mathfrak{b}', \varphi', \mathfrak{f}', \mathfrak{g}, u). \tag{100}$$

Now it follows from (99) and (95):

$$d\Phi' = -\mathfrak{d}\cdot d\mathfrak{e}' + \mathfrak{h}\cdot d\mathfrak{b}' + \rho d\varphi' - \mathfrak{v}\cdot d\mathfrak{f}' + \mathfrak{k}'\cdot d\mathfrak{g} - w'du,$$

therefore:

$$\mathfrak{d} = -\frac{\partial\Phi'}{\partial\mathfrak{e}'},\quad \mathfrak{h} = \frac{\partial\Phi'}{\partial\mathfrak{b}'},\quad \rho = \frac{\partial\Phi'}{\partial\varphi'},\quad \mathfrak{v} = -\frac{\partial\Phi'}{\partial\mathfrak{f}'},\quad \mathfrak{k}' = \frac{\partial\Phi'}{\partial\mathfrak{g}},\; w' = -\frac{\partial\Phi'}{\partial u}. \tag{101}$$

In the case of an electron at rest the quantities $\mathfrak{b}', \mathfrak{f}', u$ are to be set to constant zero, and the three remaining ones depend only on the distance r from the center. I set:

$$\left.\begin{aligned} X &= \mathfrak{e}' = -e^{+\gamma\omega}\frac{d\varphi}{dr}, \\ Y &= \varphi' = e^{+\gamma\omega}\varphi, \\ Z &= \mathfrak{g} = \frac{d\omega}{dr}. \end{aligned}\right\} \tag{102}$$

Thus we have a function Φ' that depends only on three variables:

$$\Phi'(X, Y, Z).$$

But because also:

$$X = -\frac{dY}{dr} + \gamma YZ,$$

we really have in Φ' only two unknown variables Y and Z and the derivative dY/dr of one of them.

For these two unknown variables Y and Z we then have the following two differential equations:

$$\frac{1}{r^2}\frac{d}{dr}(r^2\mathfrak{d}) = \rho,$$

$$\frac{1}{r^2}\frac{d}{dr}(r^2\mathfrak{k}) = -\gamma H,$$

or:

$$\left.\begin{aligned} &\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{\partial\Phi'}{\partial X}\right)+\frac{\partial\Phi'}{\partial Y} \;=\; 0, \\ &\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{\partial\Phi'}{\partial Z}\right)+\gamma\cdot\left(\Phi'-X\,\frac{\partial\Phi'}{\partial X}-Y\,\frac{\partial\Phi'}{\partial Y}-Z\,\frac{\partial\Phi'}{\partial Z}\right) \;=\; 0. \end{aligned}\right\} \tag{103}$$

[34] | These two equations should replace equation (34) of II, p. 15 when one wants to discuss the problem of the electron with gravitation taken into account. Incidentally, the unknown Z and its derivatives can also be eliminated from both equations according to the usual procedure of differential calculus. This yields an equation of third order for the unknown $Y = e^{\gamma\omega}\varphi$, whereas (34) was an equation of second order for φ.

The World Matrix

41. As before (I, p. 643 [p. 525 in the original], we can use the world function Φ defined in the previous section to construct the world matrix. Namely, from equation (98) and (91):

$$W \;=\; \Phi+\mathfrak{e}\cdot\mathfrak{d}-\varphi\rho+uw. \tag{104}$$

The world matrix can now be constructed according to exactly the same scheme as in equation (16), simply by including the four-vector of gravitation:

$$\left|\begin{array}{c} \Phi-\mathfrak{b}\cdot\mathfrak{h}+\mathfrak{e}_x\mathfrak{d}_x+\mathfrak{h}_x\mathfrak{b}_x+\mathfrak{f}_x\mathfrak{v}_x-\mathfrak{g}_x\mathfrak{k}_x, \\ \mathfrak{e}_x\mathfrak{d}_y+\mathfrak{h}_x\mathfrak{b}_y+\mathfrak{f}_x\mathfrak{v}_y-\mathfrak{g}_x\mathfrak{k}_y, \quad \mathfrak{e}_x\mathfrak{d}_z+\mathfrak{h}_x\mathfrak{b}_z+\mathfrak{f}_x\mathfrak{v}_z-\mathfrak{g}_x\mathfrak{k}_z, \\ -i(\mathfrak{d}_y\mathfrak{b}_z-\mathfrak{d}_z\cdot\mathfrak{b}_y-\mathfrak{f}_x\rho+\mathfrak{g}_xw), \\ \mathfrak{e}_y\mathfrak{d}_x+\mathfrak{h}_y\mathfrak{b}_x+\mathfrak{f}_y\mathfrak{v}_x-\mathfrak{g}_y\mathfrak{k}_x, \\ \Phi-\mathfrak{b}\cdot\mathfrak{h}+\mathfrak{e}_y\mathfrak{d}_y+\mathfrak{h}_y\mathfrak{b}_y+\mathfrak{f}_y\mathfrak{v}_y-\mathfrak{g}_y\mathfrak{k}_y, \\ \mathfrak{e}_y\mathfrak{d}_z+\mathfrak{h}_y\mathfrak{b}_z+\mathfrak{f}_y\mathfrak{v}_z-\mathfrak{g}_y\mathfrak{k}_z, \quad -i(\mathfrak{d}_z\mathfrak{b}_x-\mathfrak{d}_x\mathfrak{b}_z-\mathfrak{f}_y\rho+\mathfrak{g}_yw), \\ \mathfrak{e}_z\mathfrak{d}_x+\mathfrak{h}_z\mathfrak{b}_x+\mathfrak{f}_z\mathfrak{v}_x-\mathfrak{g}_z\mathfrak{k}_x, \quad \mathfrak{e}_z\mathfrak{d}_y+\mathfrak{h}_z\mathfrak{b}_y+\mathfrak{f}_z\mathfrak{v}_y-\mathfrak{g}_z\mathfrak{k}_y, \\ \Phi-\mathfrak{b}\cdot\mathfrak{h}+\mathfrak{e}_z\mathfrak{d}_z+\mathfrak{h}_z\mathfrak{b}_z+\mathfrak{f}_z\mathfrak{v}_z-\mathfrak{g}_z\mathfrak{k}_z, \\ -i(\mathfrak{d}_x\mathfrak{b}_y-\mathfrak{d}_y\mathfrak{b}_x-\mathfrak{f}_z\rho+\mathfrak{g}_zw), \\ -i(\mathfrak{e}_y\mathfrak{h}_z-\mathfrak{e}_z\mathfrak{h}_y-\mathfrak{v}_x\varphi+\mathfrak{k}_xu), \\ -i(\mathfrak{e}_z\mathfrak{h}_x-\mathfrak{e}_y\mathfrak{h}_z-\mathfrak{v}_y\varphi+\mathfrak{k}_yu), \\ -i(\mathfrak{e}_x\mathfrak{h}_y-\mathfrak{e}_y\mathfrak{h}_x-\mathfrak{v}_z\varphi+\mathfrak{k}_zu), \quad \Phi+\mathfrak{e}\cdot\mathfrak{d}-\varphi\rho+uw. \end{array}\right| \tag{105}$$

When the operation Δtv is applied to this matrix, the fourth row yields the energy principle; the first three rows will lead to the equations of motion of a material particle. |

For the most general case that all variables (97) occur in H, and by quite elementary calculational steps, of hardly greater complexity than those mentioned in section 25, one can prove the equations: [35]

$$[\mathfrak{e}\cdot\mathfrak{d}]+[\mathfrak{h}\cdot\mathfrak{b}]+[\mathfrak{f}\cdot\mathfrak{v}]+[\mathfrak{k}\cdot\mathfrak{g}] = 0, \tag{106}$$

$$[\mathfrak{e}\cdot\mathfrak{h}]+[\mathfrak{b}\cdot\mathfrak{d}]+(\rho\mathfrak{f}-\varphi\mathfrak{v})+(u\mathfrak{k}-w\mathfrak{g}) = 0, \tag{107}$$

which are the generalizations of equations (54) and (55), so that

$$\mathfrak{e}_x\mathfrak{d}_y+\mathfrak{h}_x\mathfrak{b}_y+\mathfrak{f}_x\mathfrak{v}_y-\mathfrak{g}_x\mathfrak{k}_y = \mathfrak{d}_x\mathfrak{e}_y+\mathfrak{b}_x\mathfrak{h}_y+\mathfrak{v}_x\mathfrak{f}_y-\mathfrak{k}_x\mathfrak{g}_y \quad \text{etc.}$$
$$\mathfrak{d}_y\mathfrak{b}_z-\mathfrak{d}_z\mathfrak{b}_y-\rho\mathfrak{f}_x+w\mathfrak{g}_x = \mathfrak{e}_y\mathfrak{h}_z-\mathfrak{e}_z\mathfrak{h}_y-\varphi\mathfrak{v}_x+u\mathfrak{k}_x \quad \text{etc.}$$

The world matrix (105) is symmetric across the diagonal.

Calculation of the Force Acting on a Mass Particle

42. To calculate the ponderomotive force on a mass particle we proceed exactly as in 26. We consider a volume V containing the mass particle that on the one hand is large enough that the law of superposition, valid in vacuo, is already satisfied on its surface S; but which is on the other hand small enough that the largely extended field, which causes the force action, can be well approximated as homogeneous in the interior, if the particle is imagined to be absent.

For the actions of gravity, the principle of superposition states that the differential equations (86) and (87) are linear in $\mathfrak{g}$ and u. Therefore this condition must be fulfilled:

In vacuo $(\mathfrak{g}, iu)$ *and* $(\mathfrak{k}, iw)$ *differ only by a constant factor.*

The factor of proportionality depends only on how we define the units.[18] We will make the convention that *in vacuo*:

$$\mathfrak{g} = \mathfrak{k}. \tag{108}$$

| Accordingly the world function Φ must be representable on the surface S of the volume V in the form: [36]

$$\Phi = \frac{1}{2}(\mathfrak{b}^2-\mathfrak{e}^2)+\frac{1}{2}(\mathfrak{g}^2-u^2)+\Phi_1, \tag{109}$$

where Φ_1 can be neglected as vanishingly small, and with an error that decreases as V is taken to be increasingly large.

Let $\mathfrak{s}$ be the energy current, then the value of the momentum of motion $\mathfrak{G}$ contained in V is given by:

18 So that we can disregard the factor $e^{-\gamma\omega}$, we assume that the gravitational potential ω is so small that $e^{-\gamma\omega}$ cannot be distinguished from 1. We will see in 47. that this assumption does not restrict the general validity of the proofs.

$$\mathfrak{G} = \int_V \mathfrak{s}\, dV.$$

The inertial mass M contained in the volume V amounts to:

$$M = \int_V W dV,$$

where W denotes the density of energy. Therefore the velocity $\mathfrak{q}$ of the considered particle is:

$$\mathfrak{q} = \frac{\mathfrak{G}}{M}.$$

Then the force $\mathfrak{P}$ acting on the particle is calculated to be:

$$\mathfrak{P} = \frac{d\mathfrak{G}}{dt} = \int_V \frac{\partial \mathfrak{s}}{\partial t} dV + \int_V (\mathfrak{q} \cdot \nabla)\mathfrak{s}\, dV.$$

Here the first term on the right side signifies the change in time of the momentum of motion in the stationary volume V, in which the particle moves so that $\mathfrak{s}$ changes; and the second term means the change in momentum calculated for a volume V in which the particle remains at rest and $\mathfrak{s}$ is unchanged, but where V is displaced with velocity $\mathfrak{q}$. The two together yield the change in time of $\mathfrak{G}$ in a co-moving volume V rigidly attached to the particle.

We now substitute for $\partial \mathfrak{s}/\partial t$ the values calculated from equation (58) and obtain:

$$\mathfrak{P} = \int_S \mathfrak{p}_N dS + \int_S \mathfrak{s} q dS \cdot (N, \mathfrak{q}).$$

by a single integration (cf. equation (62)). |

[37] Since the principle of superposition is valid on S, the components of $\mathfrak{p}_N$ as well as those of $\mathfrak{s}$ are expressions of second degree in the state variables; that is, $\mathfrak{P}$ is composed additively of an expression that contains only the electromagnetic field quantities (the force of the electromagnetic field), and of an expression that contains only the gravitational quantities $\mathfrak{g}$ and u (the gravitational force acting on the particle). Since we already know the first expression we are interested here only in the second. So we calculate with the expressions:

$$\begin{aligned}
\mathfrak{p}_{1x} &= -\frac{1}{2}(\mathfrak{g}_x^2 - \mathfrak{g}_y^2 - \mathfrak{g}_z^2) - \frac{1}{2}u^2,\\
\mathfrak{p}_{1y} &= -\mathfrak{g}_x\mathfrak{g}_y,\\
\mathfrak{p}_{1z} &= -\mathfrak{g}_x\mathfrak{g}_z,\\
\mathfrak{s}_x &= \mathfrak{g}_x u,
\end{aligned}$$

which result from the matrix (105) if one sets $\mathfrak{k} = \mathfrak{g}$, $u = w$ and the value (109) for Φ, and furthermore omits all terms in $\mathfrak{b}$ and $\mathfrak{e}$.

Further, according to the principle of superposition we can now think of the several components of the variables of state composed additively of two quantities each; one which corresponds to the particle's proper field, to be denoted by the index 0, and a quantity belonging to the largely extended field in which the particle moves, denoted by the index 1:

$$\mathfrak{g} = \mathfrak{g}_0 + \mathfrak{g}_1, \quad u = u_0 + u_1.$$

For the tensor components $\mathfrak{p}$ and the components of $\mathfrak{s}$ we obtain sums of three expressions each, the first of which is composed only of quantities with index 0, the second of quantities with mixed indices, and the third of quantities with index 1:

$$\begin{aligned}
\mathfrak{p}_{1x} &= -\frac{1}{2}(\mathfrak{g}_{x0}^2 - \mathfrak{g}_{y0}^2 - \mathfrak{g}_{z0}^2 + u_0^2) - (\mathfrak{g}_{x0}\mathfrak{g}_{x1} - \mathfrak{g}_{y0}\mathfrak{g}_{y1} - \mathfrak{g}_{z0}\mathfrak{g}_{z1} + u_0u_1)\\
&\qquad - \frac{1}{2}(\mathfrak{g}_{x1}^2 - \mathfrak{g}_{y1}^2 - \mathfrak{g}_{z1}^2 + u_1^2),\\
\mathfrak{p}_{1y} &= -\mathfrak{g}_{x0}\mathfrak{g}_{y0} - (\mathfrak{g}_{x1}\mathfrak{g}_{y0} + \mathfrak{g}_{x0}\mathfrak{g}_{y1}) - \mathfrak{g}_{x1}\mathfrak{g}_{y1},\\
\mathfrak{p}_{1z} &= -\mathfrak{g}_{x0}\mathfrak{g}_{z0} - (\mathfrak{g}_{x1}\mathfrak{g}_{z0} + \mathfrak{g}_{x0}\mathfrak{g}_{z1}) - \mathfrak{g}_{x1}\mathfrak{g}_{z1},\\
\mathfrak{s}_x &= u_0\mathfrak{g}_{x0} + (u_1\mathfrak{g}_{x0} + u_0\mathfrak{g}_{x1}) + u_1\mathfrak{g}_{x1}.
\end{aligned}$$

| Correspondingly, $\mathfrak{P}$ decomposes into three terms as well, which one could [38] denote by $\mathfrak{P}_{00}, \mathfrak{P}_{01}, \mathfrak{P}_{11}$. $\mathfrak{P}_{00}$ would be obtained by annulling the extended field in which the particle is moving ($\mathfrak{g}_1 = 0$, $u_1 = 0$). But because the particle's proper field is in internal equilibrium with itself, the particle can move only with constant velocity in a field-free space ($\mathfrak{g}_1 = 0$, $u_1 = 0$), therefore $\mathfrak{P}_{00} = 0$. In exactly the same way we find $\mathfrak{P}_{11} = 0$. So to calculate $\mathfrak{P}$ there remain only the terms that we have characterized as $\mathfrak{P}_{01}$. For example, the result for the x-component of the force is:

$$\mathfrak{P}_x = -\mathfrak{g}_{x1}\int_S(\mathfrak{g}_{x0}dS_x + \mathfrak{g}_{y0}dS_y + \mathfrak{g}_{z0}dS_z) - u_1\int_S u_0 dS_x$$

$$+\mathfrak{g}_{y1}\int_S(\mathfrak{g}_{y0}dS_x - \mathfrak{g}_{x0}dS_y) + \mathfrak{g}_{z1}\int_S(\mathfrak{g}_{z0}dS_x - \mathfrak{g}_{x0}dS_z)$$

$$+ u_1\int_S \mathfrak{g}_{x0}(\mathfrak{q}_x dS_x + \mathfrak{q}_y dS_y + \mathfrak{q}_z dS_z) + \mathfrak{g}_{x1}\int_S u_0(\mathfrak{q}_x dS_x + \mathfrak{q}_y dS_y + \mathfrak{q}_z dS_z).$$

Here the assumption was used that the extended field $\mathfrak{g}_1, u_1$ may be treated as constant in the interior of the volume V. Further we have set for brevity:

$$dS\cos(N, x) = dS_x, \qquad dS\cos(N, y) = dS_y, \qquad dS\cos(N, z) = dS_z.$$

From the property of the vector $\mathfrak{g}$ that (from equation (87)):

$$\operatorname{curl}\mathfrak{g} = 0,$$

it follows:

$$\int_S(\mathfrak{g}_{y0}dS_x - \mathfrak{g}_{x0}dS_y) = 0,$$

$$\int_S(\mathfrak{g}_{z0}dS_x - \mathfrak{g}_{x0}dS_z) = 0.$$

[39] So this eliminates the third and fourth term in the sum for $\mathfrak{P}_x$ written above. I again | change the remaining terms into volume integrals, making simultaneously multiple use of $\mathfrak{k}_0 = \mathfrak{g}_0$, $w_0 = u_0$ which are valid on S:

$$\mathfrak{P}_x = -\mathfrak{g}_{x1}\int\left(\frac{\partial \mathfrak{k}_{x0}}{\partial x} + \frac{\partial \mathfrak{k}_{y0}}{\partial y} + \frac{\partial \mathfrak{k}_{z0}}{\partial z}\right)dV - u_1\int\frac{\partial u_0}{\partial x}dV$$
$$+ u_1\int\left(\frac{\partial \mathfrak{g}_{x0}}{\partial x}\mathfrak{q}_x + \frac{\partial \mathfrak{g}_{x0}}{\partial y}\mathfrak{q}_y + \frac{\partial \mathfrak{g}_{x0}}{\partial z}\mathfrak{q}_z\right)dV + \mathfrak{g}_{x1}\int\left(\frac{\partial w_0}{\partial x}\mathfrak{q}_x + \frac{\partial w_0}{\partial y}\mathfrak{q}_y + \frac{\partial w_0}{\partial z}\mathfrak{q}_z\right)dV.$$

But we have:

$$\frac{\partial \mathfrak{g}_{x0}}{\partial x}\mathfrak{q}_x + \frac{\partial \mathfrak{g}_{x0}}{\partial y}\mathfrak{q}_y + \frac{\partial \mathfrak{g}_{x0}}{\partial z}\mathfrak{q}_z = \frac{d\mathfrak{g}_{x0}}{dt} - \frac{\partial \mathfrak{g}_{x0}}{\partial t} = \frac{d\mathfrak{g}_{x0}}{dt} + \frac{\partial u_0}{\partial x}$$

and:

$$\frac{\partial w_0}{\partial x}\mathfrak{q}_x + \frac{\partial w_0}{\partial y}\mathfrak{q}_y + \frac{\partial w_0}{\partial z}\mathfrak{q}_z = \frac{dw_0}{dt} - \frac{\partial w_0}{\partial t}.$$

so that the result is:

$$\mathfrak{P}_x = -\mathfrak{g}_{x1}\int\left(\frac{\partial\mathfrak{k}_{x0}}{\partial x}+\frac{\partial\mathfrak{k}_{y0}}{\partial y}+\frac{\partial\mathfrak{k}_{z0}}{\partial z}+\frac{\partial w_0}{\partial t}\right)dV+\int\left(w_1\frac{d\mathfrak{g}_{x0}}{dt}+\mathfrak{g}_{x1}\frac{dw_0}{dt}\right)dV.$$

The second term is vanishingly small compared to the first. For since w_1 and $\mathfrak{g}_{x1}$ are vanishingly small compared to w_0 and $\mathfrak{g}_{x0}$ in the interior of the volume V, that second term is negligible compared to the term:

$$\frac{d}{dt}\int w\mathfrak{g}_x dV,$$

which occurs in another formula for $\mathfrak{P}_x$, namely:

$$\mathfrak{P}_x = \frac{d\mathfrak{G}_x}{dt} = \frac{d}{dt}\int(\mathfrak{d}_y\mathfrak{b}_z-\mathfrak{d}_z\mathfrak{b}_y-\mathfrak{f}_x\rho+\mathfrak{g}_x w)dV$$

and so that term must be negligible compared to the value of $\mathfrak{P}_x$ in general. Thus we have:

$$\mathfrak{P}_x = -\mathfrak{g}_{x1}\int_V\left(\frac{\partial\mathfrak{k}_{x0}}{\partial x}+\frac{\partial\mathfrak{k}_{y0}}{\partial y}+\frac{\partial\mathfrak{k}_{z0}}{\partial z}+\frac{\partial w_0}{\partial t}\right)dV.$$

In the following I again omit the indices and put according to (86):

$$\frac{\partial\mathfrak{k}_{x0}}{\partial x}+\frac{\partial\mathfrak{k}_{y0}}{\partial y}+\frac{\partial\mathfrak{k}_{z0}}{\partial z}+\frac{\partial w_0}{\partial t} = -\gamma H,$$

| where H now denotes the Hamiltonian function of the particle's proper field. I also [40]
put $\mathfrak{g}_1 = \mathfrak{g}$, that is by $\mathfrak{g}$ I mean the field strength of the extended field in which the particle is moving. Then:

$$\mathfrak{P} = \gamma\mathfrak{g}\int_V H dV. \tag{110}$$

We have to define the *gravitational mass* m_g *of the particle* as:

$$m_g = \int_V H dV \tag{111}$$

and then:

$$\mathfrak{P} = \gamma m_g \mathfrak{g}. \tag{112}$$

We can therefore state the following proposition:

In a gravitational field there is only one type of action of force, the action of gravity, and there is nothing that would be related to it as the magnetic action of force is related to the electric one. But the gravitational mass of a material particle depends on its state of motion, in contrast to the constancy of the electric charge.

For, we have seen on p. 663 [p. 26 in the original] that:

$$\int H dV = \sqrt{1-q^2}E_0,$$

$$m_g = \sqrt{1-q^2}m_0, \tag{113}$$

if we understand m_0 to be the gravitational mass of the particle at rest. Moreover it is straightforward to verify the validity of the following proposition:

For a massive particle at rest the gravitational mass and the inertial mass are identical.

Both are $m_0 = E_0$. To what extent they differ in a moving body will be seen in a later section (45.).

Now I consider two particles at rest or in slow motion, having masses m_{g1} and m_{g2}. In the surrounding empty space the field strength of gravitation is calculated to be, due to $\operatorname{div}\mathfrak{k} = -\gamma H$ and, in vacuo, $\mathfrak{g} = \mathfrak{k}$:

$$\mathfrak{g}_1 = \gamma\frac{m_{g1}}{4\pi r_1{}^2}, \quad \mathfrak{g}_2 = \gamma\frac{m_{g2}}{4\pi r_2{}^2},$$

[41] | where r_1 and r_2 shall denote the radius vectors from the particles, and where the field lines point toward the particle generating the field. If the two are at a distance $r = r_1 = r_2$ from each other they accordingly attract each other with a force of equal magnitude:

$$\mathfrak{P}_g = \frac{\gamma^2}{4\pi}\frac{m_{g1}m_{g2}}{r^2}. \tag{114}$$

In our theory of gravitation the law of equality of action and reaction as well as Newton's law of attraction are valid.

Both laws are a necessary consequence of the principle of superposition that holds in vacuo.

Because the superposition principle also implies that the gravitational fields of very many mass particles, which merely represent the sinks of the vector $\mathfrak{k}$, simply add together, it follows *that the attractive effect of a body in the universe is altered in no way by interposing another body; rather the effect of the second body superposes unchanged; in other words, gravity cannot be shielded.*

The energy density, calculated according to the formula:

$$W = \Phi + \mathfrak{e}\cdot\mathfrak{d} - \varphi\rho + uw$$

becomes the following in vacuo, where $\mathfrak{e} = \mathfrak{d}, \mathfrak{b} = \mathfrak{h}, \rho = 0, \mathfrak{v} = 0, \mathfrak{k} = \mathfrak{g}, w = u$:

$$W = \frac{1}{2}(\mathfrak{e}^2 + \mathfrak{b}^2 + \mathfrak{g}^2 + u^2).$$

The energy density of the gravitational field in vacuo is a positive quantity.

This theorem, which is remarkable considering the attractive nature of the gravitational field, has already been derived by M. Abraham from his ansatz for the equations of gravitation.

Finally we want to calculate the numerical value of the universal constant γ from the above Newtonian law of attraction (114). In formula (114) the two gravitational masses m_{g1} and m_{g2} are to be expressed in ergs. | First let us give them in the usual fashion in grams by putting: [42]

$$m_1 = \frac{m_{g1}}{c^2}, \quad m_2 = \frac{m_{g2}}{c^2},$$

where c is the speed of light $(3 \cdot 10^{10})$. The law of attraction then takes the form:

$$\mathfrak{P}_g = \frac{\gamma^2 c^4}{4\pi} \frac{m_1 m_2}{r^2}. \tag{114a}$$

We denote what is usually called the *gravitational constant* by κ, so we have:

$$\kappa = \frac{\gamma^2 c^4}{4\pi}. \tag{115}$$

Therefore:

$$\gamma = \frac{\sqrt{4\pi\kappa}}{c^2}.$$

When we substitute:

$$\kappa = 6,648 \cdot 10^{-8},$$

we obtain:

$$\gamma = 1,016 \cdot 10^{-24}.$$

The Inertial Mass of a Material Particle

43. If E_0 is the energy of a particle at rest, then the energy of the same particle in motion with speed q follows from Laue's theorem:[19]

$$E = \frac{E_0}{\sqrt{1-q^2}}.$$

According to (91) and (98) the energy density is:

$$\begin{aligned} W &= H + \mathfrak{b} \cdot \mathfrak{h} - \mathfrak{f} \cdot \mathfrak{v} + wu \\ &= \Phi + \mathfrak{e} \cdot \mathfrak{d} - \varphi\rho + wu, \end{aligned}$$

19 M. Laue, *Das Relativitätsprinzip*, p. 170.

hence, considering (64) and (65):

$$E_0 = \int(\overline{H_0} + \overline{w_0 u_0})dV_0 = \int(\overline{\Phi_0} + \overline{w_0 u_0})dV_0. \tag{116}$$

These formulas now take the place of the formulas (66). |

[43] The following formula can be easily derived from equations (85), (86):

$$\mathfrak{k}\cdot\mathfrak{g} - wu = \operatorname{div}(\mathfrak{k}\omega) + \frac{\partial(w\omega)}{\partial t} + \gamma\omega H.$$

This implies for a material particle *at rest*, by integration over the whole volume V_0 it occupies:

$$\int_{V_0}\overline{\mathfrak{k}_0\cdot\mathfrak{g}_0}\,dV_0 - \int_{V_0}\overline{w_0 u_0}\,dV_0 = \gamma\int_{V_0}\overline{\omega_0 H_0}\,dV_0. \tag{117}$$

In place of the three equations on p. 659 [p. 10 in the original] we obtain the following equations from Laue's theorem, taking into account the gravitational terms:

$$\begin{aligned}
\int(\overline{\mathfrak{v}_0\cdot\mathfrak{h}_0} - \overline{\mathfrak{b}_{x0}\mathfrak{h}_{x0}} - \overline{\mathfrak{e}_{x0}\mathfrak{d}_{x0}} - \overline{\mathfrak{f}_{x0}\mathfrak{v}_{x0}} + \overline{\mathfrak{k}_{x0}\mathfrak{g}_{x0}} + \overline{w_0 u_0})dV_0 &= E_0,\\
\int(\overline{\mathfrak{v}_0\cdot\mathfrak{h}_0} - \overline{\mathfrak{b}_{y0}\mathfrak{h}_{y0}} - \overline{\mathfrak{e}_{y0}\mathfrak{d}_{y0}} - \overline{\mathfrak{f}_{y0}\mathfrak{v}_{y0}} + \overline{\mathfrak{k}_{y0}\mathfrak{g}_{y0}} + \overline{w_0 u_0})dV_0 &= E_0,\\
\int(\overline{\mathfrak{v}_0\cdot\mathfrak{h}_0} - \overline{\mathfrak{b}_{z0}\mathfrak{h}_{z0}} - \overline{\mathfrak{e}_{z0}\mathfrak{d}_{z0}} - \overline{\mathfrak{f}_{z0}\mathfrak{v}_{z0}} + \overline{\mathfrak{k}_{z0}\mathfrak{g}_{z0}} + \overline{w_0 u_0})dV_0 &= E_0.
\end{aligned}$$

Addition, taking note of (65), results in:

$$E_0 = -\frac{1}{3}\int(\overline{\mathfrak{e}_0\cdot\mathfrak{d}_0} - \overline{\mathfrak{b}_0\cdot\mathfrak{h}_0} - \overline{\mathfrak{k}_0\cdot\mathfrak{g}_0})dV_0 + \int\overline{w_0 u_0}\,dV_0 \tag{118}$$

in place of (71).

In the field of an electron we have $\mathfrak{h}_0 = 0,\ u_0 = 0$ *consequently:*

$$E_0 = -\frac{1}{3}\int\mathfrak{e}_0\cdot\mathfrak{d}_0\,dV_0 + \frac{1}{3}\int\mathfrak{k}_0\cdot\mathfrak{g}_0\,dV_0, \tag{119}$$

and here we have, from (64) and (115):

$$\int\mathfrak{e}_0\cdot\mathfrak{d}_0\,dV_0 = \int\varphi_0\rho\,dV_0, \tag{120}$$

$$\int\mathfrak{k}_0\cdot\mathfrak{g}_0\,dV_0 = \gamma\int\omega_0 H_0\,dV_0. \tag{121}$$

By formula (119) the proofs that we gave on p. 662 [p. 13 in the original] can no longer be carried through rigorously. Nevertheless it should be rather likely that the peculiar statements about the signs of $\mathfrak{e}$ and φ in the interior of the electron can in fact be maintained. |

Gravity of Moving Massive Particles [44]

44. Let $\mathfrak{g}_x, \mathfrak{g}_y, \mathfrak{g}_z, u$ be the gravitational field of a material particle, whose gravitational mass shall be m_0 when it is at rest. By (85), $\mathfrak{g}$ can always be derived from a gravitational potential ω:

$$\mathfrak{g}_x = \frac{\partial\omega}{\partial x},\quad \mathfrak{g}_y = \frac{\partial\omega}{\partial y},\quad \mathfrak{g}_z = \frac{\partial\omega}{\partial z},\quad u = -\frac{\partial\omega}{\partial t}\ \text{[1]}$$

and here ω is a four-dimensional scalar, that is, an invariant under Lorentz transformations. Let the particle move with speed q in the direction of the positive z-axis. We want to transform to rest, that is, we want to associate with the point x, y, z at time t a point x_0, y_0, z_0 according to the following equations:

$$x_0 = x,\quad y_0 = y,\quad z_0 = \frac{z-qt}{\sqrt{1-q^2}},\quad t_0 = \frac{t-qz}{\sqrt{1-q^2}}.$$

Let the center of the material particle be the point associated with $x_0 = 0, y_0 = 0, z_0 = 0$ which therefore has the coordinates $x = 0$, $y = 0$, $z = qt$. The distance of the point x, y, z from the center of the mass particle is:

$$r = \sqrt{x^2+y^2+(z-qt)^2},$$

and the distance of the associated point x_0, y_0, z_0 from the center $(0, 0, 0)$ of the particle in its rest frame is:

$$r_0 = \sqrt{x_0^2+y_0^2+z_0^2}\,.$$

We can also calculate this quantity as a function of x, y, z, t:

$$r_0 = \sqrt{x^2+y^2+\frac{(z-qt)^2}{1-q^2}}. \tag{122}$$

If we denote by ϑ the angle between the positive z-axis and the radius vector r, then:

$$z - qt = r\cos\vartheta$$

and thus we have:

$$r_0 = r\frac{\sqrt{1-q^2\sin^2\vartheta}}{\sqrt{1-q^2}} = r\sqrt{1+\frac{q^2}{1-q^2}\cos^2\vartheta}.$$

[45] | Let us introduce the following abbreviation:

$$\left.\begin{aligned} \sqrt{1+\frac{q^2}{1-q^2}\cos^2\vartheta} &= p, \\ r_0 &= rp. \end{aligned}\right\} \tag{123}$$

Now the gravitational potential ω for a particle at rest is easily calculated, namely:

$$\omega = \frac{\gamma m_0}{4\pi r_0}.$$

Since ω is invariant under Lorentz transformations it follows using (122) and (123) that the potential for a moving particle is given at the point (x, y, z) by the formula:

$$\omega = \frac{\gamma m_0}{4\pi\sqrt{x^2+y^2+\frac{(z-qt)^2}{1-q^2}}} = \frac{\gamma m_0}{4\pi r p}. \tag{124}$$

This gives immediately the gravitational field of the moving particle:

$$\left.\begin{aligned} \mathfrak{g}_x &= -\frac{\gamma}{4\pi p^3}\cdot\frac{m_0}{r^2}\cdot\frac{x}{r}, \\ \mathfrak{g}_y &= -\frac{\gamma}{4\pi p^3}\cdot\frac{m_0}{r^2}\cdot\frac{y}{r}, \\ \mathfrak{g}_z &= -\frac{1}{1-q^2}\cdot\frac{\gamma}{4\pi p^3}\cdot\frac{m_0}{r^2}\cdot\frac{z-qt}{r}, \\ u &= -\frac{q}{1-q^2}\cdot\frac{\gamma}{4\pi p^3}\cdot\frac{m_0}{r^2}\cdot\frac{z-qt}{r}, \end{aligned}\right\} \tag{125}$$

or, when we denote by $\mathfrak{g}_\rho$ the component of the field normal to the direction of motion (z):

$$\left.\begin{aligned} \mathfrak{g}_\rho &= -\frac{\gamma m_0}{4\pi r^2}\cdot\frac{\sin\vartheta}{p^3}, \\ \mathfrak{g}_z &= -\frac{\gamma m_0}{4\pi r^2}\cdot\frac{\cos\vartheta}{(1-q^2)p^3}, \\ u &= -\frac{\gamma m_0 q}{4\pi r^2}\cdot\frac{\cos\vartheta}{(1-q^2)p^3}, \end{aligned}\right\} \tag{126}$$

where the value of p is to be substituted from formula (123).

The formulas (126) clearly imply the following:

The gravitational field lines in the vicinity of a material body, which extend from it in a straight | radial direction when it is at rest, acquire a curved form when the body is in motion; in addition the field acquires sources and sinks in the vicinity of the moving particle. [46]

The last part is easily seen: in empty space we have $\mathfrak{g} = \mathfrak{k}$ and $u = w$; since here u and also $\partial u / \partial t$ differ from zero, this is also true of $\operatorname{div}\mathfrak{g}$, for:

$$\operatorname{div}\mathfrak{g} = \operatorname{div}\mathfrak{k} = -\frac{\partial w}{\partial t} = -\frac{\partial u}{\partial t}.$$

The order of magnitude of $\operatorname{div}\mathfrak{g}$ is that of q^2, as one can easily check.

The strange distortion of the gravitational field becomes noticeable only when q takes on quite significant values. The equation for a line of force is:

$$dz : d\rho = \frac{z - qt}{1 - q^2} : \rho,$$

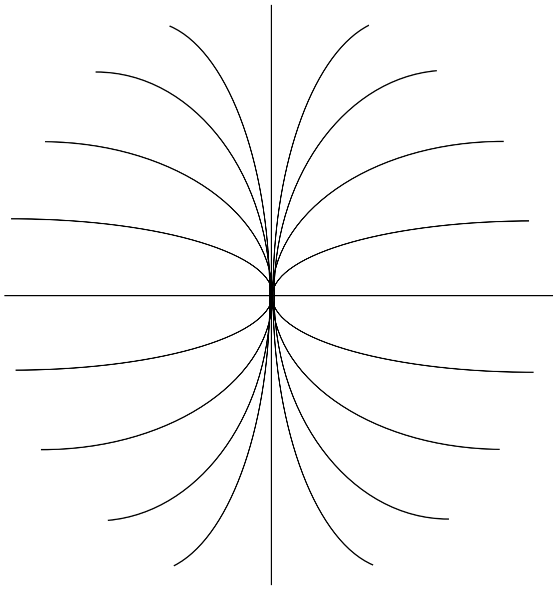


Figure 1: Shape of the gravitational field lines of a rapidly moving particle

$\left(q = \sqrt{\frac{1}{2}} = 212000 \text{ km/sec}\right).$

| therefore: [47]

$$z - qt = a \cdot \rho^{\frac{1}{1 - q^2}},$$

where a is a parameter labeling the line of force. One sees immediately that for lesser values of q the line of force hardly deviates from the radius vector, then a is given by $\operatorname{cotg}\vartheta$. But if $q^2 = \frac{1}{2}$, for example, then the lines of force have already become parabolas. Then the gravitational field has the appearance shown in the drawing above (Fig. 1). As q approaches the speed of light (the value 1), the curves open up more and more, so that the gravitational field is increasingly concentrated about the equatorial plane. Simultaneously the field strength decreases steadily toward zero, so that $\mathfrak{g}_\rho$ converges to zero as $\sqrt{(1-q^2)^3}$, and $\mathfrak{g}_z$ and u as $\sqrt{(1-q^2)}$.

45. Let us consider a body whose elementary particles are all completely at rest relative to each other, as it may be at absolute zero temperature. For this body the inertial mass is identical with the gravitational mass, let us denote it as rest mass m_0.

Now let the elementary particles in this body start to vibrate, due to an increase in temperature, for example. Let the average speed of a particle be q, then:

$$\frac{E_0}{\sqrt{1-q^2}} \sim E_0 + \frac{1}{2}E_0 q^2$$

is the average energy of a moving particle, if E_0 is its rest energy. The gravitational mass of the particle is given by (113):

$$E_0\sqrt{1-q^2} \sim E_0 - \frac{1}{2}E_0 q^2.$$

So the gravitational mass m_g of the whole body being considered decreases as the internal motion of its elementary particles increases. In fact we can estimate the change in gravitational mass if we know the magnitude of the internal motion. Let it be Q ergs, then:

$$m_g \sim m_0 - Q,$$

[48] | if the mass is specified in ergs, or:

$$m_g \sim m_0 - \frac{Q}{9 \cdot 10^{20}},$$

if the mass is calculated in grams. If we impart our body a motion of velocity v, then the total velocity q' of an elementary particle, moving with velocity q *relative* to the body in a direction at angle ϑ with respect to the direction of v, becomes by the addition theorem of velocities:[20]

$$q'^2 = \frac{q^2 + v^2 + 2qv\cos\vartheta - q^2v^2\sin^2\vartheta}{(1+q\,v\cos\vartheta)^2}.$$

The energy of this particle can be calculated:

20 M. Laue, *Das Relativitätsprinzip*. p. 43.

$$\frac{E_0}{\sqrt{1-q'^2}} = \frac{E_0(1+qv\cos\vartheta)}{\sqrt{1-q^2}\sqrt{1-v^2}}.$$

We assume that the particles in the body oscillate at random, so that each direction of q occurs equally often. Then for a large number N of particles at each moment a fraction:

$$dN = N\frac{2\pi\sin\vartheta d\vartheta}{4\pi} = \frac{1}{2}N\sin\vartheta d\vartheta$$

moves such that the direction of motion q of these dN particles makes an angle with respect to the direction of v lying between ϑ and $\vartheta + d\vartheta$. We obtain the energy of all of the N particles by multiplying the energy value of a particle, just calculated, by the number dN, and integrating over ϑ from 0 to π. Since:

$$\int_0^\pi \cos\vartheta\sin\vartheta d\vartheta = 0; \qquad \int_0^\pi \sin\vartheta d\vartheta = 2,$$

the total energy of the N particles is obtained as:

$$\frac{E_0}{\sqrt{1-q^2}\sqrt{1-v^2}} = \frac{E_0}{\sqrt{1-q^2}}\left(1+\frac{1}{2}v^2+\ldots\right).$$

| Thus the inertial mass m of a body, as is well known, simply equals its total energy content, even when its elementary particles are in vibration [49]

$$m = \frac{E_0}{\sqrt{1-q^2}} \sim E_0 + \frac{1}{2}E_0 q^2.$$

Let us again call the energy of internal motion Q, then:

$$m \sim m_0 + Q \quad \text{erg}$$

or:

$$m \sim m_0 + \frac{Q}{9\cdot 10^{20}} \quad \text{grams}.$$

Inertial mass and gravitational mass of a body are completely identical only if the body's elementary particles execute no internal motion. Hidden motion of the elementary particles cause an increase of the inertial mass and a decrease of the gravitational mass.

Since hidden motion is certainly always present in any matter, increasing with increasing temperature, it further follows:

For any substance, the ratio of gravitational to inertial mass, and therefore also the so-called gravitational constant, is a function of the temperature, which decreases with increasing temperature.

For not excessively large thermal motion we can use the approximate values for m_g and m that we just calculated:

$$\frac{m_g}{m} = \frac{m_0 - Q}{m_0 + Q} = 1 - 2\frac{Q}{m},$$

if m and Q are both calculated in ergs, or:

$$\frac{m_g}{m} = 1 - 2\frac{Q}{9 \cdot 10^{20} \cdot m},$$

if Q is in ergs and m in grams; $Q:m$ is that part of the heat contents of a unit of mass of the body that represents the kinetic energy of the molecular motion.

The change of the gravitational constant with temperature is of different amounts in different materials, such that it is larger the larger the part of the material's specific heat that corresponds to the kinetic energy of molecular motion. |

[50] In general the specific heat of bodies is larger the smaller the atomic weights of its constituents. Therefore the propositions that follow from our theory of gravity might be tested experimentally by determining the acceleration of gravity once with a pendulum whose bob consists of Lithium, and again under exactly the same conditions with a pendulum whose bob consists of Lead. The second pendulum should give a larger value for the acceleration of gravity than the first. To assess the feasibility of the experiment let us calculate the variation of the ratio m_g/m for that substance which must exhibit it to the greatest extent, namely hydrogen gas. It has a specific heat c_v of 2.5 cal per gram and per degree Celsius, that is converted into erg $1.05 \cdot 10^8$. This implies:

$$\frac{m_g}{m} = 1 - \frac{2.1 \cdot 10^8}{9 \cdot 10^{20}} \cdot \Theta = 1 - 2.3 \cdot 10^{-13} \cdot \Theta,$$

where Θ is the absolute temperature at which the measurement is performed. This yields at:

$$\Theta = 300^\circ \qquad 1 - 7 \cdot 10^{-11},$$
$$\Theta = 6000^\circ \qquad 1 - 14 \cdot 10^{-10}.$$

Thus it would have to be possible to determine the acceleration of gravity accurately to a fraction 10^{-11} or 10^{-12} in order to find differences for different pendulum bob materials, when observing at room temperature. Similar accuracy would be required when searching for a variability of the gravitational constant, say by astronomical means. For even though the higher temperatures of many celestial bodies could play a role here, one must also note the greater atomic weights of the materials that constitute most celestial bodies.

Gravitational mass and inertial mass are indistinguishable in practice.

Longitudinal Waves in the Aether

46. It seems that the most interesting consequence of the theory of gravitation developed here, which incidentally was also pointed out already by | M. Abraham[21] when [51]
he set up his ansatz, is the prediction of longitudinal waves in the aether. This can be immediately seen from the form of the equations (85) and (86). The equations (85) are absolutely identical with the equations of motion of a compressible and perfectly elastic fluid that executes infinitesimal, non-vortical motions, if one views ω as the velocity potential; and equation (86) is the so-called continuity equation if one can set $H = 0$, which is permitted, at least in vacuo. Here it must be supposed that $(\mathfrak{g}, iu)$ and (k, iw) are proportional to each other, which by the superposition principle is the case in vacuo. Then, for ω the wave equation results:

$$\frac{\partial^2\omega}{\partial x^2}+\frac{\partial^2\omega}{\partial y^2}+\frac{\partial^2\omega}{\partial z^2}-\frac{\partial^2\omega}{\partial t^2} = 0$$

from which it can be seen that the speed of longitudinal waves in the aether is 1, that is equal to the speed of light waves.

The analogy to longitudinal waves in a compressible fluid suggests that longitudinal waves in the aether must radiate from oscillating material particles. For, when a material particle oscillates, then (1) the sinks of the vector $\mathfrak{g}$ move back and forth periodically and (2) the amount of the sink varies periodically, since the particle's gravitational mass reaches a minimum at the time of greatest motion, and a maximum at the moment when our particle is at rest. Thus two different longitudinal spherical waves originate around an oscillating material particle, and one can already note that the second type has twice the frequency of the first.

We see from this that light waves, if emitted by oscillating electrons, and x-rays, which originate upon sudden deceleration of moving electrons, must always be accompanied by radiation of longitudinal waves. However, the same cannot be immediately said of such light waves that consist of exploded | dipoles (as in section 32). [52]
Let us now calculate how great the intensity of gravitational waves emitted by an oscillating material particle is.

For simplicity we will assume that the vectors $\mathfrak{h}$ and $\mathfrak{v}$ are zero in the whole surrounding of the particle when it is at rest, and also that $u = 0$. Then the energy of the particle at rest is always:

$$E = \int H dV.$$

If this particle is at an equilibrium position, then in particular:

$$E_0 = \int H_0 dV.$$

21 M. Abraham, *Physik. Zeitschr.* 13. p. 1. 1912.

At other positions we have $E > E_0$, and to calculate E we may have to take into account the further surroundings of the particle.

As the particle moves through its equilibrium position with speed q, its energy is:

$$E = \frac{E_0}{\sqrt{1-q^2}},$$

but from (69):

$$\int H dV = E_0\sqrt{1-q^2} = E(1-q^2).$$

If E is the total energy of oscillation, which stays constant during the oscillation (apart from radiation damping), then at the moment of maximum amplitude, when $q = 0$:

$$\int H dV = E,$$

and at the moment of crossing the equilibrium position, when q reaches its maximum:

$$\int H dV = E(1-q^2).$$

We will now make use of the following abbreviations:

$$\left.\begin{aligned} \int H dV &= \mu, \\ E &= \mu_0. \end{aligned}\right\} \tag{127}$$

[53] | If we think of μ as the massive particle, we can visualize the whole oscillation process thus: (1) the center of mass of μ moves with a periodic velocity, which we will call q and (2) concurrently the mass of the particle μ changes periodically; we can calculate to sufficient accuracy:

$$\mu = \mu_0(1-q^2), \tag{128}$$

where μ_0 is the value for $q = 0$. For the intensity of the emitted waves it will be inessential whether, in doing this, the particle also undergoes any kind of change of shape.

In the following let us confine attention to the case of linear oscillation of the particle, where therefore q does not change direction. We will put:

$$q = a \sin 2\pi n t. \tag{129}$$

In place of the vector of gravity $\mathfrak{g}$ and u we first calculate quantities defined somewhat differently, which we will call $\mathfrak{g}'$ and u', and which shall satisfy the following equations:

$$\left.\begin{aligned} \mathfrak{g}' &= \nabla\omega', \quad u' = -\frac{\partial\omega'}{\partial t}, \\ \operatorname{div}\mathfrak{g}' + \frac{\partial u'}{\partial t} &= -\gamma H, \end{aligned}\right\} \tag{130}$$

In vacuo, where $\mathfrak{g}$ and k, and u and w are identical to each other, the definition of $\mathfrak{g}'$ and u' as well as ω' agrees with that of $\mathfrak{g}$, u and ω, but not in the interior of the material particle, because there k and $\mathfrak{g}$, and u and w may not be regarded as identical. From (130) one derives the well-known differential equation for ω':

$$\frac{\partial^2\omega'}{\partial x^2} + \frac{\partial^2\omega'}{\partial y^2} + \frac{\partial^2\omega'}{\partial z^2} - \frac{\partial^2\omega'}{\partial t^2} = -\gamma H, \tag{131}$$

which agrees completely with the differential equation obeyed by the scalar potential of the electric field about a moving charged particle, if $+\gamma H$ is taken as the density of electric charge. But this equation can be easily integrated by the method given by E. Wiechert[22] if one confines attention to | regions of space whose distances from the [54]
center of mass of the moving particle are infinitely large compared to the particle's dimensions. Namely:

$$\omega'(x, y, z, t) = \frac{\gamma}{4\pi}\left(\frac{\mu}{r(1 - q\cos\vartheta)}\right)_{x_1, y_1, z_1, t_1}, \tag{132}$$

where:

$$\mu = \int H dV \qquad \text{at time } t_1.$$

Here x_1, y_1, z_1 denotes that point on the orbit described by the particle's center of mass from which a light wave would just arrive at the point in space (x, y, z) under consideration at the time t, and t_1 is the moment at which the particle's center of mass is just passing (x_1, y_1, z_1). r is the connecting segment from (x_1, y_1, z_1) to (x, y, z). Since we have put the speed of light equal to unity we have:

$$t_1 = t - r.$$

Let ϑ be the angle between the radius vector r, directed from (x_1, y_1, z_1) to (x, y, z), and the direction of motion q. According to (129) q is given as a function of t_1: $q = a\sin 2\pi n \cdot t_1$ (t now has a different meaning). If we take the direction of motion of the oscillating particle as the direction of the z-axis, then x_1, y_1 are constants and z_1 is a function of t_1:

22 E. Wiechert, *Ann. d. Phys.* 4. p. 682. 1901.

$$z_1 = \mathfrak{f}(t_1), \ \frac{dz_1}{dt_1} = q = \mathfrak{f}'(t_1).$$

According to (129) we have to set:

$$z_1 = -\frac{a}{2\pi n}\cos 2\pi n t_1, \ \frac{dz_1}{dt_1} = q = a\sin 2\pi n t_1.$$

The equation:

$$t_1 = t - r = t - \sqrt{(x-x_1)^2 + (y-y_1)^2 + (z-z_1)^2},$$

with the constant values for x_1 and y_1 and $z_1 = \mathfrak{f}(t_1)$ substituted, determines the quantity t_1 implicitly as function of (x, y, z, t), and the differential quotients

$$\frac{\partial t_1}{\partial x}, \ \frac{\partial t_1}{\partial y}, \ \frac{\partial t_1}{\partial z}, \ \frac{\partial t_1}{\partial t}$$

[55] | can be calculated without effort. Since q and therefore also μ are functions of t_1, they can also be differentiated with respect to x, y, z, t, where we will use the abbreviation:

$$\frac{dq}{dt_1} = \dot{q}, \ \frac{d\mu}{dt_1} = \dot{\mu}.$$

Finally one can also perform the differentiation of r and $rq\cos\vartheta = (z - z_1)q$ and calculate now the vector $\mathfrak{g}'$ and u':

$$\mathfrak{g}_x{}' = \frac{\partial\omega'}{\partial x}, \ \mathfrak{g}_y{}' = \frac{\partial\omega'}{\partial y}, \ \mathfrak{g}_z{}' = \frac{\partial\omega'}{\partial z}, \ u' = -\frac{\partial\omega'}{\partial t}.$$

The calculations have been carried through exactly by M. Abraham in his *Theorie der Elektrizität* vol. II in § 13 p. 92ff, therefore I need only write down the results. Because everything is symmetric about the z-axis I will give only two components of $\mathfrak{g}'$: namely $\mathfrak{g}_z{}'$ and $\mathfrak{g}_\rho{}' \perp \mathfrak{g}_z{}'$:

$$\left.\begin{aligned}
\mathfrak{g}_\rho' &= -\frac{\gamma\mu}{4\pi r^2(1-q\cos\vartheta)^3}(1-q^2)\sin\vartheta - \frac{\gamma\mu}{4\pi r(1-q\cos\vartheta)^3}\dot{q}\cos\vartheta\sin\vartheta \\
&\qquad - \frac{\gamma}{4\pi r(1-q\cos\vartheta)^2}\dot{\mu}\sin\vartheta, \\
\mathfrak{g}_z' &= -\frac{\gamma\mu}{4\pi r^2(1-q\cos\vartheta)^3}(\cos\vartheta - q) - \frac{\gamma\mu}{4\pi r(1-q\cos\vartheta)^3}\dot{q}\cos^2\vartheta \\
&\qquad - \frac{\gamma}{4\pi r(1-q\cos\vartheta)^2}\dot{\mu}\cos\vartheta, \\
u' &= -\frac{\gamma\mu}{4\pi r^2(1-q\cos\vartheta)^3}q(\cos\vartheta - q) - \frac{\gamma\mu}{4\pi r(1-q\cos\vartheta)^3}\dot{q}\cos\vartheta \\
&\qquad - \frac{\gamma}{4\pi r(1-q\cos\vartheta)^2}\dot{\mu},
\end{aligned}\right\} \tag{133}$$

These expressions decompose into two parts, the first being proportional to r^{-2}, and the second to r^{-1}. The two summands of the second part contain as a factor either:

$$\dot{q} = 2\pi n a \cos 2\pi n t = \frac{2\pi}{\lambda} a \cos 2\pi n t$$

or:

$$\mu = -\mu_0 q \dot{q} = -\frac{2\pi}{\lambda}\mu_0 a^2 \sin 2\pi n t \cos 2\pi n t$$

| The second part of the expressions is related to the first in order of magnitude as $1/\lambda : 1/r$, where λ is the wavelength of light corresponding to the wave number n. [56]

Let us confine attention to those wave numbers n whose wavelength of light λ is infinitely large in comparison with the dimensions of the particle, in the sense *that there are distances* r *from the particle that are infinitely small compared to* λ, *but still infinitely large compared to the dimensions of the particle*.

By this assumption there are values of r for which the formulas (133) are still valid, although r is infinitesimal compared to λ. For these *small values of* r we can calculate to good approximation:

$$\left.\begin{aligned}
\mathfrak{g}_\rho' &= -\frac{\gamma\mu}{4\pi r^2(1-q\cos\vartheta)^3}(1-q^2)\sin\vartheta, \\
\mathfrak{g}_z' &= -\frac{\gamma\mu}{4\pi r^2(1-q\cos\vartheta)^3}(\cos\vartheta - q), \\
u' &= -\frac{\gamma\cdot\mu}{4\pi r^2(1-q\cos\vartheta)^3}q(\cos\vartheta - q).
\end{aligned}\right\} \tag{134}$$

It is easy to show that these formulas agree completely with the formulas (123), which represent the field of a massive particle moving with speed q.

To see this we have to note that r stands for the distance of the point (x, y, z) from the position of the particle at the time $t_1 = t - r$. If we denote the distance between (x, y, z) and the position of the particle at time t by r', and the angle between r' and the z-axis by ϑ', we see immediately from Fig. 2 that:

$$\begin{aligned} r \sin\vartheta &= r' \sin\vartheta', \\ r \cos\vartheta &= r' \cos\vartheta' + q(t - t_1) = r' \cos\vartheta' + rq. \end{aligned}$$

An elementary calculation leads to the formula:

$$r'^2(1 - q^2 \sin^2\vartheta') = r^2(1 - q \cos\vartheta)^2$$

or:

$$r' \sqrt{\left(1 + \frac{q^2}{1 - q^2}\cos^2\vartheta'\right)} \sqrt{1 - q^2} = r(1 - q \cos\vartheta).$$

[57] | Noting that:

$$\begin{aligned} \sin\vartheta &= \frac{r'}{r}\sin\vartheta', \\ \cos\vartheta - q &= \frac{r'}{r}\cos\vartheta' \end{aligned}$$

and putting as before (121):

$$\sqrt{1 + \frac{q^2}{1 - q^2}\cos^2\vartheta'} = p,$$

we readily derive from (134) the formulas (125) where:

$$m_0 = \frac{\mu}{\sqrt{1 - q^2}}$$

in agreement with formula (113).

At distances that are large compared to the dimensions of the particle, but which still belong to its closer vicinity (r small compared to λ) the vector $\mathfrak{g}'$ and u' agrees completely with the vector of gravity $\mathfrak{g}$ and u of the moving massive particle.

Hence it follows that this is valid for all distances that are large compared to λ. Only in the interior of the particle and in the nearest vicinity can $\mathfrak{g}', u'$ be distinguished from $\mathfrak{g}, u$. But since these regions do not interest us I will simply omit the primes in the following and put:

$$\mathfrak{g}' = \mathfrak{g}, \qquad u' = u.$$

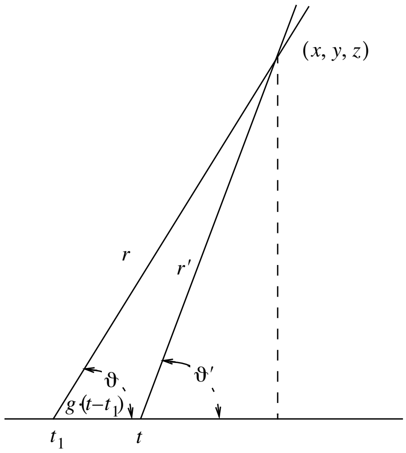


Figure 2

| At large distances, where r is infinitely large compared to λ, we have the pure longitudinal spherical waves that we desire. We need to take into account only the second part of the expressions (133): [58]

$$\mathfrak{g}_\rho = -\frac{\gamma\mu}{4\pi r}\left\{\frac{\dot{q}}{(1-q\cos\vartheta)^3}\cos\vartheta\sin\vartheta - \frac{q\dot{q}}{(1-q\cos\vartheta)^2}\frac{\sin\vartheta}{(1-q)^2}\right\}$$

by noting that:

$$\dot{\mu} = -\mu_0 q\dot{q} = -\frac{\mu}{1-q^2}q\dot{q}.$$

A minor transformation yields:

$$\mathfrak{g}_\rho = -\frac{\gamma\mu}{4\pi r(1-q\cos\vartheta)^3(1-q^2)}(\dot{q}\cos\vartheta - q\dot{q})\sin\vartheta.$$

If we transform $\mathfrak{g}_z$ and u similarly, we finally obtain the following expressions for the variables of state in the longitudinal wave *at large distances from the oscillating particle*:

$$\left.\begin{aligned} \mathfrak{g}_\rho &= -\frac{\gamma\mu}{4\pi r(1-q\cos\vartheta)^3(1-q^2)}(\dot{q}\cos\vartheta - q\dot{q})\sin\vartheta, \\ \mathfrak{g}_z &= -\frac{\gamma\mu}{4\pi r(1-q\cos\vartheta)^3(1-q^2)}(\dot{q}\cos\vartheta - q\dot{q})\cos\vartheta, \\ u &= -\frac{\gamma\cdot\mu}{4\pi r(1-q\cos\vartheta)^3(1-q^2)}(\dot{q}\cos\vartheta - q\dot{q}). \end{aligned}\right\} \tag{135}$$

This shows that $\mathfrak{g}$ is always oriented radially toward the particle, regarding the norm $\mathfrak{g} = u$. Furthermore the two waves mentioned above are clearly recognizable; the term $\dot{q}\cos\vartheta$ is due to the back and forth motion of the sink of the vector $\mathfrak{g}$, and the term $q\dot{q}$ is due to the change of the particle's gravitational mass as it is moving. Since:

$$q\dot{q} = a^2\sin 2\pi nt\cos 2\pi nt = \frac{1}{2}a^2\sin 4\pi nt$$

the second wave has twice the frequency of the first. If the particle's displacements are very large, so that q is not much less than 1, then the factor before the parenthesis causes the emitted radiation to consist of oscillations that are not pure sine waves. There is little interest for us to discuss this complication more precisely here, because motions of material particles with velocity not *far* below 1 are well known to be
[59] extremely rare. | Therefore we want to discuss exclusively the case that q is very small. Then $q\cdot\dot{q}$ can also be dropped compared to $\dot{q}\cos\vartheta$, and we retain only the first wave:

$$\begin{aligned} \mathfrak{g} &= -\frac{\gamma\mu}{4\pi r}\dot{q}\cos\vartheta, \\ u &= -\frac{\gamma\mu}{4\pi r}\dot{q}\cos\vartheta. \end{aligned} \tag{136}$$

Here μ is simply the *mass* of the particle, since for small velocities inertial and gravitational mass are identical. We substitute in (136):

$$\dot{q} = \frac{2\pi}{\lambda}a\cos 2\pi nt$$

and calculate the energy current of the wave in direction ϑ to be:

$$\mathfrak{g}u = \frac{\gamma^2\mu^2}{16\pi^2r^2}\left(\frac{2\pi}{\lambda}\right)^2 a^2\cos^2\vartheta\cos^2 2\pi nt$$

from which one obtains the intensity by integration over a unit of time:

$$J_g = \frac{\gamma^2\mu^2}{16\pi^2r^2}\left(\frac{2\pi}{\lambda}\right)^2\frac{1}{2}a^2\cos^2\vartheta. \tag{137}$$

Let us now consider a very large number N of oscillating particles, which are oriented quite at random in space, but which all oscillate linearly with amplitude a. Since in the spherical zone between ϑ and $\vartheta + d\vartheta$ their number is:

$$dN = \frac{N}{2}\sin\vartheta d\vartheta$$

we obtain the whole intensity by multiplying (137) by this value of dN and integrating over ϑ, with the result obtained in this way:

$$NJ_g = \frac{1}{6}N\frac{\gamma^2\mu^2}{16\pi^2r^2}\left(\frac{2\pi}{\lambda}\right)^2 a^2. \tag{138}$$

Let us compare this value with the corresponding value of the intensity of the emitted light. We consider a material particle with charge ε whose direction of oscillation makes an angle ϑ with the ray direction, and which oscillates with speed $q = a\sin 2\pi nt$. | The intensity of the light emitted by this particle is, in the system [60] of units used by us:[23]

$$J_e = \frac{\varepsilon^2}{16\pi^2r^2}\left(\frac{2\pi}{\lambda}\right)^2\frac{1}{2}a^2\sin^2\vartheta \tag{139}$$

and:

$$NJ_e = \frac{1}{3}N\frac{\varepsilon^2}{16\pi^2r^2}\left(\frac{2\pi}{\lambda}\right)^2 a^2. \tag{140}$$

Accordingly the ratio of the intensities for the electric and gravitational waves radiated by the same particle is:

$$\frac{\bar{J}_e}{\bar{J}_g} = 2\frac{\varepsilon^2}{\gamma^2\mu^2}. \tag{141}$$

In this formula ε means the amount of charge of the particle, calculated in a unit that is the $\sqrt{4\pi}^{\text{th}}$ part of the usual electrostatic unit, so that:

$$\varepsilon^2 = 4\pi\varepsilon_s^2,$$

if ε_s is the charge calculated in the ordinary electrostatic unit. Further, μ means the mass calculated in ergs, that is:

$$\mu = c^2m,$$

where c denotes the speed of light and m the mass in grams. Finally, if we denote the charge in electromagnetic units by e, then:

23 Cf. E. Wiechert, *Ann. d. Phys*. 4. p. 688. 1901.

$$\frac{\varepsilon_s}{c} = e.$$

Thus we obtain:

$$\frac{\bar{J}_e}{\bar{J}_g} = 2\frac{4\pi}{\gamma^2 c^2}\left(\frac{e}{m}\right)^2,$$

or, since by (113) the ordinarily so-called gravitational constant κ has the value:

$$\kappa = \frac{\gamma^2 c^4}{4\pi},$$

we have:

$$\frac{\bar{J}_e}{\bar{J}_g} = 2\frac{c^2}{\kappa}\left(\frac{e}{m}\right)^2, \tag{142}$$

$$c = 3\cdot 10^{10}, \quad \kappa = 6,65\cdot 10^{-8}, \quad e/m = 1,75\cdot 10^{7},$$

[61] | so the result for the radiation emitted by oscillating electrons is:

$$\frac{\bar{J}_e}{\bar{J}_g} = 8,3\cdot 10^{42}.$$

To appreciate this number properly let us take the square root, $\sim 3\cdot 10^{21}$. Then we see the following: the intensity of gravity radiation emitted by a radiating point at a distance of 1 cm is just as intense as the radiation of light emitted by it at the distance of $3\cdot 10^{21}$ cm, that is a distance a hundred million times the diameter of the Earth's orbit, or about 3000 light years. Here this ratio is the same for all wavelengths.

The gravitational radiation emitted by oscillating electrons (or by any oscillating mass particle) is so extraordinarily weak that it is unthinkable ever to detect it by any means whatsoever.

This makes it transparent why the longitudinal radiation of the aether apparently plays no role whatsoever in the balance of nature. It would probably be very imprudent to claim that the longitudinal waves, which certainly as such are possible at any amplitude, could not arise in appreciable intensity for other than oscillatory processes of material particles. We can only claim this much with certainty, that these processes would have to be of a highly peculiar type. So if one could ever prove the existence of gravitational waves, the processes responsible for their generation would probably be much more curious and interesting than even the waves themselves.

The Theorem on the Relativity of the Gravitational Potential

47. The theory advocated in this work differs from the theory generally adopted to date mainly because it must put the real vacua in contrast with yet another, ideal vacuum, as in the theory of gases the real gases vs. the ideal gas. For in real vacua, due to

the proximity of material bodies, traces of the states ρ, $\mathfrak{v}$ are always present and H is not absolutely zero; therefore the law of superposition of | the states of the aether, [62] which characterizes the absolute vacuum, is always valid only as a good approximation, admittedly to such a good approximation that one can hardly hope very easily to substantiate deviations from this law. However it may nevertheless be possible, in very strong electric fields where $\mathfrak{e}$ and φ are large,[24] or in very strong magnetic fields, to make observations which contradict our present-day ideas about the vacuum, and such observations should be viewed as the greatest encouragement on the path followed by me.

Such observations would concern a vacuum that already deviates rather strongly from the ideal vacuum. But among the state variables is one which appears to influence processes even in a vacuum that otherwise deserves to be called almost ideal, and that is the gravitational potential ω. If the quantities $\mathfrak{v}$ and ρ, as well as H, are so small that one can no longer speak of any noticeable influence on physical laws, but if in this good vacuum ω still has a large value, then we do not have $\mathfrak{e} = \mathfrak{d}$, $\mathfrak{b} = \mathfrak{h}$ but, as shown by (93) and (94):

$$\mathfrak{e} = e^{-\gamma\omega} \cdot \mathfrak{d}, \quad \mathfrak{b} = e^{-\gamma\omega} \cdot \mathfrak{h}.$$

From this it is evident that the dielectric constant of the vacuum is no longer 1 in a region where the gravitational potential ω is present, instead $K = e^{+\gamma\omega}$, similarly the permeability of the vacuum is no longer 1, but $M = e^{-\gamma\omega}$. But the product of the two is again $K \cdot \mathrm{M} = 1$, that is *the speed of light in a region at gravitational potential* ω *is the same as in a region with zero gravitational potential*.

But we can go much further. Let the mean value of the gravitational potential in our region be ω_0, an arbitrarily large but constant value. Then we can decompose the gravitational potential, which | is not constant due to the presence of matter and of [63] gravitational fields, into two parts:

$$\omega = \omega_0 + \omega_1.$$

The second, variable term takes on large values at most in the interior of material bodies present in the region, in vacuo itself ω_1 is small. If we define:

$$H_1 = e^{+\gamma\omega_0} H = e^{-\gamma\omega_1} H'(\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v}, \mathfrak{g}, u),$$

$$\mathfrak{e}_1 = \frac{\partial H_1}{\partial \mathfrak{d}}, \quad \mathfrak{b}_1 = -\frac{\partial H_1}{\partial \mathfrak{h}}, \quad \varphi_1 = -\frac{\partial H_1}{\partial \rho}, \quad \mathfrak{f}_1 = \frac{\partial H_1}{\partial \mathfrak{v}},$$

$$\mathfrak{k}_1 = \frac{\partial H_1}{\partial \mathfrak{g}}, \quad w_1 = -\frac{\partial H_1}{\partial u}, \quad \frac{\partial H_1}{\partial \omega_1} = -\gamma H_1,$$

24 Cf. II. p. 24 [in chap. 4 of the original, which is not included in this translation].

then in the region at gravitational potential ω_0 exactly the same equations hold for the quantities denoted with the index 1 as in a region at gravitational potential 0 for the quantities without index. From this it follows directly:

If two empty regions differ only in this, that the gravitational potential in one of them has a very large average value ω_0*, and in the other one it has the average value zero, then this does not have the least influence on the size and form of the electrons and other material elementary particles, on their charges, their laws of oscillation and other laws of motion, on the speed of light, and in general on all physical relations and processes.*

Through this theorem, which could be called the principle of the relativity of the gravitational potential, the theory of gravitation developed by me differs in principle and most sharply from the theories of A. Einstein and M. Abraham. I share the view of the latter that if this theorem were not valid, it would mean the demise of the entire principle of relativity. On the other hand I believe to have shown that the postulates I assumed lead nowhere to contradictions with experience, that in particular according to my theory only imperceptibly small deviations are to be expected from the law of proportionality of gravitational mass and inertial mass. |

[64] *Concluding Remarks*

48. Above I believe to have pursued the general theory of matter as far as is possible today. The next progress must occur through experiments, and therefore I want to discuss briefly what possibilities offer themselves to experiment.

Gravity, the preparation of whose experimental investigation was the main goal that was on my mind in this research, shows itself as obstinate as ever. It was possible to implement the theory of gravitation completely so that it is in accord with the principle of relativity as well as with all empirical facts known about gravity, and it also yields two new results that seem extremely interesting on first sight. But a closer look shows that these theoretical results provide no prospects for a successful experiment. The first result is that the ratio of gravitational to inertial mass depends on temperature, and that the dependence on temperature is much stronger for bodies of small atomic weight than for bodies of large atomic weight. Because the differences to be expected from theory for the acceleration of gravity of different substances is of the order 10^{-12} to 10^{-11}, it is experimentally useless. The second result is that there must be longitudinal waves in the aether, for which it may be worth searching. Of the processes known to us, the oscillations of atoms and electrons are relevant, which must produce longitudinal gravitational waves as they produce light. However, for electron oscillations the intensity of gravitational waves is to that of the light waves of any frequency as $1:8.3 \cdot 10^{42}$, and we must therefore deem the existence of any reagent that would react to them as totally ruled out. Thus no way can be given to search for these longitudinal waves, which by themselves are certainly highly interesting.

The next thing immediately suggested by the theory would be an investigation whether there are to be found, in very strong electric or magnetic | fields or in field-free regions at a very high potential, any deviations from the laws of Maxwell that are valid in the ideal vacuum. These would be high precision measurements, to be performed according to a theoretically well thought-out plan. Of course it is doubtful whether this will lead to success; but if there were positive results, they would give important hints to the theory how the next steps should be taken. [65]

In somewhat loose connection with the remaining theory is the conception laid out by me in the sections 31 to 36 about the quanta of action and the light of band spectra. The conception is very vague and full of hypotheses, nevertheless I believe that one could draw some conclusions from it which should give rise to new spectroscopic investigations.

EDITORIAL NOTE

[1] In the original, the second equation is misprinted as $\mathfrak{g}_y = \frac{\partial \omega}{\partial x}$.

GUSTAV MIE

REMARKS CONCERNING EINSTEIN'S THEORY OF GRAVITATION

Originally published as "Bemerkungen zu der Einsteinschen Gravitationstheorie" in Physikalische Zeitschrift 15, 1914, 3: 115–122 (received December 28, 1913) and 4: 169–176 (received 28 December 1913). Author's note: Greifswald, Physical Institute, December 24, 1913.

1. INTRODUCTION

In his interesting paper *Outline of a Generalized Theory of Relativity* etc.[1] Mr. Einstein says that by introducing a variable speed of light he has broken out of "the confines of the theory known at present as the theory of relativity," and also on other occasions he frequently contrasts his theory with the "old theory of relativity." To someone who immerses himself sufficiently deeply in the development it may become clear in what respect Mr. Einstein can speak of a new relativity; nevertheless, one may suppose that the cited passages may lead a more cursory reader to the wrong view, that one is really dealing here with a break with the theory of relativity as presently known. It is therefore certainly not without interest to present below, using

1 A. Einstein and M. Grossmann, *Entwurf einer verallgemeinerten Relativitätstheorie und einer Theorie der Gravitation*. Published as an offprint by B. G. Teubner, 1913.

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 4
Gravitation in the Twilight of Classical Physics: The Promise of Mathematics.

methods recently developed by me in a larger work about the theory of matter,[2] which strictly followed Minkowski's concept of the principle of relativity, a general theory of gravitation with a tensor potential, which includes Einstein's theory as a special case. Contrary to the rather inscrutable formulas of Einstein, the methods I use have the advantage that they yield clearly transparent expressions. In this fashion, it then becomes possible to comprehend the nature of Einstein's theory better, and in particular to clarify the so-called generalization of the principle of relativity, and furthermore to compare it with the theory of gravitation suggested by myself[3] and with
[116] that | of Nordström,[4] which deviates only slightly from mine.

2. GENERAL THEORY OF GRAVITATION WITH A TENSOR POTENTIAL

The essential difference between Einstein's theory of gravitation and my own is that in the former the gravitational field is described not by means of a four-vector, but by means of a spacetime quantity of third rank, which is related to a tensor (that is, a quantity of rank two) in the same way as a four-vector (a quantity of rank one) is related to a scalar (a quantity of rank zero). Because a tensor has 10 components, the gravitational field of Einstein has 40 components, which can be easily surveyed if each quadruple is associated with one component of a tensor, similar to associating the four components of a four-vector with a scalar. To make it intuitive I take the liberty of calling such a quantity a vector of tensors. I shall denote all the spatial components of this vector of tensors by $\mathfrak{g}$, and the temporal components by iu (cf. *Theory of Matter* III, p. 28). The four components that belong to the tensor component numbered by (μ, ν) are therefore: $\mathfrak{g}_{\mu\nu x}, \mathfrak{g}_{\mu\nu y}, \mathfrak{g}_{\mu\nu z}, iu_{\mu\nu}$ The indices μ, ν have to run over the values 1 to 4, where we have $\mathfrak{g}_{\mu\nu} = \mathfrak{g}_{\nu\mu}, u_{\mu\nu} = u_{\nu\mu}$. In addition to the vector of tensors $\mathfrak{g}_{\mu\nu}, iu_{\mu\nu}$ I introduce a second one, which I will denote by the letters $\mathfrak{f}$ and w (loc. cit. III, p. 28), that is $(\mathfrak{f}_{\mu\nu}, iw_{\mu\nu})$. This second vector of tensors shall describe the gravitational field equally well as the first; the two shall be related to each other in a similar way as the electric field strength is to the electric displacement, or the elastic stress is to the elastic strain. In an ideal vacuum, that is, at infinite distance from matter, they shall be equal, $(\mathfrak{f}_{\mu\nu}, iw_{\mu\nu}) = (\mathfrak{g}_{\mu\nu}, iu_{\mu\nu})$. In addition to the vector of tensors, for which we may choose any one of the two just named, the complete description of the state of the aether in a gravitational field requires another four dimensional tensor quantity, which I will denote by $\omega_{\mu\nu}$, and which one may call the gravitational potential. The denseness[1] of gravitational mass must be a tensor quantity as well in this theory, I will call its components $h_{\mu\nu}$ and for now make no further assumptions about how they are calculated from the state variables characterizing the

---

2 G. Mie, *Ann. d. Phys.*, Abhandlung I: 37, 511, 1912; Abhandlung II: 39, 1, 1912; Abhandlung III: 40, 1, 1913 [selections from I and III are included in this volume].

3 G. Mie, l. c. III, p. 25 ff.

4 G. Nordström, *Physik. Zeitschr.* 13, 1126, 1912; *Ann. d. Phys.* 40, 856, 1913. In the meantime Mr. Nordström has put up a second, quite different theory, which I shall briefly discuss at the end of the present paper.

material body. I now put down the following 50 equations for the 50 quantities $\mathfrak{g}_{\mu\nu}, i \cdot u_{\mu\nu}, \omega_{\mu\nu}$, by which the gravitational field is completely described (cf. 1. c. III, p. 28):

$$\left.\begin{aligned} \mathfrak{g}_{\mu\nu x} &= \frac{\partial \omega_{\mu\nu}}{\partial x} \\ \mathfrak{g}_{\mu\nu y} &= \frac{\partial \omega_{\mu\nu}}{\partial y} \\ \mathfrak{g}_{\mu\nu z} &= \frac{\partial \omega_{\mu\nu}}{\partial z} \\ u_{\mu\nu} &= -\frac{\partial \omega_{\mu\nu}}{\partial t} \end{aligned}\right\} \tag{1}$$

$$\frac{\partial \mathfrak{f}_{\mu\nu x}}{\partial x} + \frac{\partial \mathfrak{f}_{\mu\nu y}}{\partial y} + \frac{\partial \mathfrak{f}_{\mu\nu z}}{\partial z} + \frac{\partial w_{\mu\nu}}{\partial t} = -\kappa \cdot h_{\mu\nu}. \tag{2}$$

The quantity κ is a universal constant, which is denoted in the same way by Einstein, whereas I have used the letter γ for it in the *Theory of Matter*. Because eqs. (1) and (2) admit Lorentz transformations, the principle of relativity is satisfied in this theory. Incidentally, the eqs. (1) are equivalent to the following:

$$\left.\begin{aligned} \frac{\partial \mathfrak{g}_{\mu\nu x}}{\partial t} + \frac{\partial u_{\mu\nu}}{\partial x} &= 0 \\ \frac{\partial \mathfrak{g}_{\mu\nu y}}{\partial t} + \frac{\partial u_{\mu\nu}}{\partial y} &= 0 \\ \frac{\partial \mathfrak{g}_{\mu\nu z}}{\partial t} + \frac{\partial u_{\mu\nu}}{\partial z} &= 0 \end{aligned}\right\} \tag{3}$$

$$\frac{\partial \omega_{\mu\nu}}{\partial t} + u_{\mu\nu} = 0. \tag{4}$$

Multiplying eqs. (2) by $u_{\mu\nu}$, and eqs. (3) by $\mathfrak{f}_{\mu\nu x}, \mathfrak{f}_{\mu\nu y}, \mathfrak{f}_{\mu\nu z}$, and then adding it all, using (4), we have:

$$\operatorname{div} \sum u_{\mu\nu} \cdot \mathfrak{f}_{\mu\nu} + \sum \mathfrak{f}_{\mu\nu} \cdot \frac{\partial \mathfrak{g}_{\mu\nu}}{\partial t} + \sum u_{\mu\nu} \cdot \frac{\partial w_{\mu\nu}}{\partial t} - \kappa \cdot \sum h_{\mu\nu} \cdot \frac{\partial \omega_{\mu\nu}}{\partial t} = 0.$$

The summation symbols here are to be interpreted as summing over μ and ν from 1 to 4, as if the quantities numbered by μ, ν were different from those numbered by ν, μ. The sum $\Sigma u_{\mu\nu} \cdot \mathfrak{f}_{\mu\nu}$ yields an ordinary three dimensional vector that may be regarded as the energy flux (cf. loc. cit. III, p. 29). It is then easy to show that the energy principle is satisfied if there is a four dimensional scalar H, a function of all the quantities that determine the state of the aether, including also the quantities $\mathfrak{g}_{\mu\nu}, u_{\mu\nu}, \omega_{\mu\nu}$,

from which the other vector of tensors $(\mathfrak{f}_{\mu\nu}, iw_{\mu\nu})$ for the gravitational field can be
[117] derived as follows: |

$$\mathfrak{f}_{\mu\nu x} = \frac{\partial H}{\partial \mathfrak{g}_{\mu\nu x}}, \quad \mathfrak{f}_{\mu\nu y} = \frac{\partial H}{\partial \mathfrak{g}_{\mu\nu y}}, \quad \mathfrak{f}_{\mu\nu z} = \frac{\partial H}{\partial \mathfrak{g}_{\mu\nu z}}, \quad w_{\mu\nu} = -\frac{\partial H}{\partial u_{\mu\nu}} \tag{5}$$

and which in addition satisfies the differential equation

$$\frac{\partial H}{\partial \omega_{\mu\nu}} = -\kappa \cdot h_{\mu\nu}. \tag{6}$$

The proof of this assertion is extraordinarily simple; it proceeds in the same way as that for the corresponding theorem in the theory of gravitation with a scalar potential (loc. cit. III, p. 29, 30) and therefore there is no need to write it down here once again.

Incidentally, in all differentiations in (5) and (6) the quantities with the index (μ, ν) should be regarded as formally different from those with the index (ν, μ).

The quantity H is nothing other than the rest density of energy (or, equivalently, the rest density of inertial mass); in the general theory of matter it plays the role of the Hamiltonian function per unit volume, thus, for brevity, I usually call it the Hamiltonian function.

Now we want to calculate the force experienced by a mass particle in an external gravitational field $(\mathfrak{g}_{\mu\nu}, iu_{\mu\nu})$. The calculation for the gravitational field with tensor potential proceeds in exactly the same way as for the gravitational field with scalar potential (loc. cit. III, p. 35–40). For the calculation we presuppose that in the space surrounding the volume occupied by the mass particle the ratio of the vector $(\mathfrak{f}_{\mu\nu}, iw_{\mu\nu})$ to the vector $(\mathfrak{g}_{\mu\nu}, iu_{\mu\nu})$ may be regarded as constant. For the gravitational mass of the particle the calculation yields a quantity with the following 10 components:

$$m^g_{\mu\nu} = \int_V h_{\mu\nu} dV. \tag{7}$$

The integral should range over the entire volume V occupied by the mass particle. The force acting on the particle is calculated from the double sum:

$$\mathfrak{P} = \kappa \sum_{\nu=1}^{4} \sum_{\mu=1}^{4} \mathfrak{g}_{\mu\nu} m^g_{\mu\nu}. \tag{8}$$

3. IMPOSSIBILITY OF THE IDENTITY OF GRAVITATIONAL AND INERTIAL MASS

The density of inertial mass is identical with the density of energy, which is the (4,4) component of a symmetric tensor that I will, for brevity, call the energy tensor; therefore one can speak about identity or unity of essence[2] of the two masses only if the gravitational mass also occurs in the form of a tensor, a tensor that would have to be completely identical to the energy tensor. We have just learned in general terms the form of the basic equations for a theory in which the denseness of gravitational mass is a tensor: $(h_{\mu\nu})$. If we now denote the components of the energy tensor by $T_{\mu\nu}$, then *the principle of the identity of gravitational and inertial mass* is:

$$h_{\mu\nu} = T_{\mu\nu}. \tag{9}$$

If this principle were satisfied, then according to *Laue's theorem*[5] all components $m^g_{\mu\nu}$ of the gravitational mass of a material body at rest (a completely stationary system) would vanish, except for m^g_{44}, and, if I call the inertial mass of the body m_i (inertia), we would have $m^g_{44} = m_i$. The gravitational force acting on the body in the gravitational field $(\mathfrak{g}_{\mu\nu}, iu_{\mu\nu})$ would then be:

$$\mathfrak{P} = \kappa\mathfrak{g}_{44}m_i.$$

According to this, in one and the same gravitational field the forces of gravity acting on different bodies, which are in the same state of motion, would be strictly proportional to their inertial masses.

However, it is easy to show that the identity $h_{\mu\nu} = T_{\mu\nu}$ required by the principle as discussed is impossible. I will show that no function H of $\omega_{\mu\nu}$ can be found that satisfies the differential eqs. (6) in the form required by the principle of identity of the two masses:

$$\frac{\partial H}{\partial \omega_{\mu\nu}} = -\kappa T_{\mu\nu}. \tag{10}$$

If eq. (10) were satisfied, then the following would also be valid:

$$\frac{\partial T_{\mu\nu}}{\partial \omega_{\kappa\lambda}} = \frac{\partial T_{\kappa\lambda}}{\partial \omega_{\mu\nu}}. \tag{11}$$

Now I have shown in my paper on the theory of matter (Part III, p. 34, eq. (105)) that one can very simply write down the energy tensor by means of a quantity Φ, which like H is a function of the state of the aether at the spacetime point under consideration. Of course, when calculating with Φ one has to choose different quantities to describe the state of the aether than when calculating with H. Without regard to their physical meaning, the correctly chosen quantities will be ordered simply according to their association with the four coordinate axes and denoted | by [118]
x_1, x_2, x_3, x_4; y_1, y_2, y_3, y_4; z_1, z_2, z_3, z_4 etc. The components of the energy tensor can then be represented by the following schema:

5 M. v. Laue, *Das Relativitätsprinzip*, 2. ed., p. 209. Published by Vieweg & Sohn. Braunschweig 1913.

$$\left.\begin{aligned} T_{11} &= \Phi - \frac{\partial\Phi}{\partial x_1}x_1 - \frac{\partial\Phi}{\partial y_1}y_1 - \dots \\ T_{21} &= \quad - \frac{\partial\Phi}{\partial x_1}x_2 - \frac{\partial\Phi}{\partial y_1}y_2 - \dots . \end{aligned}\right\} \tag{12}$$

If we denote the state variables used in the calculation with H by $\xi_1, \xi_2, \xi_3, \xi_4$; $\eta_1, \eta_2, \eta_3, \eta_4$; $\zeta_1, \zeta_2, \zeta_3, \zeta_4$ etc., then the following differential equation follows from the definition of Φ (*Theory of Matter*, Part III, formulas (98) and (93) on page 32 and 30):

$$\frac{\partial}{\partial\omega_{\mu\nu}}H(\xi_1, \xi_2, \dots, \omega_{11}, \omega_{12}, \dots) = \frac{\partial}{\partial\omega_{\mu\nu}}\Phi(x_1, x_2, \dots, \omega_{11}, \omega_{12}, \dots).$$

Accordingly, by differentiating (12) and using (10) we obtain:

$$\begin{aligned} \frac{1}{\kappa}\frac{\partial T_{11}}{\partial\omega_{21}} &= -T_{21} + \frac{\partial T_{21}}{\partial x_1}x_1 + \frac{\partial T_{21}}{\partial y_1}y_1 + \dots \\ \frac{1}{\kappa}\frac{\partial T_{21}}{\partial\omega_{11}} &= \qquad\quad \frac{\partial T_{11}}{\partial x_1}x_2 + \frac{\partial T_{11}}{\partial y_1}y_2 + \dots . \end{aligned}$$

But it is easy to see that

$$\frac{\partial T_{11}}{\partial x_1}x_2 + \frac{\partial T_{11}}{\partial y_1}y_2 + \dots = \frac{\partial T_{21}}{\partial x_1}x_1 + \frac{\partial T_{21}}{\partial y_1}y_1 + \dots .$$

Therefore eq. (10) also leads to:

$$\frac{\partial T_{11}}{\partial\omega_{21}} + \kappa T_{21} = \frac{\partial T_{21}}{\partial\omega_{11}} . \tag{13}$$

Equation (11) and eq. (13) can be simultaneously satisfied only if $T_{21} = 0$. Exactly the same proof applies for any arbitrary $T_{\mu\nu}$ $(\mu \neq \nu)$. So the principle of identity of the two masses leads to the conclusion that all off-diagonal components of the energy tensor are equal to zero. But that would only be possible if the energy tensor were in reality a scalar. This proves the impossibility of the identity of the two masses.

The principle of identity of gravitational and inertial mass is impossible also in a theory in which the gravitational potential and the density of gravitational mass are four-dimensional tensors.

Indeed a glance at the formulas (15), (18) as well as (19) on p. 16 and 17 of Einstein's treatise shows that also in Einstein's theory the tensor for the denseness of gravitational mass found in formulas (15) and (18) is quite different from the energy tensor given by (19). It is therefore an error when Mr. Einstein speaks in the cited treatise of a "*physical unity of essence of gravitational and inertial mass*" in his theory, and of the validity of the "*equivalence hypothesis*," according to which "*the identity of gravitational and inertial mass is satisfied exactly*."

Nevertheless, there still remains the possibility that the gravitational and inertial mass in large bodies consisting of molecules can be made strictly proportional to each other by means of a series of auxiliary assumptions leading to a compensation of the deviations in the separate elementary particles when integrating over the whole volume. Indeed it seems to me that Mr. Einstein had only this remaining possibility in mind in his Vienna lecture,[6] when he postulated the equality of inertial and gravitational mass for "closed systems" (§ 2, postulate 2).

By itself it is probably rather immaterial whether one can, in such a somewhat artificial way, get the two masses to be mathematically exactly equal in closed systems, or whether they are only approximately equal, once one has abandoned the identity of the two masses in principle; whereby then, after all, the thoughts about general relativity of motion, of which Mr. Einstein spoke in such detail in his lecture, are abandoned as well.

Since Mr. Einstein puts so much weight on the validity of the theorem of the equality of the two masses in his theory, at least for closed systems, we are forced to go into the details of this theorem when discussing his theory.

4. SPECIAL ASSUMPTIONS OF EINSTEIN'S THEORY

In the cited paper Mr. Einstein uses the notation $g_{\mu\nu}$ for the components of the gravitational potential. The quantities that we denoted in eq. (1) by $\omega_{\mu\nu}$ differ from the $g_{\mu\nu}$ only by a constant factor:

$$g_{\mu\nu} = -2\kappa \cdot \omega_{\mu\nu}. \tag{14}$$

In a region infinitely distant from all matter, that is, in an ideal vacuum, the tensor $(g_{\mu\nu})$ is supposed to degenerate into the scalar -1, so that its several components take the following values: |

$$\begin{matrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{matrix}$$ [119]

Following Minkowski I always put the speed of light in an ideal vacuum equal to 1. In order to arrive at the special theory of Einstein one must make the following assumptions:

1. The Hamiltonian function H can be split into two terms, $H = H_e + H_g$, both of which depend on the components of the gravitational potential, that is, on the

6 *Physik. Zeitschr.* 14, 1249, 1913 [in this volume].

quantities $g_{\mu\nu}$; but additionally, the second one, H_g, contains only the components of the vector of tensors of the gravitational field $(\mathfrak{g}_{\mu\nu}, iu_{\mu\nu})$, and the first one, H_e, contains only the remaining state variables, e.g., the electromagnetic field quantities etc.

2. The dependence of the quantity H_e on the $g_{\mu\nu}$ shall take the form of the following expression:

$$\left.\begin{aligned} H_e &= \sqrt{g_{11}X_{11} + g_{22}X_{22} + \cdot\cdot + 2g_{12}X_{12} + \cdot\cdot} \\ H_e &= \sqrt{\sum_\mu \sum_\nu g_{\mu\nu}X_{\mu\nu}}\,, \end{aligned}\right\} \tag{15}$$

where the $X_{\mu\nu}$ no longer contain the $g_{\mu\nu}$, but only the other state variables of the aether. The $X_{\mu\nu}$ form the components of a four-dimensional tensor, H_e is then a four-dimensional scalar.

3. In the case that the material particle we consider is at rest, the tensor $(X_{\mu\nu})$ reduces to a tensor all of whose components are zero except X_{44}. Following Einstein we shall denote this value X_{44}^0 by $-\rho^2$:

$$X_{44}^0 = -\rho^2.$$

So the quantity ρ is a four-dimensional scalar. If the particle moves with velocity $\mathfrak{q}$, then we define in the usual way the following quantity $\mathfrak{V}$ as the velocity four vector:

$$\mathfrak{V}_1 = \frac{\mathfrak{q}_x}{\sqrt{1-\mathfrak{q}^2}}\,, \qquad \mathfrak{V}_2 = \frac{\mathfrak{q}_y}{\sqrt{1-\mathfrak{q}^2}}\,, \qquad \mathfrak{V}_3 = \frac{\mathfrak{q}_z}{\sqrt{1-\mathfrak{q}^2}}\,,$$

$$\mathfrak{V}_4 = \frac{i}{\sqrt{1-\mathfrak{q}^2}}\,, \qquad {\mathfrak{V}_1}^2 + {\mathfrak{V}_2}^2 + {\mathfrak{V}_3}^2 + {\mathfrak{V}_4}^2 = -1\,.$$

It then follows directly from the principle of relativity that the components of the tensor $(X_{\mu\nu})$ take the following values in the case that the particle we consider is in motion:

$$X_{\mu\nu} = +\rho^2\mathfrak{V}_\mu\mathfrak{V}_\nu. \tag{16}$$

Accordingly, the first term of the Hamiltonian function has the value:

$$\left.\begin{aligned} H_e &= \rho\sqrt{(g_{11}{\mathfrak{V}_1}^2 + g_{22}{\mathfrak{V}_2}^2 + \cdot\cdot + 2g_{12}\mathfrak{V}_1\mathfrak{V}_2 + \cdot\cdot)} \\ H_e &= \rho\sqrt{\sum_\mu \sum_\nu g_{\mu\nu}\mathfrak{V}_\mu\mathfrak{V}_\nu}\,, \end{aligned}\right\} \tag{17}$$

H_e and ρ are four-dimensional scalars as is the square root term in eq. (17),[3] so that we have:

$$\sum_{\mu}\sum_{\nu} g_{\mu\nu}\mathfrak{B}_{\mu}\mathfrak{B}_{\nu} = -g^{0}_{44}. \tag{18}$$

4. The second term of the Hamiltonian function H_g is a homogeneous function of second degree in the components of the vector of tensors $(\mathfrak{g}_{\mu\nu}, iu_{\mu\nu})$. If we write for simplicity the letters: $\mathfrak{g}_{\mu\nu1}, \mathfrak{g}_{\mu\nu2}, \mathfrak{g}_{\mu\nu3}, \mathfrak{g}_{\mu\nu4}$ for the quantities $\mathfrak{g}_{\mu\nu x}, \mathfrak{g}_{\mu\nu y}, \mathfrak{g}_{\mu\nu z}, iu_{\mu\nu}$, this means that H_g is an expression of the following form

$$H_g = \frac{1}{2}\cdot\sum_{\kappa,\,\mu,\,\lambda,\,\nu,\,\alpha,\,\beta} G_{\kappa\mu\lambda\nu\alpha\beta}\mathfrak{g}_{\kappa\lambda\alpha}\mathfrak{g}_{\mu\nu\beta}, \tag{19}$$

where each of the six indices is to be summed over separately from 1 to 4. The coefficients $G_{\kappa\mu\lambda\nu\alpha\beta}$ are functions of the gravitational potential, that is of the quantities $g_{\mu\nu}$, about which we have to make certain stipulations in order to come to Einstein's theory.

5. The tensor $(\gamma_{\mu\nu})$ defined by the following equations:

$$\left.\begin{aligned} g_{1\nu}\gamma_{1\nu}+g_{2\nu}\gamma_{2\nu}+g_{3\nu}\gamma_{3\nu}+g_{4\nu}\gamma_{4\nu} &= 1 \\ g_{1\mu}\gamma_{1\nu}+g_{2\mu}\gamma_{2\nu}+g_{3\mu}\gamma_{3\nu}+g_{4\mu}\gamma_{4\nu} &= 0, \qquad \mu\neq\nu \end{aligned}\right\} \tag{20}$$

is called the *inverse tensor* [*reziproker Tensor*] of $(g_{\mu\nu})$. Furthermore, let:

$$g = \begin{vmatrix} g_{11} & g_{12} & g_{13} & g_{14} \\ g_{21} & g_{22} & g_{23} & g_{24} \\ g_{31} & g_{32} & g_{33} & g_{34} \\ g_{41} & g_{42} & g_{43} & g_{44} \end{vmatrix}. \tag{21}$$

Then $\gamma_{\mu\nu}$ can be calculated as the cofactor [*adjungierte Unterdeterminante*] of $g_{\mu\nu}$ divided by g.[7] Let us now put:

$$G_{\kappa\mu\lambda\nu\alpha\beta} = 2\kappa\sqrt{g}\gamma_{\kappa\mu}\gamma_{\lambda\nu}\gamma_{\alpha\beta}. \tag{22}$$

5. THE FUNDAMENTAL EQUATIONS OF EINSTEIN'S THEORY

Before we substitute the expression, that results according to the above stipulations for H, into the general eqs. (1) and (2) in accordance with (5) and (6), we want to put down a few simple formulas which result directly from the eqs. (20) defining $\gamma_{\mu\nu}$, and which are very convenient for the following calculations. Differentiation of (20) with respect to any coordinate x_β yields:[4]

7 I take the sign of g as positive. Einstein has some signs different from me, because he puts $x_4 = t$ whereas I put $x_4 = it$.

$$g_{1\lambda}\frac{\partial\gamma_{1\nu}}{\partial x_\beta}+g_{2\lambda}\frac{\partial\gamma_{2\nu}}{\partial x_\beta}+g_{3\lambda}\frac{\partial\gamma_{3\nu}}{\partial x_\beta}+g_{4\lambda}\frac{\partial\gamma_{4\nu}}{\partial x_\beta} = -\left(\gamma_{1\nu}\frac{\partial g_{1\lambda}}{\partial x_\beta}+\gamma_{2\nu}\frac{\partial g_{2\lambda}}{\partial x_\beta}+\gamma_{3\nu}\frac{\partial g_{3\lambda}}{\partial x_\beta}+\gamma_{4\nu}\frac{\partial g_{4\lambda}}{\partial x_\beta}\right) \tag{23}$$

for $\lambda = \nu$ as well as for $\lambda \neq \nu$. |

[120] Now we write down the 4 eqs. (23) for a particular value ν, putting sequentially $\lambda = 1, 2, 3, 4$, we multiply each equation by $\gamma_{\mu\lambda}$, where μ is another constant number which may be different from or equal to ν, and then we add, with the result:

$$\frac{\partial\gamma_{\mu\nu}}{\partial x_\beta} = -\sum_\kappa\sum_\lambda \gamma_{\mu\lambda}\gamma_{\kappa\nu}\frac{\partial g_{\kappa\lambda}}{\partial x_\beta}. \tag{24}$$

By differentiating (20) with respect to any g_{mn}, treating g_{mn} and g_{nm} as different variables, we obtain:[5]

$$\left.\begin{aligned}\sum_{\kappa=1}^{4} g_{\kappa\lambda}\frac{\partial\gamma_{\kappa\nu}}{\partial g_{mn}} &= 0, \quad \lambda\neq n\\ \sum_{\kappa=1}^{4} g_{\kappa n}\frac{\partial\gamma_{\kappa\nu}}{\partial g_{mn}}+\gamma_{m\nu} &= 0.\end{aligned}\right\} \tag{25}$$

For fixed ν, we multiply the four equations for $\lambda = 1, 2, 3, 4$ by $\gamma_{\mu\lambda}$ and add to obtain:

$$\frac{\partial\gamma_{\mu\nu}}{\partial g_{mn}}+\gamma_{m\nu}\gamma_{\mu n} = 0. \tag{26}$$

Finally let us differentiate the determinant g with respect to g_{mn}, where again g_{mn} and g_{nm} are treated as different variables:

$$\frac{\partial g}{\partial g_{mn}} = g\cdot\gamma_{mn}. \tag{27}$$

From (27) it follows that:

$$\frac{\partial\sqrt{g}}{\partial g_{mn}} = \frac{1}{2}\sqrt{g}\gamma_{mn}. \tag{28}$$

After these preparations we now calculate first the vector of tensors $(\mathfrak{k}_{mn}, iw_{mn})$ and then the tensor of gravitational mass h_{mn}.

I will call the four quantities $\mathfrak{k}_{mnx}, \mathfrak{k}_{mny}, \mathfrak{k}_{mnz}, iw_{mn}$ for simplicity $\mathfrak{k}_{mn1}, \mathfrak{k}_{mn2}, \mathfrak{k}_{mn3}, \mathfrak{k}_{mn4}$. From (5) and (19) one then has:

$$\mathfrak{k}_{mn\alpha} = \frac{\partial H}{\partial\mathfrak{g}_{mn\alpha}} = \sum_{\mu,\nu,\beta} G_{m\mu n\nu\alpha\beta}\mathfrak{g}_{\mu\nu\beta}.$$

Here again the $\mathfrak{k}_{mn}$ and $\mathfrak{k}_{nm}$ as well as the $\mathfrak{g}_{mn}$ and $\mathfrak{g}_{nm}$ are regarded as different variables, but after the differentiation we set $G_{m\mu n\nu\alpha\beta} = G_{\mu m\nu n\beta\alpha}$. If we write, from (1) and (14),

$$\mathfrak{g}_{mn\beta} = -\frac{1}{2\kappa}\frac{\partial g_{mn}}{\partial x_\beta},$$

then we find, using formulas (22) and (24),

$$\mathfrak{k}_{mn\alpha} = \sum_{\beta=1}^{4}\sqrt{g}\gamma_{\alpha\beta}\frac{\partial\gamma_{mn}}{\partial x_\beta}. \tag{29}$$

This is the same vector of tensors appearing in Einstein's paper in the differential equation for the gravitational field, (15) and (18) on p. 16 and 17. It is easily seen that H_g can be written in the following form:

$$H_g = \frac{1}{2}\sum_{m,n,\alpha}\mathfrak{k}_{mn\alpha}\mathfrak{g}_{mn\alpha}. \tag{30}$$

Now we turn to the calculation of h_{mn}. From (6) and (14) we have:

$$h_{mn} = -\frac{1}{\kappa}\frac{\partial H}{\partial\omega_{mn}} = +2\frac{\partial H}{\partial g_{mn}}. \tag{31}$$

Since $H = H_e + H_g$, h_{mn} also splits into a sum: $h_{mn} = \overset{e}{h}_{mn} + \overset{g}{h}_{mn}$. From (16) and (17) we have

$$\left.\begin{aligned} \overset{e}{h}_{mn} &= \frac{\rho\mathfrak{V}_m\mathfrak{V}_n}{\sqrt{\sum_{\mu,\nu}g_{\mu\nu}\mathfrak{V}_\mu\mathfrak{V}_\nu}} \\ \overset{e}{h}_{mn} &= \frac{\rho}{\sqrt{-\overset{0}{g}_{44}}}\mathfrak{V}_m\mathfrak{V}_n. \end{aligned}\right\} \tag{32}$$

So $\overset{e}{h}_{mn}$ has the form $\overset{e}{h}_{mn} = h\mathfrak{V}_m\mathfrak{V}_n$, where $h = \rho/\sqrt{-\overset{0}{g}_{44}}$ is a four-dimensional scalar. To compare (32) with Einstein's expression for $\overset{e}{h}_{mn}$, I put (Einstein, eq. $(1'')$, p. 6):

$$\sqrt{\sum_{\mu,\nu}g_{\mu\nu}dx_\mu dx_\nu} = ds;$$

if furthermore dV is the volume element at the spacetime point under consideration, then the rest volume dV' of the mass particle occupying dV is

$$dV' = \frac{dV}{\sqrt{1-\mathfrak{q}^2}},$$

and I define:

$$dm = \rho dV' = \rho \frac{dV}{\sqrt{1-\mathfrak{q}^2}}.$$

Further, following Einstein (p. 10), I put

$$dV_0 = \sqrt{g} \cdot \frac{dV}{\frac{ds}{dt}},$$

and

$$\rho_0 = \frac{dm}{dV_0} = \frac{\rho}{\sqrt{g}\sqrt{1-\mathfrak{q}^2}} \frac{ds}{dt}.$$

We then have:

$$h^e_{mn} = \rho_0 \sqrt{g} \frac{dx_m}{ds} \frac{dx_n}{ds},$$

or, as in Einstein (p. 10),[8] putting

$$\Theta_{mn} = -\rho_0 \frac{dx_m}{ds} \frac{dx_n}{ds},$$

[121] I we have:

$$h^e_{mn} = -\sqrt{g} \cdot \Theta_{mn} \tag{33}$$

in agreement with Einstein's eqs. (15) and (18).

Now I calculate the second term of h_{mn}. From (19) and (31) we have

$$h^g_{mn} = \sum_{\kappa, \mu, \lambda, \nu, \alpha, \beta} \frac{\partial G_{\kappa\mu\lambda\nu\alpha\beta}}{\partial g_{mn}} \mathfrak{g}_{\kappa\lambda\alpha} \mathfrak{g}_{\mu\nu\beta},$$

where

$$G_{\kappa\mu\lambda\nu\alpha\beta} = 2\kappa\sqrt{g}\gamma_{\kappa\mu}\gamma_{\lambda\nu}\gamma_{\alpha\beta}.$$

When differentiating, g_{mn} and g_{nm} are to be treated as different. The formulas (26) and (28) yield:

$$\frac{\partial G_{\kappa\mu\lambda\nu\alpha\beta}}{\partial g_{mn}} = \kappa\sqrt{g}\gamma_{mn}\gamma_{\kappa\mu}\gamma_{\lambda\nu}\gamma_{\alpha\beta} - 2\kappa\sqrt{g}\gamma_{\alpha n}\gamma_{m\beta}\gamma_{\kappa\mu}\gamma_{\lambda\nu} -$$
$$-2\kappa\sqrt{g}(\gamma_{\alpha\beta}\gamma_{\lambda n}\gamma_{m\nu}\gamma_{\kappa\mu} + \gamma_{\alpha\beta}\gamma_{\kappa n}\gamma_{m\mu}\gamma_{\lambda\nu}).$$

---

8 I choose the negative sign to make Θ_{44} positive.

If one further puts

$$\mathfrak{g}_{\kappa\lambda\alpha} = -\frac{1}{2\kappa}\frac{\partial g_{\kappa\lambda}}{\partial x_\alpha}, \qquad \mathfrak{g}_{\mu\nu\beta} = -\frac{1}{2\kappa}\frac{\partial g_{\mu\nu}}{\partial x_\beta},$$

and notes that $g_{\mu\nu} = g_{\nu\mu}$, $\gamma_{\mu\nu} = \gamma_{\nu\mu}$, then, with the aid of formula (24), one finds:

$$\begin{aligned} h^g_{mn} = &-\frac{\sqrt{g}}{4\kappa}\sum_{\mu,\nu,\alpha,\beta}\gamma_{mn}\gamma_{\alpha\beta}\frac{\partial g_{\mu\nu}}{\partial x_\alpha}\frac{\partial\gamma_{\mu\nu}}{\partial x_\beta} + \frac{\sqrt{g}}{2\kappa}\sum_{\mu,\nu,\alpha,\beta}\gamma_{\alpha m}\gamma_{\beta n}\frac{\partial g_{\mu\nu}}{\partial x_\alpha}\frac{\partial\gamma_{\mu\nu}}{\partial x_\beta} \\ &+ \frac{\sqrt{g}}{\kappa}\sum_{\mu,\nu,\alpha,\beta}\gamma_{m\nu}\gamma_{\alpha\beta}\frac{\partial g_{\mu\nu}}{\partial x_\alpha}\frac{\partial\gamma_{\mu n}}{\partial x_\beta}, \end{aligned} \tag{34}$$

or from formula (23)

$$\begin{aligned} h^g_{mn} = &-\frac{\sqrt{g}}{4\kappa}\sum_{\mu,\nu,\alpha,\beta}\gamma_{mn}\gamma_{\alpha\beta}\frac{\partial g_{\mu\nu}}{\partial x_\alpha}\frac{\partial\gamma_{\mu\nu}}{\partial x_\beta} + \frac{\sqrt{g}}{2\kappa}\sum_{\mu,\nu,\alpha,\beta}\gamma_{\alpha m}\gamma_{\beta n}\frac{\partial g_{\mu\nu}}{\partial x_\alpha}\frac{\partial\gamma_{\mu\nu}}{\partial x_\beta} \\ &- \frac{\sqrt{g}}{\kappa}\sum_{\mu,\nu,\alpha,\beta}g_{\mu\nu}\gamma_{\alpha\beta}\frac{\partial\gamma_{m\nu}}{\partial x_\alpha}\frac{\partial\gamma_{\mu n}}{\partial x_\beta}. \end{aligned}$$

This last form is used by Einstein; he collects together the first two terms calling their sum $-\sqrt{g}\cdot\vartheta_{mn}$ (Einstein, p. 15, formula (13)), and writes:

$$h^g_{mn} = -\sqrt{g}\left(\vartheta_{mn} + \frac{1}{\kappa}\sum_{\mu,\nu,\alpha,\beta}\gamma_{\alpha\beta}g_{\mu\nu}\frac{\partial\gamma_{m\nu}}{\partial x_\alpha}\frac{\partial\gamma_{\mu n}}{\partial x_\beta}\right). \tag{35}$$

Written in Einstein's notation, the density of the gravitational mass is then, taken together:

$$h_{mn} = -\sqrt{g}\left(\Theta_{mn} + \vartheta_{mn} + \frac{1}{\kappa}\sum_{\mu,\nu,\alpha,\beta}\gamma_{\alpha\beta}g_{\mu\nu}\frac{\partial\gamma_{m\nu}}{\partial x_\alpha}\frac{\partial\gamma_{\mu n}}{\partial x_\beta}\right). \tag{36}$$

Taking into account formula (29),

$$\mathfrak{f}_{mn\alpha} = \sum_\beta \sqrt{g}\gamma_{\alpha\beta}\frac{\partial\gamma_{mn}}{\partial x_\beta},$$

one sees immediately that Einstein's fundamental eq. (15) and (18) on p. 16 and p. 17 is nothing other than our eq. (2):

$$\sum_\alpha \frac{\partial\mathfrak{f}_{mn\alpha}}{\partial x_\alpha} = -\kappa h_{mn}.$$

Einstein's theory of gravitation is a special case of the general theory of gravitation with a tensor potential described in section 2. Thus it fits perfectly into the framework of the ordinary theory of relativity.

6. THE ENERGY TENSOR

In order to set up the energy tensor, Einstein adds another auxiliary assumption to those enumerated in Section 4, which will become particularly important in the course of our investigation. Let us define a new four-vector $\mathfrak{W}$ as follows:

$$\left.\begin{aligned} \mathfrak{W}_\nu &= \frac{\partial H_e}{\partial \mathrm{V}_\nu} = \frac{\rho \sum\limits_\mu g_{\mu\nu}\mathfrak{V}_\mu}{\sqrt{\sum\limits_\mu \sum\limits_\nu g_{\mu\nu}\mathfrak{V}_\mu \mathfrak{V}_\nu}}, \\ \mathfrak{W}_\nu &= \frac{\rho}{\sqrt{-g_{44}^0}} \sum_{\mu=1}^{4} g_{\mu\nu}\mathfrak{V}_\mu \,. \end{aligned}\right\} \tag{37}$$

The new auxiliary assumption of Einstein is:

$$\begin{aligned} &\frac{\partial}{\partial x}(\mathfrak{W}_\mu \cdot \mathfrak{V}_1) + \frac{\partial}{\partial y}(\mathfrak{W}_\mu \cdot \mathfrak{V}_2) + \frac{\partial}{\partial z}(\mathfrak{W}_\mu \cdot \mathfrak{V}_3) - i\frac{\partial}{\partial t}(\mathfrak{W}_\mu \cdot \mathfrak{V}_4) \\ &\qquad\qquad = -\kappa \sum_{m,n} \mathfrak{g}_{mn\mu} h_{mn}^e = \sum_{m,n} \frac{\partial H_e}{\partial g_{mn}} \frac{\partial g_{mn}}{\partial x_\mu}. \end{aligned} \tag{38}$$

Einstein calls this auxiliary assumption the law of energy-momentum conservation. It is found as eq. (10) on p. 10 in his treatise, and there takes the form

$$\sum_{\mu,\nu} \frac{\partial}{\partial x_\nu}(\sqrt{g} g_{\sigma\mu}\Theta_{\mu\nu}) - \frac{1}{2}\sum \sqrt{g}\frac{\partial g_{\mu\nu}}{\partial x_\sigma}\Theta_{\mu\nu} = 0. \quad |$$

[122] Noting that as a consequence of our eqs. (32) and (33):

$$\mathfrak{W}_\sigma \mathfrak{V}_\nu = -\sqrt{g}\sum_\mu g_{\sigma\mu}\Theta_{\mu\nu},$$

and that according to (31) and (33)

$$\frac{\partial H_e}{\partial g_{\mu\nu}} = -\frac{1}{2}\sqrt{g}\Theta_{\mu\nu},$$

one sees immediately that this equation is identical with (38).

Assuming (38) to be correct, one easily finds that the components of a tensor defined as follows:

$$\left.\begin{aligned} T_{\alpha\alpha} &= \mathfrak{W}_\alpha \mathfrak{V}_\alpha + H_g - \sum_\mu \sum_\nu \mathfrak{g}_{\mu\nu\alpha}\mathfrak{f}_{\mu\nu\alpha}, \\ T_{\alpha\beta} &= \mathfrak{W}_\alpha \mathfrak{V}_\beta - \sum_\mu \sum_\nu \mathfrak{g}_{\mu\nu\alpha}\mathfrak{f}_{\mu\nu\beta} \end{aligned}\right\} \tag{39}$$

satisfy the differential equations

$$\frac{\partial T_{\alpha 1}}{\partial x}+\frac{\partial T_{\alpha 2}}{\partial y}+\frac{\partial T_{\alpha 3}}{\partial z}-i\frac{\partial T_{\alpha 4}}{\partial t} = 0; \tag{40}$$

for the proof one uses eqs. (38), (1), (2), (5), (6). So the tensor (39) is the energy tensor, the eqs. (40) are the energy-momentum equations. A bit of calculation shows that the eqs. (40) are identical with the eqs. (19) on p. 17 of Einstein:

$$\sum_{\mu,\nu}\frac{\partial}{\partial x_\nu}(\sqrt{g}g_{\sigma\mu}(\Theta_{\mu\nu}+\vartheta_{\mu\nu})) = 0, \qquad (\sigma = 1, 2, 3, 4.)$$

In a gravitation-free space, where $g_{\nu\nu} = -1$, $g_{\mu\nu} = 0$, $\mathfrak{g}_{\mu\nu\alpha} = 0$, the components of the energy tensor take on the following values:

$$T_{\alpha\alpha}{}' = -\rho\mathfrak{V}_\alpha{}^2, \qquad T_{\alpha\beta}{}' = -\rho\mathfrak{V}_\alpha\mathfrak{V}_\beta.$$

Thus, according to eqs. (16), the tensor $(X_{\mu\nu})$ introduced on p. 17 differs from the energy tensor $(T_{\mu\nu}{}')$ in gravitation free space only by the factor $-\rho$,

$$X_{\mu\nu} = -\rho T_{\mu\nu}{}'.$$

By decomposing the components of the energy tensor into two terms in the same way as $H = H_e + H_g$,

$$T_{\alpha\beta} = T^e_{\alpha\beta}+T^g_{\alpha\beta},$$

where

$$T^e_{\alpha\beta} = \mathfrak{W}_\alpha\mathfrak{V}_\beta$$

and

$$T^g_{\alpha\beta} = -\sum_{\mu,\nu}\mathfrak{g}_{\mu\nu\alpha}\mathfrak{k}_{\mu\nu\beta},$$

or

$$T^g_{\alpha\alpha} = H_g-\sum_{\mu,\nu}\mathfrak{g}_{\mu\nu\alpha}\mathfrak{k}_{\mu\nu\alpha},$$

one notes directly that (16) and (37) imply the theorem:

$$H_e = T^e_{11}+T^e_{22}+T^e_{33}+T^e_{44}. \tag{41}$$

The auxiliary assumptions (16) and (38) of Einstein's theory imply that that part of the Hamiltonian function H_e which does not contain the field strength of gravity becomes identical to the sum of the diagonal terms of the part of the energy tensor that is devoid of gravitational field strength.

[169] 7. THEOREM OF THE RELATIVITY OF THE GRAVITATIONAL POTENTIAL

The assumptions of Einstein's theory are perhaps in part of rather secondary importance, but they are all made according to one principle, namely that the resulting equations admit other linear transformations in addition to the Lorentz transformations. I believe one may characterize the two propositions:

1. *that the gravitational potential is a four-dimensional tensor,*
2. *that the general equations of the aether dynamics admit other linear transformations in addition to those of Lorentz,*

as the two essential or main assumptions of Einstein's theory, compared to which the other assumptions play a subsidiary role as auxiliary assumptions.

To grasp the true nature of the second proposition, which Einstein regards as a generalization of the principle of relativity, it will be necessary to go into it rather precisely, although in doing so repetition of several calculations of Messrs. A. Einstein and M. Grossmann will be unavoidable.

Let us imagine a material system located in empty space, far distant from all other matter, that is at a place where the gravitational potential has the scalar value -1, and that we have complete knowledge of the processes and their laws in this system. Further, we imagine this same material system transported into the vicinity of a very large body, the Earth for example, where the gravitational potential is no longer equal to -1, but instead is represented by a tensor. Because the gravitational potential enters into the function H and thereby also into the equations of the aether dynamics, presumably all processes in the material system are influenced by the mere presence of a gravitational potential that differs from -1. Now the question is, what is the nature of this influence of the gravitational potential. This question is one that I, too, have already asked in my theory of the scalar gravitational potential, and I have given an answer for the case of that theory (loc. cit. III, p. 61ff).

At the second location the field strength of gravitation shall also be so small that the changes of the gravitational potential do not reach any appreciable values at the boundaries of a region containing the material system and extending infinitely far in comparison with that system; that is, the gravitational potential may be considered constant on the boundary. Let us denote by $(\overset{1}{g}_{\mu\nu})$ this constant value of the gravitational potential at infinite distance from the material system being considered. By the following equations we will then define 16 transformation coefficients $a_{\mu\nu}$, which
[170] together form a four dimensional and in general asymmetric | matrix $(a_{\mu\nu} \neq a_{\nu\mu})$:

$$a_{\mu 1}a_{\nu 1} + a_{\mu 2}a_{\nu 2} + a_{\mu 3}a_{\nu 3} + a_{\mu 4}a_{\nu 4} = -\overset{1}{g}_{\mu\nu}. \tag{42}$$

Since these are only 10 equations, 6 of the coefficients $a_{\mu\nu}$ can of course be chosen arbitrarily. I will denote the inverse matrix of the matrix $(a_{\mu\nu})$ by $(\alpha_{\mu\nu})$, so the elements $\alpha_{\mu\nu}$ are defined as follows:

$$\left.\begin{aligned} a_{1\mu}\alpha_{1\mu} + a_{2\mu}\alpha_{2\mu} + a_{3\mu}\alpha_{3\mu} + a_{4\mu}\alpha_{4\mu} &= 1, \\ a_{1\mu}\alpha_{1\nu} + a_{2\mu}\alpha_{2\nu} + a_{3\mu}\alpha_{3\nu} + a_{4\mu}\alpha_{4\nu} &= 0, \qquad \mu \neq \nu \end{aligned}\right\} \tag{43}$$

Now I introduce in place of the components of the gravitational potential $g_{\mu\nu}$ in the interior and in the closer vicinity of said material system the following linear functions of the $g_{\mu\nu}$, which I will call $g_{\mu\nu}'$:

$$g_{\mu\nu}' = \sum_{\kappa,\lambda} \alpha_{\kappa\mu}\alpha_{\lambda\nu} g_{\kappa\lambda}. \tag{44}$$

The ten quantities $g_{\mu\nu}'$ taken together again form a four-dimensional tensor, which originated from the tensor $(g_{\mu\nu})$ by deformation and rotation, as it were; at infinity, where $(g_{\mu\nu})$ reaches $(\overset{1}{g}_{\mu\nu})$, $(g_{\mu\nu}')$ becomes, by formulas (42) and (43), the scalar -1. Further I calculate $(\gamma_{\mu\nu}')$, the inverse tensor to $(g_{\mu\nu}')$, defined by:

$$\left.\begin{aligned} g_{\mu 1}'\gamma_{\mu 1}' + g_{\mu 2}'\gamma_{\mu 2}' + g_{\mu 3}'\gamma_{\mu 3}' + g_{\mu 4}'\gamma_{\mu 4}' &= 1, \\ g_{\mu 1}'\gamma_{\nu 1}' + g_{\mu 2}'\gamma_{\nu 2}' + g_{\mu 3}'\gamma_{\nu 3}' + g_{\mu 4}'\gamma_{\nu 4}' &= 0, \;\; \mu \neq \nu. \end{aligned}\right\}$$

It is easily seen that the $\gamma_{\mu\nu}'$ can be represented as linear functions of the components $\gamma_{\mu\nu}$ of the inverse tensor to $g_{\mu\nu}$:

$$\gamma_{\mu\nu}' = \sum_{\kappa,\lambda} a_{\kappa\mu} a_{\lambda\nu} \gamma_{\kappa\lambda}. \tag{45}$$

Using formulas (43) one can easily verify the equations of definition of the $\gamma_{\mu\nu}'$. Conversely, if one wants to calculate the $g_{\mu\nu}$ from the $g_{\mu\nu}'$ and the $\gamma_{\mu\nu}$ from the $\gamma_{\mu\nu}'$, one has:

$$g_{\mu\nu} = \sum_{\kappa,\lambda} a_{\mu\kappa} a_{\nu\lambda} g_{\kappa\lambda}', \qquad \gamma_{\mu\nu} = \sum_{\kappa,\lambda} \alpha_{\mu\kappa}\alpha_{\nu\lambda}\gamma_{\kappa\lambda}'. \tag{46}$$

In place of the rectangular coordinate system (x_1, x_2, x_3, x_4) we introduce further an oblique-angled one (x_1', x_2', x_3', x_4'), which moreover has different units of lengths on the different coordinate axes, by making the following substitutions:

$$x_\nu = \alpha_{\nu 1}x_1' + \alpha_{\nu 2}x_2' + \alpha_{\nu 3}x_3' + \alpha_{\nu 4}x_4'. \tag{47}$$

We then have:

$$x_\nu' = a_{1\nu}x_1 + a_{2\nu}x_2 + a_{3\nu}x_3 + a_{4\nu}x_4. \tag{48}$$

Further we denote by g the determinant of the $g_{\mu\nu}$, as in eq. (21) above, similarly by g_1 and g' the determinants of the $\overset{1}{g}_{\mu\nu}$ and the $g_{\mu\nu}'$. Then it follows directly from (42) and (46) that:

$$g = g_1 g'. \tag{49}$$

We further define:

$$\mathfrak{g}_{\mu\nu\alpha}{}' = -\frac{1}{2\kappa}\frac{\partial g_{\mu\nu}{}'}{\partial x_\alpha{}'}, \qquad \mathfrak{f}_{\mu\nu\alpha}{}' = \sqrt{g'}\sum_\beta \gamma_{\alpha\beta}{}'\frac{\partial \gamma_{\mu\nu}{}'}{\partial x_\beta{}'}, \tag{50}$$

and by analogy to (30):

$$H_g{}' = \frac{1}{2}\sum_{\mu,\nu,\alpha}\mathfrak{g}_{\mu\nu\alpha}{}'\mathfrak{f}_{\mu\nu\alpha}{}'. \tag{51}$$

By a simple calculation it can then be shown that:

$$H_g{}' = \frac{1}{\sqrt{g_1}}H_g. \tag{52}$$

If we regard $H_g{}'$ as a function of the transformed quantities $g_{\mu\nu}{}'$ and $\mathfrak{g}_{\mu\nu\alpha}{}'$, then (51) implies:

$$\mathfrak{f}_{\mu\nu\alpha}{}' = \frac{\partial H_g{}'}{\partial \mathfrak{g}_{\mu\nu\alpha}{}'}. \tag{53}$$

Next we introduce a new velocity vector $\mathfrak{V}'$, obtained from $\mathfrak{V}$ by the following transformation equations:

$$\frac{\mathfrak{V}_\mu}{s} = \alpha_{\mu 1}\mathfrak{V}_1{}' + \alpha_{\mu 2}\mathfrak{V}_2{}' + \alpha_{\mu 3}\mathfrak{V}_3{}' + \alpha_{\mu 4}\mathfrak{V}_4{}', \tag{54}$$

or:

$$s\mathfrak{V}_\mu{}' = a_{1\mu}\mathfrak{V}_1 + a_{2\mu}\mathfrak{V}_2 + a_{3\mu}\mathfrak{V}_3 + a_{4\mu}\mathfrak{V}_4, \tag{55}$$

where s is to denote the following quantity:

$$s^2 = \sum_{\mu,\nu}\overset{1}{g}_{\mu\nu}\mathfrak{V}_\mu\mathfrak{V}_\nu = \frac{1}{\sum_{\mu,\nu}\overset{1}{\gamma}_{\mu\nu}\mathfrak{V}_\mu{}'\mathfrak{V}_\nu{}'}. \tag{56}$$

By squaring and adding the eqs. (55) one finds, taking note of (42):

$$\mathfrak{V}_1{}'^2 + \mathfrak{V}_2{}'^2 + \mathfrak{V}_3{}'^2 + \mathfrak{V}_4{}'^2 = -1. \tag{57}$$

It is easily seen that

$$\sum_{\mu,\nu} g_{\mu\nu}\mathfrak{V}_\mu\mathfrak{V}_\nu = s^2\sum_{\mu,\nu} g_{\mu\nu}{}'\mathfrak{V}_\mu{}'\mathfrak{V}_\nu{}',$$

hence by (16):

$$H_e = \rho\, s\sqrt{\sum_{\mu,\nu} g_{\mu\nu}{}'\mathfrak{V}_\mu{}'\mathfrak{V}_\nu{}'}.$$

Let us next define:

$$\rho' = \frac{\rho\, s}{\sqrt{g_1}} = \frac{\rho}{\sqrt{g_1}}\sqrt{\sum_{\mu,\nu} g^1_{\mu\nu}\mathfrak{V}_\mu\mathfrak{V}_\nu}\,, \tag{58}$$

$$H_e' = \rho'\sqrt{\sum_{\mu,\nu} g_{\mu\nu}'\mathfrak{V}_\mu'\mathfrak{V}_\nu'}\,, \tag{59}$$ [171]

then we have:

$$H_e' = \frac{1}{\sqrt{g_1}}H_e, \tag{60}$$

and from (52):

$$H' = H_e' + H_g' = \frac{1}{\sqrt{g_1}}H. \tag{61}$$

Now we define, by analogy to (37):

$$\mathfrak{W}_\nu' = \frac{\partial H_e'}{\partial \mathfrak{V}_\nu'} = \frac{\rho'\sum_\mu g_{\mu\nu}'\mathfrak{V}_\mu'}{\sqrt{\sum_{\mu,\nu} g_{\mu\nu}'\mathfrak{V}_\mu'\mathfrak{V}_\nu'}}, \tag{62}$$

then a simple calculation shows that:

$$\mathfrak{W}_\nu' = \frac{s}{\sqrt{g_1}}(\alpha_{1\nu}\mathfrak{W}_1 + \alpha_{2\nu}\mathfrak{W}_2 + \alpha_{3\nu}\mathfrak{W}_3 + \alpha_{4\nu}\mathfrak{W}_4). \tag{63}$$

After these preliminaries we now quickly come to the conclusion. From the definition (53) of the $\mathfrak{f}_{\mu\nu\alpha}'$ it follows that:

$$\sum_\alpha \frac{\partial \mathfrak{f}_{\mu\nu\alpha}'}{\partial x'_\alpha} = \frac{1}{\sqrt{g_1}}\sum_{\kappa,\lambda} a_{\kappa\mu}a_{\lambda\nu}\sum_\alpha \frac{\partial \mathfrak{f}_{\kappa\lambda\alpha}}{\partial x_\alpha}.$$

Substituting according to (2), (6), (14):

$$\sum_\alpha \frac{\partial \mathfrak{f}_{\kappa\lambda\alpha}}{\partial x_\alpha} = -\kappa h_{\kappa\lambda} = -2\kappa\frac{\partial H}{\partial g_{\kappa\lambda}},$$

by virtue of (46) and (61) results in:

$$\frac{\partial \mathfrak{f}_{\mu\nu x}'}{\partial x'} + \frac{\partial \mathfrak{f}_{\mu\nu y}'}{\partial y'} + \frac{\partial \mathfrak{f}_{\mu\nu z}'}{\partial z'} + \frac{\partial w_{\mu\nu}'}{\partial t'} = -\kappa h_{\mu\nu}', \qquad h_{\mu\nu}' = 2\frac{\partial H'}{\partial g_{\mu\nu}'}. \tag{64}$$

The definition of the $\mathfrak{W}_\nu'$ and $\mathfrak{V}_\nu'$, (55) and (63) implies:

$$\sum_{\nu} \frac{\partial}{\partial x_{\nu}'}(\mathfrak{W}_{\mu}' \cdot \mathfrak{B}_{\nu}') = \frac{1}{\sqrt{g_1}} \sum_{\lambda} \alpha_{\lambda\mu} \sum_{\nu} \frac{\partial}{\partial x_{\nu}}(\mathfrak{W}_{\lambda} \cdot \mathfrak{B}_{\nu}).$$

From this equation a small calculation according to (38) gives:

$$\left.\begin{aligned} \frac{\partial}{\partial x'}(\mathfrak{W}_{\mu}' \cdot \mathfrak{B}_1') + \frac{\partial}{\partial y'}(\mathfrak{W}_{\mu}' \cdot \mathfrak{B}_2') + \frac{\partial}{\partial z'}(\mathfrak{W}_{\mu}' \cdot \mathfrak{B}_3') - i\frac{\partial}{\partial t'}(\mathfrak{W}_{\mu}' \cdot \mathfrak{B}_4') \\ = \sum_{m,\,n} \frac{\partial H_e'}{\partial g_{mn}'} \frac{\partial g_{mn}'}{\partial x_{\mu}'}. \end{aligned}\right\} \quad (65)$$

We have now obtained each and every equation of Einstein's theory of gravitation in terms of the transformed quantities; and it turned out that in terms of the latter they have exactly the same form as the original equations in the non-transformed quantities. Considering that the transformed gravitational potential $(g_{\mu\nu}')$ has the scalar value -1 at infinity, whereas the non-transformed $(g_{\mu\nu})$ becomes $(g^1_{\mu\nu})$ at infinity, we see that the transformation property just proved signifies the same as the following theorem:

The Theorem of the Relativity of the Gravitational Potential. *If two empty spaces differ only in that the gravitational potential in one of them has the scalar value* -1, *but in the other an arbitrary tensor value* $g^1_{\mu\nu}$, *then all physical processes in the two spaces proceed in exactly the same fashion, provided that space and time in the first space is specified by means of an ordinary orthogonal system of coordinates* (x, y, z, it), *whereas in the second they are specified by means of a certain oblique system of coordinates* (x', y', z', it') *defined by the eqs. (42) and (48).*

From this it is clearly seen that the "generalized theorem of relativity" plays exactly the same role in Einstein's theory as what I call the "theorem of the relativity of the gravitational potential" (loc. cit. III, p. 61) does in my theory. However, the transformations in Einstein's theory are much more complicated than the extremely simple transformation that is valid in my theory, which is represented by the formulas on p. 63 of my treatise III. This is obvious because calculation with tensors is generally more complicated than with scalars. However, in one respect the difference of Einstein's relativity theorem as opposed to mine is really significant. Whereas only the quantities that specify the state of the aether are transformed in my theory, while the coordinates remain unchanged, in Einstein's theory the coordinates are transformed as well. Therefore the transformations of the two theorems of relativity, that of motion and that of gravitational potential, are very similar to each other in Einstein's theory, and that is probably the reason why Einstein could initially regard his theorem as a generalization of the principle of relativity of motion.[9] From the point of
[172] | view of a physicist the main distinction between the two theories is that by my relativity theorem the laws of nature are not affected at all by the gravitational potential, whereas by Einstein's relativity theorem they are affected as if the units of length and time are changed by the presence of a gravitational potential that differs from -1. So,

according to Einstein's theory the speed of light, the frequencies of spectral lines, the dimensions of atoms and of bodies composed of them, are supposed to change with the gravitational potential, whereas according to my theory nothing of all this should be observable.

8. EQUALITY OF INERTIAL AND GRAVITATIONAL MASS OF CLOSED SYSTEMS IN THE THEORIES OF EINSTEIN AND MIE

Einstein has shown in the report on his lecture held at the Vienna Naturforscherversammlung how one can prove that the two masses of closed systems are equal.[10] If one substitutes the values (32) for h^e_{mn} and (34) for h^g_{mn} into the fundamental eqs. (2) and simultaneously notes (29), these equations become:

$$\sum_{\alpha}\frac{\partial \mathfrak{k}_{mn\alpha}}{\partial x_\alpha} = -\kappa\frac{\rho}{\sqrt{-g^0_{44}}}\mathfrak{V}_m\mathfrak{V}_n + \kappa\sum_{\mu,\nu,\alpha}\left(\gamma_{\alpha m}\mathfrak{g}_{\mu\nu\alpha}\mathfrak{k}_{\mu\nu n} - \frac{1}{2}\gamma_{mn}\mathfrak{g}_{\mu\nu\alpha}\mathfrak{k}_{\mu\nu\alpha}\right)$$
$$-\sum_{\mu,\nu,\alpha}\gamma_{m\nu}\frac{\partial g_{\mu\nu}}{\partial x_\alpha}\mathfrak{k}_{\mu n\alpha}.$$

Now take the four equations that correspond to a fixed value of n, as m ranges over the sequence of numbers $1, 2, 3, 4$, multiply each equations by g_{mn}, and add. This yields, with the use of (20):

$$\sum_{\alpha}\frac{\partial}{\partial x_\alpha}\sum_{m} g_{mn}\mathfrak{k}_{mn\alpha} = -\kappa\left(\mathfrak{W}_n\mathfrak{V}_n + H_g - \sum_{\mu,\nu}\mathfrak{k}_{\mu\nu n}\,\mathfrak{g}_{\mu\nu n}\right)$$

or from (39):

$$\sum_{\alpha}\frac{\partial}{\partial x_\alpha}\sum_{m} g_{mn}\mathfrak{k}_{mn\alpha} = -\kappa T_{nn}.$$

Similarly by multiplying the four equations one after the other by g_{mp} and adding one obtains:

$$\sum_{\alpha}\frac{\partial}{\partial x_\alpha}\sum_{m} g_{mp}\mathfrak{k}_{mn\alpha} = -\kappa\left(\mathfrak{W}_p\mathfrak{V}_n - \sum_{\mu,\nu}\mathfrak{g}_{\mu\nu p}\mathfrak{k}_{\mu\nu n}\right) = -\kappa T_{pn}.$$

---

9 In the introduction to the treatise *Outline of a Generalized Theory of Relativity* etc., Mr. Einstein states the hypothesis "that a homogeneous gravitational field can physically be completely replaced by a state of acceleration of the reference system."[6] Apparently he has the mistaken notion, that this hypothesis (the equivalence hypothesis) is the foundation of the theory developed by him. That would indeed be a more general relativity of motion. In his Vienna lecture Mr. Einstein only demands of the theory of gravitation that the "observable laws of nature do not depend on the absolute magnitude of the gravitational potential" (postulate 4, this journal 14, 1250, 1913) [in this volume]. That is the relativity principle of the gravitational potential, which is what Einstein's theory really satisfies.

10 This journal, 14, 1258, eq. (7b), 1913 [in this volume].

So the following holds in general:

$$\sum_{\alpha} \frac{\partial}{\partial x_{\alpha}} \sum_{m} g_{mp} \mathfrak{f}_{mn\alpha} = -\kappa T_{pn}\,, \quad p = n \quad \text{or} \quad p \neq n. \tag{66}$$

By integrating the eqs. (66) over the volume occupied by the closed system and over a time during which the components of the gravitational as well as the inertial mass of the system do not change, one obtains:

$$\overset{1}{g}_{1\mu} \overset{g}{m}_{1\nu} + \overset{1}{g}_{2\mu} \overset{g}{m}_{2\nu} + \overset{1}{g}_{3\mu} \overset{g}{m}_{3\nu} + \overset{1}{g}_{4\mu} \overset{g}{m}_{4\nu} = \overset{i}{m}_{\mu\nu}. \tag{67}$$

Here $\overset{i}{m}_{\mu\nu}$, the (μ, ν) component of the inertial mass, denotes the following integral:

$$\overset{i}{m}_{\mu\nu} = \int T_{\mu\nu} dV, \tag{68}$$

and further $\overset{1}{g}_{\mu\nu}$ denote the components of the gravitational potential on the boundary of the volume occupied be the closed system. As in Section 7 (p. 169), we assume that the potential $(\overset{1}{g}_{\mu\nu})$ can be regarded as constant on the entire boundary surface.

The following equations are easily derived from (67):

$$\overset{g}{m}_{\mu\nu} = \overset{1}{\gamma}_{\mu 1} \overset{i}{m}_{1\nu} + \overset{1}{\gamma}_{\mu 2} \overset{i}{m}_{2\nu} + \overset{1}{\gamma}_{\mu 3} \overset{i}{m}_{3\nu} + \overset{1}{\gamma}_{\mu 4} \overset{i}{m}_{4\nu}. \tag{69}$$

Let us for brevity call the component $\overset{i}{m}_{44}$ the inertial mass m_i of the system:

$$m_i = \int T_{44} dV. \tag{70}$$

If the system moves through space with velocity $\mathfrak{q}$ it is easy to derive from *Laue's theorem* that

$$\left.\begin{aligned} \overset{i}{m}_{\mu\nu} &= -m_i \mathfrak{q}_{\mu} \mathfrak{q}_{\nu}, \\ \mathfrak{q}_1 = \mathfrak{q}_x, \quad \mathfrak{q}_2 &= \mathfrak{q}_y, \quad \mathfrak{q}_3 = \mathfrak{q}_z, \quad \mathfrak{q}_4 = i. \end{aligned}\right\} \tag{71}$$

And it follows from (69) that:

$$\left.\begin{aligned} \overset{g}{m}_{\mu\nu} &= -m_i \mathfrak{p}_{\mu} \mathfrak{q}_{\nu}, \\ \mathfrak{p}_{\mu} &= \overset{1}{\gamma}_{\mu 1} \mathfrak{q}_1 + \overset{1}{\gamma}_{\mu 2} \mathfrak{q}_2 + \overset{1}{\gamma}_{\mu 3} \mathfrak{q}_3 + \overset{1}{\gamma}_{\mu 4} \mathfrak{q}_4. \end{aligned}\right\} \tag{72}$$

The weights of two material bodies that are moving in the same gravitational field with the same velocities are mathematically exactly proportional to their inertial masses.

So we have the theorem to which Einstein still attaches such great importance, once the principle of the identity of the two masses had to be dropped. But from the procedure of the proof it is easy to recognize that this theorem has nothing to do with the actual main assumptions of Einstein's theory, which I mentioned on p. 714

[p. 169 in the original]; that rather it is based mainly on the inessential assumptions that Einstein introduced as supplements into the theory. Most of all, | to prove the theorem one absolutely must adopt the assumption (41): [173]

$$H_e = T^e_{11} + T^e_{22} + T^e_{33} + T^e_{44}$$

and the assumption (19) or (30):

$$H_g = \frac{1}{2}\sum \mathfrak{g}_{\mu\nu\alpha}\mathfrak{f}_{\mu\nu\alpha}$$

as correct.

The role played in the proof by these two assumptions is recognized most clearly if they are also introduced into the theory of gravitation that I have suggested. In formulating this theory I followed the principle of making no arbitrary auxiliary assumptions if possible, but developing the consequences purely from a single main assumption. This is the assumption that the gravitational mass is completely identical with the rest mass. Certain quite definite reasons speak against introducing the auxiliary assumptions under discussion, as we shall see. Let us, however, temporarily ignore these reasons and adopt both assumptions as correct. So we put:

$$\begin{aligned} H &= H_e + H_g, \quad T_{\alpha\beta} = T^e_{\alpha\beta} + T^g_{\alpha\beta} \\ H_e &= T^e_{11} + T^e_{22} + T^e_{33} + T^e_{44}. \\ H_g &= \frac{1}{2}\mathfrak{g}\mathfrak{f}. \end{aligned}$$

In the theory of the scalar potential we then have (loc. cit. III, p. 34, eq. (105)):

$$T^g_{\alpha\alpha} = H_g - \mathfrak{g}_\alpha\mathfrak{f}_\alpha, \quad T^g_{\alpha\beta} = -\mathfrak{g}_\alpha\mathfrak{f}_\beta.$$

A quite simple calculation yields, using the eq. (85) of my treatise III, p. 28:

$$\frac{\partial T^g_{\alpha 1}}{\partial x} + \frac{\partial T^g_{\alpha 2}}{\partial y} + \frac{\partial T^g_{\alpha 3}}{\partial z} - i\frac{\partial T^g_{\alpha 4}}{\partial t} = \frac{\partial H_g}{\partial \omega}\mathfrak{g}_a - \mathfrak{g}_a\left(\frac{\partial \mathfrak{f}_x}{\partial x} + \frac{\partial \mathfrak{f}_y}{\partial y} + \frac{\partial \mathfrak{f}_z}{\partial z} + \frac{\partial w}{\partial t}\right).$$

But we have:

$$\frac{\partial \mathfrak{f}_x}{\partial x} + \frac{\partial \mathfrak{f}_y}{\partial y} + \frac{\partial \mathfrak{f}_z}{\partial z} + \frac{\partial w}{\partial t} = -\kappa h,$$

where h means the density of gravitational mass (eq. (86) loc. cit. III, p. 28) and further (eq. (93) loc. cit. III, p. 30):

$$-\kappa h = \frac{\partial H}{\partial \omega}.$$

If we also split h into two terms $h = h_e + h_g$, where:

$$h_e = -\frac{1}{\kappa}\frac{\partial H_e}{\partial \omega}, \quad h_g = -\frac{1}{\kappa}\frac{\partial H_g}{\partial \omega},$$

then it is easily seen that

$$\frac{\partial T^g_{\alpha 1}}{\partial x} + \frac{\partial T^g_{\alpha 2}}{\partial y} + \frac{\partial T^g_{\alpha 3}}{\partial z} - i\frac{\partial T^g_{\alpha 4}}{\partial t} = \kappa h_e \vartheta_\alpha.$$

But since:

$$\frac{\partial T_{\alpha 1}}{\partial x} + \frac{\partial T_{\alpha 2}}{\partial y} + \frac{\partial T_{\alpha 3}}{\partial z} - i\frac{\partial T_{\alpha 4}}{\partial t} = 0,$$

it follows from this:

$$\frac{\partial T^e_{\alpha 1}}{\partial x} + \frac{\partial T^e_{\alpha 2}}{\partial y} + \frac{\partial T^e_{\alpha 3}}{\partial z} - i\frac{\partial T^e_{\alpha 4}}{\partial t} = -\kappa h_e \vartheta_\alpha, \tag{73}$$

an equation that is the exact analogue of eq. (38) of Einstein's theory. We can regard (73) as the equation of motion of a particle having the inertial mass

$$m_i = \int T^e_{44} dV \tag{74}$$

under the influence of the gravitational force in addition to the forces that correspond to the state variables of the aether occurring in the $T^e_{\alpha\beta}$. The gravitational mass of a particle is to be reckoned as:

$$m_g = \int h_e dV. \tag{75}$$

Now I introduce the main assumption on which my theory is based:

$$\left.\begin{aligned} h &= H \\ h_e &= H_e, \quad h_g = H_g. \end{aligned}\right\} \tag{76}$$

Following eq. (41) I put:

$$h_e = H_e = T^e_{11} + T^e_{22} + T^e_{33} + T^e_{44},$$

so that, if for simplicity we assume the body to be at rest, *Laue's theorem* implies:

$$\left.\begin{aligned} \int h_e dV &= \int T^e_{44} dV \\ m_g &= m_i. \end{aligned}\right\} \tag{77}$$

That is because Laue's theorem is separately applicable to each of the two terms in the energy tensor $T^e_{\alpha\beta}$ and $T^g_{\alpha\beta}$, since in the interior of a complete stationary system each of the two components of the energy current $T^e_{4\beta}$ and $T^g_{4\beta}$ must vanish separately.

If Einstein's auxiliary assumptions $H_e = T^e_{11} + T^e_{22} + T^e_{33} + T^e_{44}$ *and* $H_g = \frac{1}{2}\mathfrak{g}\mathfrak{k}$ *were to be introduced into the theory of gravitation suggested by me, then also in this theory the weights of two bodies that move with the same velocity in the same gravitational field would be proportional to their inertial masses.*

The theorem about the equality of the two inertial masses of closed systems is not at all a consequence of the two main assumptions of Einstein's theory of gravitation, the assumption of a tensor potential and the assumption of a peculiar transformation property of the basic equations; but it follows | from the inessential, incidental auxiliary assumptions of the theory. [174]

9. INTERNAL CONTRADICTION IN EINSTEIN'S AUXILIARY ASSUMPTIONS

Because the trace $T_{11} + T_{22} + T_{33} + T_{44}$ is a four-dimensional scalar, transformation to a coordinate system, in which the closed material system under consideration is at rest, yields:

$$\begin{aligned}&\int(T_{11} + T_{22} + T_{33} + T_{44})dV \\ &= \int(T^0_{11} + T^0_{22} + T^0_{33} + T^0_{44})\sqrt{1-\mathfrak{q}^2}dV_0 \\ &= \sqrt{1-\mathfrak{q}^2}\int T^0_{44}dV_0,\end{aligned}$$

where $\mathfrak{q}$ denotes the velocity with which the material system moves in the original coordinate system. Because T^0_{44} equals the denseness of the rest energy, that is H, it follows that:

$$\int(T_{11} + T_{22} + T_{33} + T_{44})dV = \int HdV. \tag{78}$$

But by assumption (41):

$$T^e_{11} + T^e_{22} + T^e_{33} + T^e_{44} = H_e,$$

it follows from (78) that:

$$\int(T^g_{11} + T^g_{22} + T^g_{33} + T^g_{44})dV = \int H_g dV. \tag{79}$$

Equation (79) is in direct contradiction with the auxiliary assumption (19). Namely, if one writes this auxiliary assumption in the form (30) then (39) results in:

$$T^g_{11} + T^g_{22} + T^g_{33} + T^g_{44} = 2H_g. \tag{80}$$

The two auxiliary assumptions (19) and (41) of Einstein's theory are mutually contradictory.

It is remarkable that these two assumptions are necessary precisely for the proof of the theorem of the equality of the two masses. It should probably not be difficult to eliminate the internal contradiction from Einstein's theory. *However one may well suppose that the removal of the internal contradiction will be accompanied by the failure of the theorem of the equality of the two masses*.

From the general theoretical investigations that I made concerning the nature of matter one can discern that assumption (41) is by itself untenable, even apart from the contradiction with assumption (19). I can say that this realization was indeed the reason for me to abandon *ab initio* the theorem about the equality of the two masses of a closed system; or else considerations such as those presented in Section 8 would rather quickly have come to mind.

APPENDIX

10. NORDSTRÖM'S TWO THEORIES OF GRAVITATION

Mr. Gunnar Nordström has published two different theories of gravitation, both of which he obtained by suitable modifications of Abraham's equations of gravitation (which are not in accord with the principle of relativity). The first of these was first published by him toward the end of the year 1912.[11] There the rest density of energy is decomposed into three terms:

$$H = H_e + H_p + H_g,$$

of which the second depends only on the elastic tensions of matter, the third only on the field strength of gravitation, and the first on all the remaining state variables. Mr. Nordström calls H_e in particular the rest density of the matter's inertial mass and puts:

$$\frac{\partial H}{\partial \omega} = -\kappa H_e.$$

Accordingly in this theory of Nordström's we have:

$$H = e^{-\kappa\omega} H_e' + H_p + H_g,$$

where H_e' as well as H_p and H_g no longer depend on the gravitational potential ω.

Evidently this theory is rather similar to the one suggested by me. I have:

$$H = e^{-\kappa\omega} H',$$

11 This journal 13, 1126, 1912; *Ann. d. Phys*. 40, 856, 1913 [both in this volume].

where H' no longer depends on ω, so I have avoided the somewhat artificial decomposition of H into three terms.

Nordström's notation deviates strongly from mine, one has to put:

$$H_e = \nu, \quad \kappa = \frac{g}{c^2}, \quad \omega = -\Phi, \quad \mathfrak{g}_x = \mathfrak{f}_x = -\frac{\partial \Phi}{\partial x} \quad \text{etc.,}$$

in order to obtain Nordström's equations, in addition one must note that I have set the speed of light in an ideal vacuum equal to 1, which Nordström calls c.

Nordström's second theory[12] appeared only recently. Incidentally, this is the theory about which Mr. Einstein spoke in his Vienna lecture,[13] whereas in the | discussion I meant the older theory, the only one published to that date. The second theory contains the two assumptions, that H is to be split into two terms: [175]

$$H = H_e + H_g,$$

of which only the second depends on the field strength of gravity $(\mathfrak{g}, iu)$, and that:

$$H_g = \frac{1}{2}(\mathfrak{g}^2 - u^2).$$

The gravitational potential ω shall reside only in H_e, as in Nordström's first theory. A further assumption is made about the density of inertial mass h (loc. cit. eqs. (1), (2), (14), (15)):

$$h = \frac{\nu}{1 - \kappa\omega}$$

$$\nu = \overset{e}{T}_{11} + \overset{e}{T}_{22} + \overset{e}{T}_{33} + \overset{e}{T}_{44}.$$

Finally (loc. cit. eq. (27)) the quantity:

$$\frac{\nu}{(1 - \kappa\omega)^2}$$

shall not depend on ω.

However, these assumptions incorporate a grave internal contradiction. Namely, by noting, as I proved in my treatise III on p. 30, that the energy principle can be satisfied only if

$$-\kappa h = \frac{\partial H}{\partial \omega}$$

---

12 *Ann. d. Phys.* 42, 533, 1913 [in this volume].

13 In his lecture *On the Present State of the Problem of Gravitation*, Mr. Einstein mentioned of all theories other than his own only this second theory of Nordström's. The comments on Abraham's theory found in the report in this journal 14, 1250, 1913 [in this volume] did not come up in the lecture itself. I want to mention this here in order to explain my remarks at the beginning of the discussion (this journal 14, 1262, 1913).

or, since Nordström's H_g is independent of ω:

$$-\kappa h = \frac{\partial H_e}{\partial \omega}$$

one sees that Nordström's assumptions imply:

$$\frac{1}{(1-\kappa\omega)^3}\frac{\partial H_e}{\partial \omega} = -\kappa H_e' = -\kappa \frac{T_{11}^e + T_{22}^e + T_{33}^e + T_{44}^e}{(1-\kappa\omega)^4}$$

where H_e' is to denote an ω-independent quantity. Integration results in:

$$H_e = \frac{1}{4}(1-\kappa\omega)^4 H_e' = \frac{1}{4}(T_{11}^e + T_{22}^e + T_{33}^e + T_{44}^e).$$

We therefore have:

$$\int H_e dV = \frac{1}{4}\int (T_{11}^e + T_{22}^e + T_{33}^e + T_{44}^e) dV,$$

where the integral is to be taken over the volume of a closed system.

But according to Laue's theorem we must have:

$$\int H_e dV = \int (T_{11}^e + T_{22}^e + T_{33}^e + T_{44}^e) dV.$$

Accordingly the assumptions made by Mr. Nordström must somehow lead to contradictions with the energy principle, which of course must not happen. I have not further explored whether and how this error can be eliminated from Nordström's ansatz, and therefore I do not want to discuss this theory here any further, although I believe that it would be quite interesting when consistently developed.

The notation is the same as in Nordström's older papers. In some formulas Φ is replaced by $\Phi - \Phi_0$, for example:

$$1 - \kappa\omega = 1 + \frac{g}{c^2}(\Phi - \Phi_0)$$

and further he sets:

$$\frac{\kappa}{1-\kappa\omega} = \frac{g}{1 + \frac{g}{c^2}(\Phi - \Phi_0)} = g(\Phi)$$

as well as:

$$\frac{1-\kappa\omega}{\kappa} = \frac{c^2}{g} + \Phi - \Phi_0 = \Phi'.$$

11. CONCLUSION: SUMMARY

In his Vienna lecture Mr. Einstein was very articulate about the ultimate goal of his investigations.

1. In his research an attempt is made to enlarge the theory of relativity; in particular the principle of relativity, which at first is valid only for uniform motion, is to be extended to accelerated motion, at least to uniform acceleration. As Mr. Einstein himself emphasized, this amounts to demanding covariance of the laws of nature not only with respect to linear substitutions, but also with respect to nonlinear substitutions.

2. The generalization of the principle of relativity is to be achieved by allowing the complete replacement of the accelerated motion of a material system by a gravitational field. As Mr. Einstein puts it, a physicist from his standpoint can characterize the gravitational field as "fictitious," because a suitable transformation of the basic equations of physics can always make the gravitational field disappear at the location in question, by replacing it with an equivalent state of acceleration. Conversely one can of course equally well designate an acceleration of the system as fictitious. This hypothesis of the equivalence of gravitational field and acceleration is of course realizable only if inertial and gravitational mass are identical in their nature.

Even acknowledging the extremely ingenious and painstaking workmanship that Mr. Einstein has devoted to the achievement of the stated goal, one can nevertheless say nothing more than that his attempt has had only a negative result. |

1. The degree of generalization of the relativity principle achieved in Einstein's [176]
work concerns only linear transformations, so it has nothing whatever to do with accelerated motion. In the present analysis I have demonstrated that this "generalization" means nothing more than that besides the relativity of motion there exists a relativity of the gravitational potential. This second relativity is valid in my theory as well, and there it is in fact achieved by extremely simple means.

2. The equivalence hypothesis seems to me untenable already for this reason, that there is no such thing as the identity of inertial and gravitational mass in Einstein's theory. Certainly, introducing several auxiliary assumptions produces the proposition that the gravitational and inertial mass of closed systems are strictly proportional to each other. But this proposition is by no means a consequence of the transformation properties of the fundamental equations, as it ought to be according to the equivalence hypothesis, it would be equally valid in my theory if the auxiliary assumptions just mentioned were to be imported also into it. Further, it has turned out that these auxiliary assumptions contain an internal contradiction, and thereby the proposition of the equality of the two masses becomes untenable even in the modest form it eventually assumed.

As a positive result of the present investigation I count the demonstration that in any theory in which gravitational mass is a four-dimensional tensor, an identity of the tensor of inertial mass with the tensor of gravitational mass is impossible, come what may. As far as I can see this surely establishes quite generally that a principle of the identity of the two masses cannot be valid. Whether one can attain from this a general demonstration of the impossibility of Einstein's equivalence hypothesis cannot be said without more detailed investigation, but to me it seems quite possible. In any event I am inclined to believe that the failure of Einstein's attempt is to be explained by the impossibility of success. In the discussion at Einstein's lecture I have pointed out that a generalization of the relativity principle as intended by Einstein will probably always lead to contradictions with the general principles of inquiry in physics (this journal 14, 1264). Now it would be interesting if the impossibility of generalization could be demonstrated from a different point of view by rigorous mathematics. In this context a proposition announced by Mr. Einstein (this journal 14, 1257) seems to me significant: according to it no system of fundamental equations can be devised that would be covariant in their entirety for arbitrary substitutions.

EDITORIAL NOTES

[1] *Dichtigkeit* is translated as "denseness," in order to respect the distinction Mie draws between the tensor quantity *Dichtigkeit* and the scalar *Dichte* ("density").

[2] Mie uses the term *Wesensgleichheit*, translated as "unity of essence," alluding to Einstein's use of this term to describe the relation between inertia and gravitation.

[3] In the original, Mie mistakenly refers to eq. (16) rather than eq. (17).

[4] In eq. (23), the subscript $_\lambda$ of the first occurrence of $g_{1\lambda}$ is missing in the original.

[5] In the original text the summation in the second line in the following equations runs over λ; here it has been corrected to κ.

[6] Here, Mie leaves out Einstein's qualification that the gravitational field is infinitesimally extended.

GUSTAV MIE

THE PRINCIPLE OF THE RELATIVITY OF THE GRAVITATIONAL POTENTIAL

Originally published as "Das Prinzip von der Relativität des Gravitationspotentials" in Arbeiten aus den Gebieten der Physik, Mathematik, Chemie. Festschrift Julius Elster und Hans Geitel zum sechzigsten Geburtstag, Braunschweig: Friedr. Vieweg & Sohn 1915, pp. 251–268. Author's note: Greifswald (Physical Institute of the University) April 27, 1915.

1. Gravity has long eluded theoretical investigation, and, apart from the meager knowledge gained by experience, the main reason is that the gravitational field exhibits the peculiarity that the field itself is amplified when it performs external work. It is therefore difficult to design the theory so that it does not conflict with the energy principle. Admittedly the magnetic field of two current-carrying conductors shows the same peculiarity. However, in that case one readily recognizes in the apparatus that provides the current the source of energy both for the energy increase of the magnetic field and for the energy carried off as work. For the gravitational field such an external energy source is absent, and therefore one formerly used to assume in gravitational theories that the energy of the field is negative, so that upon amplification of the field a positive energy is released and is manifested as work gained. Every theory of gravitation that is built upon the scheme of Maxwell's equations must make the assumption of a negative energy. But this assumption is untenable, because a field whose energy is negative cannot be in stable equilibrium, but is always unstable. Namely, whereas an electrostatic field, for example, exhibits that distribution of lines of force for which the energy has the smallest possible value, a field that is similarly constituted but with negative energy has of course precisely the largest possible value at equilibrium. It is therefore to be expected that when the equilibrium is slightly perturbed, the field will continually release energy to the exterior | while simultaneously [252]
moving further and further away from its equilibrium state. So in this way one does not attain a satisfactory theory of gravity.

A simple, feasible way leading out of this difficulty was first indicated by M. Abraham.[1] This way consists of including in the state variables, upon which the

1 M. Abraham, *Ann. d. Phys.* 38, 1056 (1912).

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 4
Gravitation in the Twilight of Classical Physics: The Promise of Mathematics.

amount of energy per cubic centimeter depends, the *potential* of the field, formerly taken to be only a "mathematical construct", rather than only the *field strength* of the gravitational field, formerly considered exclusively. Now, when two gravitating masses approach each other, there is on the one hand an increase in the energy of the field, counted as positive exactly like that of an electric field of similar appearance, but at the same time there is a change, a decrease to be exact, in the internal energy of the approaching material bodies, because in them the potential of the gravitational field becomes different. So the two gravitating masses release a part of their internal energy as a consequence of the change in their gravitational potentials, and thus provide the source for the work gained due to the attraction as well rather than only the increase in energy of the gravitational field. Thus, in the case of the gravitational field, matter under the influence of a changed potential performs the same task as the current source in the case of the magnetic field between current-carrying conductors mentioned above.

However, it is important to note that this procedure cannot be carried through for a field whose potential is a four vector, like that of the electromagnetic field. Therefore M. Abraham has derived the gravitational field from a potential that is invariant under Lorentz transformations, so it is a four dimensional scalar. This results in a theory without further difficulties. From the investigations of Messrs. A. Einstein and M. Grossmann,[2] it follows that one can achieve the same by deriving the gravitational field from a potential that is a four dimensional tensor. However, the theory of the tensorial gravitational potential is significantly more complicated than the scalar one, and since in spite of its complications it does not exhibit any advantages whatever over the theory of a scalar potential, I prefer to stay with the scalar potential. |

[253] Thus the potential plays a very important role in the theory of gravitation. It may be concluded from my investigations on the theory of matter[3] that the four potential of the electromagnetic field, no less than the scalar potential of the gravitational field, has to be counted among the state variables on which the energy depends. But this dependence must be such that when the electromagnetic potential is changed in the region where a material particle is located, the *net* change in energy of the particle is either nil or an infinitely small amount of higher order. For the work due to the action of electromagnetic forces is obtained with great accuracy as equal to the sum of the energy change experienced by the field due to displacement of the bodies that generate the field and the energy provided by the sources of electricity used in doing this. Therefore the energy of the material particles is to be regarded as constant, independent of the field strength and the potential prevailing in their vicinity. So it has been possible to develop a theory of the electromagnetic phenomena, adequate for a large class of empirical facts, in which the four potential does not occur except as a purely mathematical construct. It is different in the case of gravity. The total energy of the

---

2 A. Einstein und M. Grossmann, *Entwurf einer verallgemeinerten Relativitätstheorie und einer Theorie der Gravitation*. Leipzig, B. G. Teubner, 1913.

3 G. Mie, *Ann. d. Phys., Abhandl.* I, 37, 511 (1912); II, 39, 1 (1912); III, 40, 1 (1913).

material particles depends on the gravitational potential prevailing in their surroundings to such an extent that the sign of the action of the force is thereby reversed. Accordingly it seems that in order to set up a theory of gravitation one must also specify how the gravitational potential enters into the physics of the aether. For this reason—in contrast to the potential-free theory of electromagnetism—countless theories of gravitation are possible, which all show the same form of the basic equations and differ only in the way that the gravitational potential enters into the expression for the energy. Indeed several theories of that kind have already been proposed, but one could of course add to them any number of others. But this procedure is hardly satisfactory because following it one can never completely avoid pulling quite arbitrary assumptions out of thin air. And this without knowing anything empirical about how the gravitational potential enters into the expression for the energy! Thus the
consequences | derived from the theories so obtained will also be given little credibil- [254]
ity. Therefore in the following I want to investigate how far one can get without arbitrary specializations, presupposing as correct only a few quite general principles, which have a certain inherent probability due to their simplicity.

2. The first principle upon which I base the theory is *the principle of relativity.* By this I mean the proposition that all basic equations of the physics of the aether admit the Lorentz transformation. This principle has been accepted in all newer gravitational theories except that of Abraham. Why Mr. Abraham considers his special assumptions more important than the principle of relativity is something I cannot fathom.

3. Secondly I presuppose *Hamilton's principle*. The basic equations of the physics of the aether, whatever their detailed appearance, can at any rate always be divided into two groups, so that exactly as many state variables occur in each of the two groups of equations as are necessary and sufficient for a complete description of the state of the aether. So, for this description one can choose at will either the variables of the first group of equations or equally well those of the second group. In my papers I have differentiated between the two types of state variables as intensive and extensive quantities [*Intensitätsgrößen* and *Quantitätsgrößen*]. The variables of the two groups can be coordinated with each other into pairs of conjugate [*entsprechende*] variables. Hamilton's principle amounts in essence to the proposition that the state variables of one group can be calculated from those of the other group if one knows only a single function of them, the Hamiltonian. To calculate a desired state variable from given variables of the other group one has to take the partial derivative of the Hamiltonian with respect to the conjugate state variable; this partial derivative is the desired quantity. In most cases it turns out to be advantageous to consider the intensive quantities as the primary state variables. I have always denoted the Hamiltonian function of the intensive quantities by Φ and have called it "the world function."

If Hamilton's principle is valid, the equations of motion of mechanics and the energy principle can easily be derived from it as consequences. If one prefers not to
presuppose it, then it | is at least highly questionable whether the energy principle can [255]
be maintained. In the gravitational theory of G. Nordström[4] the special assumptions

are chosen in such a way that Hamilton's principle is not valid, so that no world function exists. It is incomprehensible why Mr. Nordström prefers his special assumptions over Hamilton's extraordinarily clear and simple principle.[5]

Concerning the variables that are supposed to determine the world aether, and on which the world function Φ will accordingly depend, we shall assume the guiding principle that we shall try to get by with as few variables as possible. We can safely leave the question of whether it will prove necessary as science progresses to increase the number of state variables undecided; probably the propositions that we can derive from our general principles will not be significantly modified by this. At the present state of science the following mutually independent variables of state (intensive quantities) suffice: 1. electric field strength $\mathfrak{e}$; 2. magnetic induction $\mathfrak{b}$; 3. electromagnetic four potential $\varphi, \mathfrak{f}$; 4. four vector of the field strength of gravity $\mathfrak{g}, \gamma$; 5. gravitational potential ω. The world function is therefore: $\Phi(\mathfrak{e}, \mathfrak{b}, \varphi, \mathfrak{f}, \mathfrak{g}, \gamma, \omega)$.

The corresponding extensive quantities are: 1. electric displacement $\mathfrak{d}$; 2. magnetic field strength $\mathfrak{h}$; 3. electric charge and electric current $\rho, \mathfrak{v}$; 4. four vector of excitation of the gravitational field $\mathfrak{k}, \chi$; 5. density of gravitational mass h.

Hamilton's principle leads to the following equations:

$$\mathfrak{d} = -\frac{\partial\Phi}{\partial\mathfrak{e}}, \mathfrak{h} = \frac{\partial\Phi}{\partial\mathfrak{b}}, \rho = \frac{\partial\Phi}{\partial\varphi}, \mathfrak{v} = -\frac{\partial\Phi}{\partial\mathfrak{f}}, \mathfrak{k} = \frac{\partial\Phi}{\partial\mathfrak{g}}, \chi = -\frac{\partial\Phi}{\partial\gamma}, h = -\frac{\partial\Phi}{\partial\omega}. \quad (1)$$

[256] | Here I generally use the same notation as in my earlier papers (cf. Theory of Matter III, p. 30). Except that previously I chose the less practical notation u, w in place of the letters γ, χ; further instead of h I previously used γH or (Physik. Zeitschr. 15, 175 (1914)) κh, where γ resp. κ means the gravitational constant $1,016 \cdot 10^{-24}$. Let us denote the gravitational mass of the particle by m_g, that is:

$$m_g = \int h dV, \quad (2)$$

where the integral is to be taken over the entire volume occupied by the particle (cf. Theory of Matter III p. 6); then the force $\mathfrak{P}$ acting on it in a field $\mathfrak{g}$ is:

$$\mathfrak{P} = m_g \mathfrak{g} \ . \quad (3)$$

4. My third assumption is *the principle of the relativity of the gravitational potential*.

---

4 G. Nordström, *Ann. d. Phys*. 42, 533 (1913).

5 In answer to the objection I raised (*Physik. Zeitschr.* 15, 175 (1914)), that then the energy principle must also fail, Mr. G. Nordström has recently tried to show (*Physik. Zeitschr.* 15, 375 (1914)) that in spite of the disagreement with Hamilton's principle his theory does not have to conflict with the energy principle. However, the proof of this has not yet really been established, because Mr. Nordström has not actually put down the basic equations of the theory, but he has only indicated how one might set them up. It seems to me by no means certain that one can proceed according to his indications without using Hamilton's principle, for Mr. Herglotz, who sets up basic equations in his Mechanics of Continuous Media (*Ann. d. Phys*. 36, 493 (1911)) of the type indicated by Mr. Nordström, has certainly considered it necessary to base his investigations on Hamilton's principle.

Introducing the potentials as independent state quantities leads to a peculiar difficulty. It forces us to assume that the properties of matter and the laws of material processes depend on the potentials that prevail at the location where these properties and processes are observed. On the other hand no one has ever been aware of such an influence of the potentials, and if it exists at all it must at least be quite insignificant. Otherwise, although it may have seldom been looked for,[6] one should think that it would already have been noticed on other occasions. So we are confronted with the dilemma that on the one hand the theory absolutely demands an influence of the potentials on physical processes, and that on the other hand experience negates this influence to such an extent that it has become second nature to regard the potentials as purely mathematical, calculational constructs.

This dilemma can be eliminated, initially for the gravitational potential, using the principle at hand in a way that is as simple as it is perfect. The principle declares: |

In two regions of different gravitational potential exactly the same processes can run according to exactly the same laws if one only thinks of the units of measurement as changing in a suitable way with the value of the gravitational potential. [257]

I shall show that this principle can be realized even as I obtain its mathematical formulation. We assume that the gravitational potential ω has a value Ω that *differs from zero* in an ideal vacuum, at an infinite distance from any matter. Ω is a universal constant of the aether, like the speed of light, the constant of gravity etc.; a Lorentz transformation does not change its value because ω is a four dimensional scalar. I note incidentally that by contrast the four potential of the electromagnetic field $(\varphi, \mathfrak{f})$ must be zero in a vacuum. For, if it had a non-zero value there, this would change upon any Lorentz transformation. One would then have universal constants that would depend on the choice of the spacetime coordinate system, in other words the principle of relativity would not be strictly valid. In the same way the field strengths $(\mathfrak{b}, -i\mathfrak{e})$ and $(\mathfrak{g}, i\gamma)$ must of course also vanish in a pure vacuum. Let us now transform all quantities of state in such a way that each is multiplied by a constant factor:

$$\left.\begin{array}{ll} \omega = a \cdot \omega', & (\mathfrak{g}, i\gamma) = b \cdot (\mathfrak{g}', i\gamma'), \\ (\mathfrak{f}, i\varphi) = c \cdot (\mathfrak{f}', i\varphi'), & (\mathfrak{b}, -i \cdot \mathfrak{e}) = d \cdot (\mathfrak{b}', -i\mathfrak{e}') \end{array}\right\} \tag{4}$$

The primed and unprimed quantities then differ only by the measurement units. If the same equations are to hold in the primed quantities as in the unprimed ones, then it is absolutely necessary that the world function Φ also experiences no changes through the introduction of new measurement units other than a constant factoring out:

$$\Phi(\mathfrak{e}, \mathfrak{b}, \varphi, \mathfrak{f}, \mathfrak{g}, \gamma, \omega) = e \cdot \Phi(\mathfrak{e}', \mathfrak{b}', \varphi', \mathfrak{f}', \mathfrak{g}', \gamma', \omega') = e \cdot \Phi'. \tag{5}$$

6 Some time ago with Prof J. Herweg I have used a good echelon grating to observe the spectrum of a mercury arc lamp, in which quite large values of the magnetic vector potential could be produced by means of nearby electric ring magnets, without generating a magnetic field in the lamp. No trace of an influence of the vector potential on the spectral lines was revealed.

For then, but only then, is it true for the state variables of the other set that they are transformed in the same way:

$$\left.\begin{array}{ll} h = \dfrac{e}{a}\cdot h', & (\mathfrak{k}, i\chi) = \dfrac{e}{b}\cdot(\mathfrak{k}', i\chi'), \\ (\mathfrak{v}, i\rho) = \dfrac{e}{c}\cdot(\mathfrak{v}', i\rho'), & (\mathfrak{h}, -i\cdot\mathfrak{d}) = \dfrac{e}{d}(\mathfrak{h}', -i\cdot\mathfrak{d}') \end{array}\right\} \tag{6}$$

[258] | If one now also puts:

$$x = \frac{a}{b}\cdot x',\quad y = \frac{a}{b}\cdot y',\quad z = \frac{a}{b}\cdot z',\quad t = \frac{a}{b}\cdot t', \tag{7}$$

and if the condition:

$$\frac{a}{b} = \frac{c}{d} \tag{8}$$

is satisfied, then, as one can prove easily, the basic equations of the physics of the aether are equally as valid in the primed quantities as in the unprimed ones. If we choose the value $\omega'_\infty = \Omega$, then we have thereby completely reduced all physical problems in a region of gravitational potential $\omega_\infty = \Omega_1 = a\Omega$ to the corresponding problems in an ideal vacuum where the gravitational potential is Ω; that is, the processes in the two regions differ only in the difference of the units of measurement.

This shows that the principle of relativity formulated above is valid if and only if the world function Φ has the property demanded by equation (5). But equation (5) is the condition that Φ is a homogeneous function of the variables.

We can state quite generally that the principle of the relativity of the gravitational potential is identical with the demand that the world function Φ *is a homogeneous function of the variables:*

$$\omega,\quad (\mathfrak{g}^\kappa, \gamma^\kappa),\quad (\mathfrak{f}^\lambda, \varphi^\lambda),\quad (\mathfrak{b}^\mu, \mathfrak{e}^\mu),$$

where κ, λ, μ *denote arbitrary positive or negative, integral or fractional numbers.*

I will call the degree of this homogeneous function ν. In the equations (4) and (5) we then have to put:

$$b = a^{\frac{1}{\kappa}},\quad c = a^{\frac{1}{\lambda}},\quad d = a^{\frac{1}{\mu}},\quad c = a^{\nu}. \tag{9}$$

If we further put (taking note of 8):

$$1 - \frac{1}{\kappa} = \frac{1}{\lambda} - \frac{1}{\mu} = \alpha, \tag{10}$$

then we have:

$$x = a^{\alpha}\cdot x',\quad y = a^{\alpha}\cdot y',\quad z = a^{\alpha}\cdot z',\quad t = a^{\alpha}\cdot t'. \tag{11}$$

As an example of a theory in which this relativity principle is valid I mention | the [259]
theory of gravitation developed by me on a previous occasion (Theory of Matter III, p. 25ff). To obtain this, one has to specialize by setting:

$$\kappa = 1, \quad \lambda = \mu = \frac{1}{\nu},$$

and further:

$$\Omega = -\frac{\nu}{\gamma},$$

where $\gamma = 1,016 \cdot 10^{-24}$ signifies the gravitational constant. Because ω becomes zero at infinity in the notation chosen previously by me, one has to substitute $(\omega - \Omega)$ everywhere in place of ω in the formulae of my treatise III, in order to adjust them to my current notation. On then obtains the equations of treatise III, if one makes ν infinite. Namely:

$$\lim_{\nu = \infty}\left(\frac{\omega^\nu}{\Omega^\nu}\right) = e^{-\gamma \cdot (\omega - \Omega)}.$$

If we want to transform to a space having $\Omega_1 = \Omega + \omega_0$, then we have to put:

$$a = \frac{\Omega_1}{\Omega},$$

and we have:

$$\lim_{\nu = \infty} a = 1, \quad \lim_{\nu = \infty}(a^\nu) = e^{-\gamma \cdot \omega_0},$$

from which the transformation equations presented in treatise III on p. 63 follow directly.

Finally, equation (10) implies: $\alpha = 0$, $a^\alpha = 1$, so the units of length and time measurements are not changed by the transformation.

Accordingly the theory I developed previously is indeed a special case, or rather a limiting case, of theories in which the principle of relativity of the gravitational potential holds.

5. The only empirical fact that we know about gravitation to date is the *proportionality of the gravitational and inertial mass of a body.* It is interesting that this fact can also be obtained theoretically, as I will now show, by assuming the principle of relativity of the gravitational potential to be correct. |

I focus on a material body that is a complete system in the sense of Laue's theo- [260]
rem.[7] For the sake of generality I assume that the elementary particles of the body execute arbitrary random motion, whereas the body as a whole is at rest. I will mark

7 M. v. Laue, *Das Relativitätsprinzip*, 2nd Ed. p. 208. Braunschweig 1913. In the following I use the formulas developed in *Theory of Matter* III, section 27 and 28 (p. 5–11) and 43 (p. 42, 43) [in this volume].

the time average of state variables by horizontal bars above the respective mathematical symbols, as in my earlier investigations. The principle of relativity of the gravitational potential, according to which Φ is a homogeneous function of the state variables (cf. p. 257) yields

$$\nu \cdot \overline{\Phi} = \overline{\omega \cdot \frac{\partial \Phi}{\partial \omega}} + \frac{1}{\kappa} \cdot \left(\overline{\mathfrak{g} \cdot \frac{\partial \Phi}{\partial \mathfrak{g}}} + \overline{\gamma \cdot \frac{\partial \Phi}{\partial \lambda}}\right) + \frac{1}{\lambda} \cdot \left(\overline{\mathfrak{f} \cdot \frac{\partial \Phi}{\partial \mathfrak{f}}} + \overline{\varphi \cdot \frac{\partial \Phi}{\partial \varphi}}\right)$$
$$+ \frac{1}{\mu} \cdot \left(\overline{\mathfrak{e} \cdot \frac{\partial \Phi}{\partial \mathfrak{e}}} + \overline{\mathfrak{b} \cdot \frac{\partial \Phi}{\partial \mathfrak{b}}}\right),$$

or, using the relations (1):

$$\nu \overline{\Phi} = -\overline{\omega \cdot h} + \frac{1}{\kappa} \cdot (\overline{\mathfrak{g} \cdot \mathfrak{k}} - \overline{\gamma \cdot \chi}) + \frac{1}{\lambda}(\overline{\varphi \cdot \rho} - \overline{\mathfrak{f} \cdot \mathfrak{v}}) - \frac{1}{\mu} \cdot (\overline{\mathfrak{e} \cdot \mathfrak{d}} - \overline{\mathfrak{b} \cdot \mathfrak{h}}). \tag{12}$$

Let the inertial mass of the body be m, the gravitational mass m_g, then we have by equation (2) and according to the Theory of Matter III, equation (116) on p. 42:

$$m_g = \int \overline{h} \cdot dV \tag{13}$$

$$m = \int (\overline{\Phi} + \overline{\gamma \cdot \chi}) \cdot dV. \tag{14}$$

In addition we need the following equations from the *Theory of Matter* III, p. 7: equations (64) and (65), as well as p. 43: equation (117) and (118):

$$\int \overline{\mathfrak{e} \cdot \mathfrak{d}} \cdot dV = \int \overline{\varphi \cdot \rho} \cdot dV \tag{15}$$

$$\int \overline{\mathfrak{b} \cdot \mathfrak{h}} \cdot dV = \int \overline{\mathfrak{f} \cdot \mathfrak{v}} \cdot dV \tag{16}$$

$$\int (\overline{\mathfrak{g} \cdot \mathfrak{k}} - \overline{\gamma \cdot \chi}) \cdot dV = \int \overline{\omega \cdot h} \cdot dV - \omega_\infty \cdot m_g \tag{17}$$

$$3m = \int (\overline{\mathfrak{g} \cdot \mathfrak{k}} + 3 \cdot \overline{\gamma \cdot \chi} - \overline{\mathfrak{e} \cdot \mathfrak{d}} + \overline{\mathfrak{b} \cdot \mathfrak{h}}) \cdot dV. \tag{18}$$

I have everywhere substituted γ and χ for the letters u and w used previously, also h in place of γH, finally I substituted $\omega - \omega_\infty$ for ω. |
[261] If one now notes equation (10):

$$1 - \frac{1}{\kappa} = \frac{1}{\lambda} - \frac{1}{\mu} = \alpha,$$

by a simple calculation one finds from equations (12) to (18):

$$(\nu + 3\alpha) \cdot m - (\nu + 4\alpha) \cdot \int \overline{\gamma \cdot \chi} \cdot dV = -\omega_\infty \cdot m_g. \tag{19}$$

In an ideal vacuum, where $\omega_\infty = \Omega$, we have:

$$(\nu + 3\alpha) \cdot m - (\nu + 4\alpha) \cdot \int \overline{\gamma \cdot \chi} \cdot dV = -\Omega \cdot m_g. \tag{20}$$

To begin with, focus on the case that the elementary particles of the body are all at rest:

$$\gamma = -\frac{\partial\omega}{\partial t} = 0.$$

Therefore, we have in a region of gravitational potential ω_∞:

$$(\nu + 3\alpha) \cdot m = -\omega_\infty \cdot m_g, \tag{21}$$

and in an ideal vacuum:

$$(\nu + 3\alpha) \cdot m = -\Omega \cdot m_g. \tag{22}$$

Thus the ratio m_g/m is a universal constant.

For a material body whose elementary particles are motionless, the law of the proportionality of inertial and gravitational mass holds with mathematical precision.

If one wished to regard the law of the proportionality of the two masses as a kind of axiom, which must also be satisfied with mathematical precision when the elementary particles of the body execute hidden motions, then one would have to assume, in addition to the principle of the relativity of the gravitational potential, the validity of the relation:

$$\nu + 4\alpha = 0.$$

But this law certainly states only an empirical fact, and even if it is true to very high accuracy according to the experiments of Eötvös, there is no sensible reason why one should accord it a character other than an empirical, approximate one. Even Newton's laws of motion, though treated almost as axioms for hundreds of years, are only approximate propositions according to the theory of relativity. These laws, it is true, are valid to such accuracy that usually one cannot experimentally substantiate any deviations from them. They approach the truth so closely only because the experimentally attainable speeds of material bodies can be characterized as infinitesimal compared | to the speed of light. It would be quite possible that the whole validity of [262] the law of proportionality of the two masses has a quite similar reason; namely, that the speeds of the hidden motions of the elementary particles of a material body are in general infinitely small compared to the speed of light. In an earlier paper I have tried to estimate what the order of magnitude of the deviation from proportionality of the two masses, due to the thermal motion of the molecules, might be; and I found that even at temperatures of several thousands of degrees Celsius the deviation lies below an experimentally detectable magnitude (*Theory of Matter* III, p. 50). But there are no cogent reasons to assume, for example, that more intense motion takes place in the interior of atoms themselves. Occasionally it is strongly emphasized[8] that, according to research by L. Southerns to a fractional accuracy of $5 \cdot 10^{-6}$, the quotient of the two masses has the same value for radioactive uranium oxide as for lead oxide. To be sure, this fact would be of great importance if it were known that in the interior of radioactive atoms intense motions already prevail, such as those exhibited by the emitted α- and β-particles upon explosion. If this were the case, one could not be satisfied with the proposition of proportionality of the two masses as just derived, one would have to demand that it should also be valid for material bodies with intense hidden motion, for example, in such a way that $\nu + 4\alpha = 0$. But to me the hypothesis of violent inner motions in radioactive atoms seems unlikely, especially because it would be hard to understand why it did not produce any radiation of electromagnetic waves. At any rate what is simpler is the notion that also in the interior of radioactive atoms, generally only motions which are to be called very slow compared to the speed of light prevail, but that in this process an atom occasionally reaches an unstable equilibrium state and explodes, and that now its fragments gain the enormous speeds with which they fly apart. But then the result of L. Southerns is explained without further ado.

If one assumes that the hidden motions of the molecules, the atoms, and the ele-
[263] *mentary particles in the interior of the atoms that constitute a material body | are very slow compared to the speed of light, then the principle of the relativity of the gravitational potential yields the law of proportionality of the two masses as an approximate theorem of great accuracy.*

Should the presence of very rapid motions in the interior of atoms really be proved at some time, then there would still be time to examine the theory, as to whether and how it correctly reproduces the action of gravity on these atoms.

Equation (22) allows us to make certain statements about the value of the universal constant Ω. Namely, the ratio

$$\frac{m_g}{m} = \kappa$$

can be specified once one has fixed some system of units (cf. *Theory of Matter* III, p. 42, where instead of κ I used the letter γ). Choosing the erg as unit of energy and mass, the centimeter as unit of length, $1/3 \cdot 10^{10}$ seconds as unit of time, so that the speed of light equals 1, and choosing further the units of the gravitational field such that $\mathfrak{k} = \mathfrak{g}$ in an ideal vacuum, results in:

$$\kappa = 1.016 \cdot 10^{-24},$$

and therefore

$$\Omega = -\frac{(\nu + 3\alpha)}{\kappa} = -(\nu + 3\alpha) \cdot 0.985 \cdot 10^{24}. \tag{23}$$

8 M. Abraham, *Jahrb. d. Radioakt. u. Elektronik* 11, 470 (1915).

Unless $(\nu + 3\alpha)$ happens to be very small, Ω is very large, and the only relevant fact is that it is very large also compared to all changes that the gravitational potential may experience due to the vicinity of large gravitational masses.

For example, let ω_E be the gravitational potential at the surface of the Earth, and let M_g be the gravitational mass and R the radius of the Earth, then:

$$\omega_E - \Omega = \frac{M_g}{4\pi \cdot R} .$$

Substitution for these values in our chosen units

$$M_g = 1.016 \cdot 10^{-24} \cdot 9 \cdot 10^{20} \cdot 5.95 \cdot 10^{27} = 5.44 \cdot 10^{24},$$
$$R = 6.37 \cdot 10^{8} \text{ cm},$$

yields:

$$\omega_E - \Omega = 6.84 \cdot 10^{14}.$$

| So if $(\nu + 3\alpha)$ is of the order of magnitude 1, then $\omega_E - \Omega$ is of the order $10^{-9} \cdot \Omega$. Likewise the potential ω_s on the surface of the Sun yields [264]

$$\omega_s - \Omega = 2.24 \cdot 10^{18},$$

that is, about $10^{-6}\Omega$ if $(\nu + 3\alpha)$ is of the order of magnitude 1.

Let the potential on the surface of any celestial object be Ω_1, then the quantity a with which the transformation from (4) to (9) is to be executed, is:

$$a = 1 + \frac{\Omega_1 - \Omega}{\Omega} = 1 - \frac{\kappa}{(\nu + 3\alpha)} \cdot (\Omega_1 - \Omega), \tag{24}$$

so a deviates from 1 only by a very small amount. The change of the distances and times under the influence of the changed potential occurs in the ratio:

$$a^{\alpha} = 1 - \frac{\alpha}{\nu + 3\alpha} \cdot \kappa \cdot (\Omega_1 - \Omega), \tag{25}$$

and the change of the density of energy or of the inertial mass in the ratio:

$$a^{\nu} = 1 - \frac{\nu}{\nu + 3\alpha} \cdot \kappa \cdot (\Omega_1 - \Omega). \tag{26}$$

The total inertial mass of a material body changes by the ratio:

$$a^{\nu + 3\alpha} = 1 - \kappa \cdot (\Omega_1 - \Omega), \tag{27}$$

and its total gravitational mass by the ratio:

$$a^{\nu + 3\alpha - 1} = 1 - \left(1 - \frac{1}{\nu + 3\alpha}\right) \cdot \kappa \cdot (\Omega_1 - \Omega). \tag{28}$$

Thus all units change by very small amounts "of the first order" under the influence of the gravitational potential.

6. The theorem of the proportionality of the two masses can be derived in a more intuitive way from the principle of the relativity of the gravitational potential as follows.

Let a material body be located in a region where a gravitational potential ω_∞ exists. By changing masses of material at large distances from the body let the potential be brought to $\omega_\infty + \Delta\omega$, and let this change of potential occur uniformly during a time Δt, so that during Δt the constant gravitational state $\gamma = -\partial\omega/\partial t = -\Delta\omega/\Delta t$ exists at the place considered. I will denote by S a surface that surrounds the body,
[265] but at a sufficient distance from its molecules that | on it the superposition principle is valid for the fields, and so that at points on S the value of the gravitational field caused by the body is constant in time, uninfluenced by the hidden motions of its elementary particles. The surface integral of the field excitation $\mathfrak{k}$ over the surface S then yields the gravitational mass of the body m_g:

$$-\int \mathfrak{k} \cdot d\mathfrak{S} = m_g.$$

Further, during the time Δt and through every element $d\mathfrak{S}$ of the surface there flows a constant energy current of density $\gamma \cdot \mathfrak{k}$ (Theory of Matter III, p. 29). Before and after this time, $\gamma = 0$ on the surface, and therefore also no energy enters or leaves. Thus, as ω_∞ is changed to $\omega_\infty + \Delta\omega$, the body gains the net amount of energy:

$$\begin{aligned} \Delta m &= -\Delta t \cdot \int \gamma \cdot \mathfrak{k} \cdot d\mathfrak{S} = -\gamma \cdot \Delta t \int \mathfrak{k} \cdot d\mathfrak{S}, \\ \Delta m &= -\Delta\omega \cdot m_g. \end{aligned} \tag{29}$$

This is the energy change that was discussed in detail in the introduction as the cause for the attractive effect of gravity in spite of the positive field energy. If we put $\omega_\infty = a \cdot \Omega$ and $\omega_\infty + \Delta\omega = (a + \Delta a) \cdot \Omega$, we can also write equation (29) as follows:

$$\Delta m = -\Omega \cdot m_g \cdot \Delta a,$$

or

$$\Delta m = -\omega_\infty \cdot m_g \cdot \frac{\Delta a}{a} \quad . \tag{30}$$

If the principle of the relativity of the gravitational potential is valid, then Δm can be calculated in yet another way. If we denote the density of energy in an arbitrary element of volume dV of the body by W, then:

$$m = \int W \cdot dV, \tag{31}$$

where W is a homogeneous function of degree ν of the variables given on p. 257. If the body were now to experience no change due to the change of the gravitational potential other than the change in measurement units, then during the time Δt one would have to regard the expression (31) as a function of a single variable a, and the change in m would be:

$$\Delta_a m = \left(\int \frac{\partial W}{\partial a} \cdot dV \cdot \int W \cdot \frac{\partial dV}{\partial a}\right) \cdot \Delta a$$

$$\Delta_a m = (\nu + 3\alpha) \cdot \int W \cdot dV \cdot \frac{\Delta a}{a}$$

$$\Delta_a m = (\nu + 3\alpha) \cdot m \cdot \frac{\Delta a}{a} \quad . \tag{32}$$

| If the elementary particles of the body remain at rest in their equilibrium posi- [266]
tions, then the supposition just made is certainly satisfied. For the occurrence of the quantity γ during the time Δt does not cause any motion, so the elementary particles are still at rest in the equilibrium positions after Δt has passed, and the body has experienced no other changes than those subject to the changes of a. In this case we can put $\Delta_m = \Delta_a m$, and the combination of equation (31) with (32) yields:

$$-\omega_\infty \cdot m_g = (\nu + 3\alpha) \cdot m.$$

Thus for a body with motionless elementary particles we have found equation (21) in a second way.

However, if the elementary particles execute hidden motions, then it is to be expected that the occurrence of the quantity γ during the time Δt influences these motions, particularly because the superposition principle is not valid in the interior of matter, and because therefore all state variables are influenced by the value of γ. After the time Δt has passed, the average value of the hidden motion of the elementary particles will therefore have become different than before Δt due to the action of γ. We can express this intuitively by saying that a change in the gravitational potential is to be associated with a small adiabatic temperature change of the body. That is, not only is there a change in the temperature as such, which is a measure of the random motion of the molecules, but there is also a change in that quantity which we may characterize as the temperature of motion of the elementary particles inside the atom. The temperature inside the atom does not have to be associated with the body temperature proper. I will denote by ΔQ the small amount of energy that one would have to transfer to the body in order to reduce the temperature of the hidden motion of its elementary particles to the values they had before Δt, so that the net change in the body's energy would be $\Delta_a m$. This is the latent or bound energy gained by the body during an "isothermal" change of ω. Its energy change upon an "adiabatic" change is therefore:

$$\Delta m = \Delta_a m - \Delta Q. \tag{33}$$

Now I denote by $(\partial Q / \partial \omega)_{is}$ the increase of the latent energy of the body during an "isothermal" change of the gravitational potential in comparison with the increase in the potential, then we have:

$$\Delta Q = \omega_\infty \cdot \left(\frac{\partial Q}{\partial \omega}\right)_{is} \cdot \frac{\Delta a}{a}. \tag{34}$$

[267] | If we substitute the expressions of equation (30), (32), (34) into equation (33) then the result is:

$$(\nu + 3\alpha) \cdot m - \omega_\infty \cdot \left(\frac{\partial Q}{\partial \omega}\right)_{is} = -\omega_\infty \cdot m_g. \tag{35}$$

The statement on proportionality of the two masses is approximately valid if the "latent" energy ΔQ is vanishingly small in comparison with the "free" energy Δm.

Comparing equation (19), found from Laue's theorem, with (35) we find:

$$(\nu + 4\alpha) \cdot \int \overline{\gamma \cdot \chi} \cdot dV = \omega_\infty \cdot \left(\frac{\partial Q}{\partial \omega}\right)_{is}. \tag{36}$$

RESULTS AND PROSPECTS

1. The principle of the relativity of the gravitational potential is established as the simplest expression of the fact that the gravitational potential in general has no perceptible influence on material processes, although it occurs as an independent quantity of state in the basic equations of the physics of the aether. We succeeded in formulating the principle in a quite general fashion, without making special assumptions about the form in which the gravitational potential enters into the basic equations of aether physics.

2. From the principle of the relativity of the gravitational potential one can derive theoretically the well-known empirical law of the proportionality of the gravitational and inertial mass of all material bodies. It is true that this law may have only approximate validity for bodies whose elementary particles execute random hidden motions. But if the speeds of the hidden motions are very small compared to the speed of light, the accuracy to which the law is valid can be so great that one cannot find deviations experimentally. If one wanted the law to be valid in general and with mathematical precision, one would have to supplement the principle of the relativity of the gravitational potential with an extra assumption.

3. The additional term that implies deviation from the mathematically exact validity of the proportionality of the two masses can be given an interesting interpretation. If the gravitational potential experiences a change at the place where a material body

is located, maybe due to displacement of a distant large and heavy mass, then simul-
taneously with this there is a change not only of the body's energy | content, but also [268]
in general of its temperature of the molecular random motion as well as of the interatomic motion. If the body should change strictly isothermally with the gravitational potential, so that the temperatures of the hidden motions of its elementary particles all remain constant, then the change of its free energy due to the potential change must be supplemented by a a change of its latent energy, for example by radiation. This supplied or removed latent energy provides the measure of our additional term. If it is very small compared to the change of the free energy, the deviation from proportionality of the two masses is small; if this latent energy vanishes, the theorem of the proportionality is mathematically exact.

4. As the next goal of the theory of matter, the task of setting up a principle for the electromagnetic four potential that is analogous to the principle of the relativity of the gravitational potential, and thereby providing an explanation for the lack of perceptible influence of the four potential on material processes presents itself.

MAX BORN

THE MOMENTUM-ENERGY LAW IN THE ELECTRODYNAMICS OF GUSTAV MIE

Originally published as "Der Impuls-Energie-Satz in der Elektrodynamik von Gustav Mie" in Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen 1914, 1, pp. 23–36. Submitted by Mr. Hilbert during the meeting on 20th December 1913.

INTRODUCTION: THE MATHEMATICAL FORM OF MIE'S ELECTRODYNAMIC CONCEPTION OF THE WORLD

Whereas the electron theory developed by H. A. Lorentz requires certain hypotheses about the structure of the electron (e.g. the hypothesis regarding the rigidity in the usual sense, or in the context of the theory of relativity), Gustav Mie[1] set himself the task of trying to modify Maxwell's equations in such a way that the existence of electrons ("nodes" of the field) and, even more generally, the existence of material atoms and molecules follows necessarily from the new equations. The fact that without the addition of new forces, stable accumulations of charge, as represented by electrons, are incompatible with the usual differential equations of the magnetic field is closely linked to the linearity of these equations. Therefore, it was first of all necessary to relinquish linearity. Mie carried out this idea in the most general and elegant manner which can be imagined in the framework of today's physics borne from Lagrange's analytical mechanics. To illustrate the type of generalization of the fundamental equations, it is perhaps | best to start with the equation of motion of a system of [24]
masses with one degree of freedom q. If

$$\Phi(\dot{q}, q) = T - U$$

represents the Lagrangian (difference between kinetic and potential energy), then it is well known that from the variation of the Hamiltonian integral

1 G. Mie, *Grundlagen einer Theorie der Materie*. 3rd. communication in *Ann. d. Phys.* (4), vol. 37, p. 511; vol. 39, p. 1; vol. 40, p. 1.

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 4
Gravitation in the Twilight of Classical Physics: The Promise of Mathematics.

$$\int_{t_1}^{t_2} \Phi(\dot{q}, q) dt \tag{1}$$

one obtains the equations of motion in the form

$$\frac{d}{dt}\frac{\partial\Phi}{\partial\dot{q}} - \frac{\partial\Phi}{\partial q} = 0. \tag{2}$$

The transition from the usual equations of the electromagnetic field to Mie's fundamental equations can then be considered to be parallel to the transition from a quasi-elastic system where Φ has the form $\Phi = \frac{a}{2}\dot{q}^2 + \frac{b}{2}q^2$ to a system where Φ is a completely arbitrary function of $\dot{q}$ and q. In this process, the form of the differential equation (2) remains completely preserved. Indeed, in the final analysis, Mie's theory aims to show that the field equations of electron theory are variational derivatives of a variational principle completely analogous to (1), except that there are 4 functions of 4 variables, where Φ is a certain quadratic form of the field quantities, and that then, as in the mechanical example shown above, the form of the fundamental equations remains completely preserved if Φ becomes an arbitrary function of the field quantities. Therefore, one can say that the equations of Mie achieve the same for electrodynamics as the Lagrange equations of the second kind achieve for the mechanics of systems of point particles [*Punktsysteme*]. They offer a formal scheme which, through an appropriate choice of the function Φ, can be adjusted to the specific properties of the system. As the aim of the mechanistic explanation of nature in the past was to derive a Lagrangian function Φ for the interaction of atoms and to derive all physical and chemical properties of matter, so Mie now sets for himself the task of selecting his "world-function" Φ in such a way that on the basis of its differential equations the existence of the electron and the atoms, as well as the totality of their interactions follows. I would like to view this requirement of Mie as the mathematical
[25] content of that | program which considers the aim of physics to be the construction of an "electromagnetic worldview."

In the following, I would like to make a contribution to the clarification of the mathematical structure of Mie's fundamental equations. The variational problem of Mie is still not the most general one can devise for the four-dimensional continuum, and one is well advised to compare it with the most general, in order to determine what are the properties which have to be attributed to the four-dimensional continuum (the aether), in order to obtain specifically Mie's laws. It will turn out that these are *not* the properties of an *elastic* body. The four-dimensional theory of elasticity compatible with the principle of relativity has been exhaustively treated by Herglotz[2] and is obtained through a different specialization of our variational principle. Mie's four-dimensional continuum corresponds rather to the three-dimensional aether of MacCullagh,[3] who, from the assumption that the vortices of the aether and not its

2 *Ann. d. Phys.* (4), vol. 36, p. 493.

deformations store energy, obtains equations identical with Maxwell's equations for stationary electrodynamic processes. The analogy of Mie's theory with Lagrangian mechanics is manifested most clearly by considering the law of conservation of energy. It is known that for a variational problem of the form (1), there always exists an integral of the differential equations (2), expressing conservation of energy, if the independent variable t does not appear explicitly in Φ (t is then a "cyclic coordinate"). Because then one has

$$\frac{d\Phi}{dt} = \frac{\partial\Phi}{\partial\dot{q}}\ddot{q} + \frac{\partial\Phi}{\partial q}\dot{q},$$

and if one adds to this equation (2) multiplied by $\dot{q}$, one obtains

$$\frac{d\Phi}{dt} = \frac{d}{dt}\left(\dot{q}\frac{\partial\Phi}{\partial\dot{q}}\right); \tag{3}$$

If one introduces as "energy" the Legendre transform of Φ:

$$W = \Phi - \dot{q}\frac{\partial\Phi}{\partial\dot{q}},$$

then one can write equation (3) in the form

$$\frac{dW}{dt} = 0 \quad \text{or} \quad W = \text{const.} \tag{3'}$$ |

which represents the law of conservation of energy. [26]

In Mie's electrodynamics there also exists a momentum-energy conservation law which plays a significant role in all the new dynamical theories based on the principle of relativity. The law consists of 4 equations, the first three express the conservation of momentum, the last the conservation of energy. Mie obtains the last of these equations by calculation and the others on the basis of symmetry requirements demanded by the principle of relativity. I will show in the following that these 4 equations are precise generalizations of equation (3) for the case of 4 variables. The requirement for them to be valid is, as before, that the function Φ does not contain the 4 independent variables explicitly, and the proof follows along the same lines of reasoning we used when deriving equation (3). In this process, the structure of Mie's formulae for the energy quantities will emerge which, at first sight, is not readily apparent.

3 *Irish. Trans.* 21.

1. THE VARIATIONAL PRINCIPLE OF STATICS FOR A FOUR-DIMENSIONAL CONTINUUM

One will be able to describe the deformation of a four-dimensional continuum by giving the projections u_1, u_2, u_3, u_4, of the deformations of its points with respect to 4 orthogonal axes as functions of the coordinates x_1, x_2, x_3, x_4:

$$u_\alpha = u_\alpha(x_1, x_2, x_3, x_4), \qquad \alpha = 1, \ldots 4. \tag{4}$$

We further use the abbreviation

$$\frac{\partial u_\alpha}{\partial x_\beta} = a_{\alpha\beta}. \tag{5}$$

All properties of the continuum should now be determined through the function Φ of the displacements u_α and their derivatives $a_{\alpha\beta}$, and the resulting deformations should be determined by the requirement that variations of the four-dimensional integral over the four-dimensional space

$$\int\Phi(a_{11}, a_{12}, a_{13}, a_{14}; a_{21}, \ldots a_{44}; u_1, \ldots u_4)dx_1dx_2dx_3dx_4 \tag{6}$$

vanish. |

[27] If we now use the abbreviation[4]

$$\frac{\partial\Phi}{\partial a_{\alpha\beta}} = X_{\alpha\beta}, \qquad \frac{\partial\Phi}{\partial u_\alpha} = X_\alpha, \tag{7}$$

then this requirement yields the 4 differential equations:

$$\sum_\gamma \frac{\partial X_{\beta\gamma}}{\partial x_\gamma} - X_\beta = 0, \tag{8}$$

which express the requirement for equilibrium and correspond to equation (2) in the introduction.

2. FIRST SPECIAL CASE OF THE PRINCIPLE: HERGLOTZ' THEORY OF ELASTICITY

In the theory of relativity, x_1, x_2, x_3, represent the space coordinates and x_4 is the time multiplied by the imaginary unit i and the speed of light. The statics of the four-dimensional continuum is then nothing other than the dynamics of the three-dimensional one.

Therefore, the theory of elasticity, which has been adapted by Herglotz to satisfy the principle of relativity, must appear as a special case of our principle (6).

---

4 In the following all indices shall run through the values 1, 2, 3, 4, and all sums should extend over these values.

I will briefly outline how the quantities appearing in this process are to be interpreted and how the function Φ must be specified. The independent variables x_1, x_2, x_3, have to be considered as parameters ξ, η, ζ, which at a given instant fix the position of the points of the body; x_4 is set to $ic\tau$, where τ is a "time-like" parameter which otherwise is totally arbitrary. u_1, u_2, u_3, are the coordinates x, y, z, of the points of the body at an arbitrary time $t = u_4/ic$. Then, the quantities $a_{\alpha\beta}$, for $\alpha, \beta = 1, 2, 3$ are obviously determined by the strain in the body, whereas $a_{14}/a_{44}, a_{24}/a_{44}, a_{34}/a_{44}$ are the velocity components. The function Φ is now specified through the requirement that the integral (6) neither changes its value under a Lorentz transformation of the variables u_1, u_2, u_3, u_4 (rotation of the four-dimensional space) nor under a change of the time | parameter τ. Consequently, Φ is not a [28]
function of all 16 quantities $a_{\alpha\beta}$, but depends only on a combination of 6 of them, the "rest-deformations" $e_{11}, e_{22}, e_{33}, e_{23}, e_{31}, e_{12}$. These quantities, which I introduced first,[5] are a measure of the deformation of the volume element as measured by a co-moving observer. Most remarkable in this formulation is the absence of the kinetic energy. Instead, the velocities appear in the rest-deformations $e_{\alpha\beta}$. Herglotz extensively examined the laws of motion that arise from the interpretation of these quantities, equation (8), and showed that the ordinary mechanics of elastic bodies is a limiting case of this theory.

3. SECOND SPECIAL CASE OF THE PRINCIPLE: MIE'S ELECTRODYNAMICS

The theory of Mie is quite a different special case of the variational principle (6). Before we interpret the electrodynamic significance of the quantities, we want to present the characteristic specification of the function Φ, which shapes the entire theory: Φ shall only be a function of the differences

$$a_{\alpha\beta} - a_{\beta\alpha} = \frac{\partial u_\alpha}{\partial x_\beta} - \frac{\partial u_\beta}{\partial x_\alpha} \tag{9}$$

This formulation applied to three-dimensional space leads exactly to the theory of MacCullagh mentioned in the introduction. Therefore, one can interpret the formulae here in the same manner as is done there. The quantities (9) are namely the components of the infinitesimal rotation of the volume elements of the continuum, the "rotation components." In the theory of MacCullagh, the energy of the aether depends only on these rotations, but not on the deformation of the aether. It is clear that we can conceptualize Mie's theory in the same way if instead of aether we say "four-dimensional world." We leave it open whether a mechanistic interpretation in the usual sense of this formulation is possible and we restrict ourself to the assertion that it contains the entire electrodynamics of Mie (and, as a special case, also the classical electron theory).

5 *Ann. d. Phys.* (4), vol. 30, 1909, p. 1.

Let us now turn to the physical interpretation and description of the quantities
[29] appearing here. In Mie's theory, x_1, x_2, x_3, x_4 are | nothing but the coordinates and the time x, y, z, ict. Furthermore, Mie writes for

$$\begin{gathered} u_1, u_2, u_3, u_4 \\ f_x, f_y, f_z, i\varphi. \end{gathered} \tag{10}$$

These primary quantities[6] characterizing the aether correspond to the components of the four-potential in the theory of electrons.

The components of rotation (9) appear in Mie's theory as components of the 6-vector $(\mathfrak{b}, -i\mathfrak{e})$, where $\mathfrak{b}$ represents the magnetic induction and $\mathfrak{e}$ the electric field strength according to the scheme

$$(a_{\alpha\beta} - a_{\beta\alpha}) = \begin{vmatrix} 0 & -\mathfrak{b}_z & \mathfrak{b}_y & i\mathfrak{e}_x \\ \mathfrak{b}_z & 0 & -\mathfrak{b}_x & i\mathfrak{e}_y \\ -\mathfrak{b}_y & \mathfrak{b}_x & 0 & i\mathfrak{e}_z \\ -i\mathfrak{e}_x & -i\mathfrak{e}_y & -i\mathfrak{e}_z & 0 \end{vmatrix}. \tag{11}$$

This can also be written as

$$(\mathfrak{b}, -i\mathfrak{e}) = \mathfrak{Curl}(f, i\varphi) \tag{11'}$$

or as

$$\mathfrak{b} = \mathrm{curl} f, \qquad \mathfrak{e} = -\,\mathrm{grad}\varphi - \frac{\partial f}{\partial t}. \tag{11''}$$

With these symbols, Φ is seen to be a function of the components of the vectors $\mathfrak{b}, \mathfrak{e}, f$, and of the scalar φ:

$$\Phi(\mathfrak{b}_x, \mathfrak{b}_y, \mathfrak{b}_z,\ \mathfrak{e}_x, \mathfrak{e}_y, \mathfrak{e}_z;\ f_x, f_y, f_z, \varphi), \tag{6'}$$

where the fundamental assumption, that Φ depends only on the rotations $a_{\alpha\beta} - a_{\beta\alpha}$, is manifested. But at the same time, this also implies that the vectors $\mathfrak{e}, \mathfrak{b}$ satisfy the
[30] one quadruple of | Maxwell's equations, namely:

$$\mathrm{Div}(\mathfrak{b}, -i\mathfrak{e}) = 0 \tag{12}$$

---

6 In this presentation, f_x, f_y, f_z, φ, should be designated as "extensive quantities" [*Quantitätsgrößen*] since they have the character of displacement components of the four-dimensional continuum. The $\mathfrak{v}_x, \mathfrak{v}_y, \mathfrak{v}_z$, to be defined momentarily, would then be introduced as "intensive quantities" [*Intensitätsgrößen*]. That Mie proceeds here, as with the division of field-vectors into extensive and intensive quantities, in exactly the opposite way has its origin in the fact that the formulation of his expressions is closer to the physical conceptualization of electric density, displacement, field strength etc. In addition, Mie uses a different variational principle, which arises from ours via a Legendre transformation, and which readily suggests his choice of division of the quantities. Since Mie's variational principle requires additional conditions which cannot be readily incorporated into the formulation of the statics of the four-dimensional continuum, I have preferred the approach presented here.

or

$$\mathrm{curl}\,\mathfrak{e} + \frac{\partial \mathfrak{b}}{\partial t} = 0, \qquad \mathrm{div}\,\mathfrak{b} = 0; \tag{12'}$$

since these equations follow directly from $(11')$ and $(11'')$ respectively. The differential equations (8), however, are nothing but the second quadruple of Maxwell's equations. To show this we set, like Mie:[1]

$$\begin{aligned}
&\frac{\partial\Phi}{\partial\mathfrak{b}_x} = \mathfrak{h}_x, \qquad \frac{\partial\Phi}{\partial\mathfrak{e}_x} = -\mathfrak{d}_x,\\
&\frac{\partial\Phi}{\partial\mathfrak{b}_y} = \mathfrak{h}_y, \qquad \frac{\partial\Phi}{\partial\mathfrak{e}_y} = -\mathfrak{d}_y,\\
&\frac{\partial\Phi}{\partial\mathfrak{b}_z} = \mathfrak{h}_z, \qquad \frac{\partial\Phi}{\partial\mathfrak{e}_z} = -\mathfrak{d}_z,\\
&\frac{\partial\Phi}{\partial f_x} = -\mathfrak{v}_x, \qquad \frac{\partial\Phi}{\partial f_y} = -\mathfrak{v}_y, \qquad \frac{\partial\Phi}{\partial f_z} = -\mathfrak{v}_z, \qquad \frac{\partial\Phi}{\partial\varphi} = \rho.
\end{aligned} \tag{13}$$

Then, the 4 quantities X_α of the general theory become identified with the quantities $-\mathfrak{v}_x, -\mathfrak{v}_y, -\mathfrak{v}_z, -i\rho$, and the 16 quantities $X_{\alpha\beta}$ with the components of the vectors $\mathfrak{h}$ and $\mathfrak{d}$ as illustrated in the matrix equation

$$(X_{\alpha\beta}) = \begin{vmatrix} 0 & -\mathfrak{h}_z & \mathfrak{h}_y & i\mathfrak{d}_x \\ \mathfrak{h}_z & 0 & -\mathfrak{h}_x & i\mathfrak{d}_y \\ -\mathfrak{h}_y & \mathfrak{h}_x & 0 & i\mathfrak{d}_z \\ -i\mathfrak{d}_x & -i\mathfrak{d}_y & -i\mathfrak{d}_z & 0 \end{vmatrix} \tag{14}$$

With this notation the equations (8) turn into

$$\mathrm{Div}(\mathfrak{h}, -i\mathfrak{d}) = (\mathfrak{v}, i\rho) \tag{15}$$

or

$$\mathrm{curl}\,\mathfrak{h} - \frac{\partial\mathfrak{d}}{\partial t} = \mathfrak{v}, \qquad \mathrm{div}\,\mathfrak{d} = \rho. \tag{15'}$$

From these, one recognizes that ρ is the electric charge density, $\mathfrak{v}$ the convection current (charge times velocity), $\mathfrak{h}$ the magnetic field strength and $\mathfrak{d}$ the electric displacement. We also see that these quantities, according to equation (14), in terms of the picture of the | statics of the four-dimensional continuum, correspond to stresses [31]
and forces.

The equation of continuity for the electric current follows from equation (15)

$$\mathrm{Div}(\mathfrak{v}, i\rho) = 0 \tag{16}$$

or

$$\operatorname{div}\mathfrak{v} + \frac{\partial\rho}{\partial t} = 0. \tag{16'}$$

However, Φ is still an arbitrary function of its 10 arguments. We see that Maxwell's equations (12) and (15) are formally valid for any function Φ.

However, if one wants to maintain the validity of the principle of relativity, the choice of Φ has to be restricted. Obviously, Φ is then not allowed to depend explicitly on all of the 10 arguments, but only on such combinations of them as are invariant under Lorentz transformations. Mie has shown that there exist four such invariants which are independent of one another. In our representation we could for instance choose the following 4 invariants:

1. The length of the four-vectors $(f, i\varphi)$:

$$\chi = \sqrt{\varphi^2 - f^2},$$

2. The absolute magnitude of the six-vector $(\mathfrak{b}, -i\mathfrak{e})$:

$$\eta = \sqrt{\mathfrak{e}^2 - \mathfrak{b}^2},$$

3. The scalar product of the six-vector $(\mathrm{b}, -i\mathrm{e})$ with its dual vector $(-i\mathfrak{e}, \mathfrak{b})$:

$$\kappa = (\mathfrak{b}\mathfrak{e})$$

4. As the simultaneous invariant of the four-vector and of the six-vector one can take the square of the length of the four-vector obtained from the multiplication of the two original vectors:

$$\lambda^2 = ([f\mathfrak{b}] + \varphi\mathfrak{e})^2 - (f\mathfrak{e})^2.$$

[32] Φ can still be chosen as an arbitrary function of these 4 | arguments.

The aim of physical research then, as suggested by the theory of Mie, is to account, through an appropriate choice of the function $\Phi(\chi, \eta, \kappa, \lambda)$, for all the electromagnetic properties[7] of electrons and atoms.

In this, we have exactly the continuation of Lagrange's magnificent program.

The classical theory of electrons is formally a special case of Mie's theory, but not in the strict sense. Indeed one obtains its field equations by simply setting:

$$\Phi = \frac{1}{2}(\mathfrak{b}^2 - \mathfrak{e}^2) - (f\mathfrak{v}) + \varphi\rho, \tag{17}$$

where $\mathfrak{v}_x, \mathfrak{v}_y, \mathfrak{v}_z$, and ρ are considered to be *given* functions of space and time which describe the motion of the electrons.

But then Φ is no longer a function of only the 4 invariants $\chi, \eta, \kappa, \lambda$, but in addition depends explicitly on x, y, z, t, which however, is excluded in Mie's theory on

7 We are excluding gravitation here.

principle. In Mie's theory, the forces that hold electrons and atoms together should arise naturally from the formulation of Φ, whereas in the classical theory of electrons the forces have to be specifically added.

4. THE MOMENTUM-ENERGY LAW FOR THE GENERAL CASE OF THE FOUR-DIMENSIONAL CONTINUUM

The assumption of Mie just emphasized, that the function Φ is independent of x, y, z, t, is also the real mathematical reason for the validity of the momentum-energy-law.

In order to show that, we first consider, as in section 1, a general four-dimensional continuum whose equilibrium is determined by equation (8). We assert that for these differential equations, a law, analogous to the energy law $(3')$ of Lagrangian mechanics, is always valid as soon as one of the 4 coordinates x_α does not appear explicitly in Φ.

Then one obtains by differentiation of Φ with respect to x_α:

$$\frac{\partial\Phi}{\partial x_\alpha} = \sum_{\beta,\gamma} X_{\beta\gamma}\frac{\partial^2 u_\beta}{\partial x_\alpha \partial x_\gamma} + \sum_\beta X_\beta \frac{\partial u_\beta}{\partial x_\alpha},$$

and if one now adds equations (8) multiplied by the quantities $\dfrac{\partial u}{\partial x_\alpha}$ to the above equation one obtains: |

$$\frac{\partial\Phi}{\partial x_\alpha} = \sum_\gamma \frac{\partial}{\partial x_\gamma}\Bigl(\sum_\beta X_{\beta\gamma} a_{\beta\alpha}\Bigr). \tag{18}$$ [33]

This is the formula corresponding to the energy conservation law in mechanics. If Φ is independent of all 4 coordinates x_α, then (18) is valid for $\alpha = 1, 2, 3, 4$. These 4 equations are to be designated the momentum-energy theorem. They can also be summarized by the symbolic equation

$$\mathrm{Div}\, T = 0, \tag{18'}$$

if the 16 components $T_{\alpha\beta}$ of the matrix T are defined as

$$T_{\alpha\beta} = \Phi\delta_{\alpha\beta} - \sum_\gamma a_{\gamma\alpha} X_{\gamma\beta} \tag{19}$$

where

$$\delta_{\alpha\beta} = \begin{Bmatrix} 1 & \text{for} & \alpha = \beta \\ 0 & \text{for} & \alpha \neq \beta \end{Bmatrix}.$$

In the matrix calculus, (19) can be written as:

$$T = \Phi - \bar{a}X, \tag{19'}$$

where $\bar{a} = (a_{\beta\alpha})$ is the transpose of the matrix $a = (a_{\alpha\beta})$.[8]

5. SPECIAL FORM OF THE MOMENTUM-ENERGY EQUATION FOR THE CASE OF MIE'S ELECTRODYNAMICS.

The equations described by Mie as the momentum-energy equations are essentially nothing but the general equations (18) and $(18')$ respectively. A minor mathematical transformation leads to the formulae of Mie. In order to see why the transformation is necessary, it is best to consider the stress-energy tensor in the succinct symbolic form $(19')$. Keeping in mind the electrodynamic significance of the quantities $a_{\alpha\beta}$ and $X_{\alpha\beta}$ (equations (11) and (14)), we see that although the quantities $X_{\alpha\beta}$ can be expressed directly through the components of the field vectors, the $a_{\alpha\beta}$ cannot; rather, only the combinations $a_{\alpha\beta} - a_{\beta\alpha}$ whose matrix is to be denoted by $a - \bar{a}$ have a physical meaning. Therefore, we will have to transform equation $(19')$ in such a manner that it contains the difference-matrix $a - \bar{a}$. |

[34] Naturally, we may not simply add aX to T, because then $(18')$ would cease to be valid. Nevertheless, one can try to define a matrix ψ in such a way that the divergence equations $(18')$ remains valid for the matrix

$$S = \Phi + (a - \bar{a})X + \psi. \tag{20}$$

If we denote the added matrix $\psi + aX$ by ω, so that $S = T + \omega$, then we also require that

$$\mathrm{Div}\,\omega = 0. \tag{21}$$

We now show that (21) is satisfied by the matrix

$$\psi_{\alpha\beta} = -u_\alpha X_\beta \tag{22}$$

provided that the matrix X is skew-symmetric:

$$X_{\alpha\beta} = -X_{\beta\alpha}, \quad \text{or} \quad X = -\bar{X}. \tag{23}$$

Then, because of (8), we have

$$\begin{aligned}\omega_{\alpha\beta} &= \sum_\gamma a_{\alpha\gamma} X_{\gamma\beta} - u_\alpha X_\beta \\ &= \sum_\gamma \left(\frac{\partial u_\alpha}{\partial x_\gamma} X_{\gamma\beta} + u_\alpha \frac{\partial X_{\gamma\beta}}{\partial x_\gamma} \right) \\ &= \sum_\gamma \frac{\partial}{\partial x_\gamma} (u_\alpha X_{\alpha\beta})\end{aligned}$$

8 The product of two matrices is that matrix whose element with the subscripts α, β arises from multiplying the row α by the column β.

and by using (23) again

$$\sum_{\beta} \frac{\partial \omega_{\alpha\beta}}{\partial x_{\beta}} = \sum_{\beta\gamma} \frac{\partial^2 u_{\alpha} X_{\gamma\beta}}{\partial x_{\beta} \partial x_{\gamma}} = -\sum_{\beta\gamma} \frac{\partial^2 u_{\alpha} X_{\gamma\beta}}{\partial x_{\beta} \partial x_{\gamma}} = 0;$$

i.e. equation (21) is satisfied.

As a glance at the scheme (14) shows, the condition (23) is met in Mie's theory precisely because of the requirement that Φ is only a function of the differences $a_{\alpha\beta} - a_{\beta\alpha}$. Hence, one can write the energy-momentum equation in the form

$$\mathrm{Div} S = 0, \tag{24}$$

where S is defined by (20) and (22).

The mathematical structure of the law is especially transparent in this form. |

If we introduce the electromagnetic notation, then the matrix equation (20) becomes [35]

$$S = \begin{vmatrix} \Phi & 0 & 0 & 0 \\ 0 & \Phi & 0 & 0 \\ 0 & 0 & \Phi & 0 \\ 0 & 0 & 0 & \Phi \end{vmatrix} + \begin{vmatrix} 0 & -\mathfrak{b}_z & \mathfrak{b}_y & i\mathfrak{e}_x \\ \mathfrak{b}_z & 0 & -\mathfrak{b}_x & i\mathfrak{e}_y \\ -\mathfrak{b}_z & \mathfrak{b}_x & 0 & i\mathfrak{e}_z \\ -i\mathfrak{e}_x & -i\mathfrak{e}_y & -i\mathfrak{e}_z & 0 \end{vmatrix} \cdot \begin{vmatrix} 0 & \mathfrak{h}_z & \mathfrak{h}_y & i\mathfrak{d}_x \\ \mathfrak{h}_z & 0 & -\mathfrak{h}_x & i\mathfrak{d}_y \\ -\mathfrak{h}_y & \mathfrak{h}_x & 0 & i\mathfrak{d}_z \\ -i\mathfrak{d}_x & -i\mathfrak{d}_y & -i\mathfrak{d}_z & 0 \end{vmatrix}$$
$$+ \begin{vmatrix} f_x\mathfrak{v}_x & f_x\mathfrak{v}_y & f_x\mathfrak{v}_z & if_x\rho \\ f_y\mathfrak{v}_x & f_y\mathfrak{v}_y & f_y\mathfrak{v}_z & if_y\rho \\ f_z\mathfrak{v}_x & f_z\mathfrak{v}_y & f_z\mathfrak{v}_z & if_z\rho \\ i\varphi\mathfrak{v}_x & i\varphi\mathfrak{v}_y & i\varphi\mathfrak{v}_z & -\varphi\rho \end{vmatrix} \tag{20'}$$

or carrying out the multiplication:

$$S = \begin{vmatrix} \Phi - \mathfrak{b}_y\mathfrak{h}_y - \mathfrak{b}_z\mathfrak{h}_z + \mathfrak{e}_x\mathfrak{d}_x + f_x\mathfrak{v}_x, & \mathfrak{e}_x\mathfrak{d}_y + \mathfrak{h}_x\mathfrak{b}_y + f_x\mathfrak{v}_y, & \mathfrak{e}_x\mathfrak{d}_z + \mathfrak{h}_x\mathfrak{b}_z + f_x\mathfrak{v}_z, & -i(\mathfrak{d}_y\mathfrak{b}_z - \mathfrak{d}_z\mathfrak{b}_y - \rho f_x), \\ \mathfrak{e}_y\mathfrak{d}_x + \mathfrak{h}_y\mathfrak{b}_x + f_y\mathfrak{v}_x, & \Phi - \mathfrak{b}_z\mathfrak{h}_z - \mathfrak{b}_x\mathfrak{h}_x + \mathfrak{e}_y\mathfrak{d}_y + f_y\mathfrak{v}_y, & \mathfrak{e}_y\mathfrak{d}_z + \mathfrak{h}_y\mathfrak{b}_z + f_y\mathfrak{v}_z, & -i(\mathfrak{d}_z\mathfrak{b}_x - \mathfrak{d}_x\mathfrak{b}_z - \rho f_y), \\ \mathfrak{e}_z\mathfrak{d}_x + \mathfrak{h}_z\mathfrak{b}_x + f_z\mathfrak{v}_x, & \mathfrak{e}_z\mathfrak{d}_y + \mathfrak{h}_z\mathfrak{b}_y + f_z\mathfrak{v}_y, & \Phi - \mathfrak{b}_x\mathfrak{h}_x - \mathfrak{b}_y\mathfrak{h}_y + \mathfrak{e}_z\mathfrak{d}_z + f_z v_z, & -i(\mathfrak{d}_x\mathfrak{b}_y - \mathfrak{d}_y\mathfrak{b}_x - \rho f_z), \\ -i(\mathfrak{e}_y\mathfrak{h}_z - \mathfrak{e}_z\mathfrak{h}_y - \varphi\mathfrak{v}_x), & -i(\mathfrak{e}_z\mathfrak{h}_x - \mathfrak{e}_x\mathfrak{h}_z - \varphi\mathfrak{v}_y), & -i(\mathfrak{e}_x\mathfrak{h}_y - \mathfrak{e}_y\mathfrak{h}_x - \varphi\mathfrak{v}_z), & \Phi + \mathfrak{e}_x\mathfrak{d}_x + \mathfrak{e}_y\mathfrak{d}_y + \mathfrak{e}_z\mathfrak{d}_z - \varphi\rho \end{vmatrix} \tag{20''}$$

This is precisely the stress-energy matrix presented by Mie.

Mie then showed that this matrix is symmetric provided Φ depends only on the 4 invariants $\chi, \eta, \kappa, \lambda$. This proof, which is carried out by simple calculation, cannot be significantly simplified by our method of presentation.

It is perhaps not superfluous to emphasize that the energy-momentum equation of the classical electron theory does not arise as a special case by using Φ in $(20'')$ as formulated in (17), because then Φ is not independent of x, y, z, t, since v and ρ depend on position and time and thus, our line of proof becomes invalid. One can also easily see, by substituting in (24) for Φ as formulated in (17), that the result is at variance with the energy-momentum law of the classical electron theory. However, if one adds to (24) the terms that arise by differentiating (17) with respect to x, y, z, t, which arise because of their dependence on $\mathfrak{v}$ and ρ, and which cannot be written in
[36] the form of a four-dimensional | divergence, then one obtains the energy-momentum law of the electron theory in its usual form. With respect to the corresponding question in the electrodynamic theory of moving material bodies the same is to be said. None of the available formulations for the stress-energy-matrix, neither Minkowski's unsymmetric one, nor Abraham and Laue's symmetric one, fall directly under Mie's scheme, yet the same method can be employed here as well.

EDITORIAL NOTE

[1] In the last line of eqs. (13), ∂f_x is misprinted in the original as $\partial \mathfrak{b}_x$.

INCLUDING GRAVITATION IN A
UNIFIED THEORY OF PHYSICS

LEO CORRY

THE ORIGIN OF HILBERT'S AXIOMATIC METHOD[1]

1. AXIOMATICS, GEOMETRY AND PHYSICS IN HILBERT'S EARLY LECTURES

This chapter examines how Hilbert's axiomatic approach gradually consolidated over the last decade of the nineteenth century. It goes on to explore the way this approach was actually manifest in its earlier implementations.

Although geometry was not Hilbert's main area of interest before 1900, he did teach several courses on this topic back in Königsberg and then in Göttingen. His lecture notes allow an illuminating foray into the development of Hilbert's ideas and they cast light on how his axiomatic views developed.[2]

1.1 Geometry in Königsberg

Hilbert taught projective geometry for the first time in 1891 (Hilbert 1891). What already characterizes Hilbert's presentation of geometry in 1891, and will remain true later on, is his clearly stated conception of this science as a natural one in which, at variance with other mathematical domains, sensorial intuition— *Anschauung*—plays a fundamental role that cannot be relinquished. In the introduction to the course, Hilbert formulated it in the following words:

> Geometry is the science that deals with the properties of space. It differs essentially from pure mathematical domains such as the theory of numbers, algebra, or the theory of functions. The results of the latter are obtained through pure thinking... The situation is completely different in the case of geometry. I can never penetrate the properties of space by pure reflection, much as I can never recognize the basic laws of mechanics, the law of gravitation or any other physical law in this way. Space is not a product of my reflections. Rather, it is given to me through the senses. I thus need my senses in order to fathom its properties. I need intuition and experiment, just as I need them in order to figure out physical laws, where also matter is added as given through the senses.[3]

---

1 This chapter is based on extracts from (Corry 2004), in particular on chapters 2, 3, and 5.
2 An exhaustive analysis of the origins of *Grundlagen der Geometrie* based on these lecture notes and other relevant documents was first published in (Toepell 1986). Here we draw directly from this source.
3 The German original is quoted in (Toepell 1986, 21). Similar testimonies can be found in many other manuscripts of Hilbert's lectures. Cf., e.g., (Toepell 1986, 58).

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 4
Gravitation in the Twilight of Classical Physics: The Promise of Mathematics.

The most basic propositions related to this intuition concern the properties of incidence, and in order to express them conveniently it is necessary to introduce "ideal elements." Hilbert stressed that these are to be used here only as a shorthand with no metaphysical connotations.

In the closing passage of his lecture, Hilbert briefly discussed the connections between analytic and projective geometry. While the theorems and proofs of the former are more general than those of the latter, he said, the methods of the latter are much purer, self-contained, and necessary.[4] By combining synthetic and axiomatic approaches, Hilbert hinted, it should be possible, perhaps, to establish a clear connection between these two branches of the discipline.

In September of that year, Hilbert attended the *Deutsche Mathematiker-Vereinigung* meeting in Halle, where Hermann Wiener (1857–1939) lectured on the foundations of geometry.[5] The lecture could not fail to attract Hilbert's attention given his current teaching interests. Blumenthal reported in 1935 that Hilbert came out greatly excited by what he had just heard, and made his famous declaration that it must be possible to replace "point, line, and plane" with "table, chair, and beer mug" without thereby changing the validity of the theorems of geometry (Blumenthal 1935, 402–403). Seen from the point of view of later developments and what came to be considered the innovative character of *Grundlagen der Geometrie,* this may have been indeed a reason for Hilbert's enthusiasm following the lecture. If we also recall the main points of interest in his 1891 lectures, however, we can assume that Wiener's claim about the possibility of proving central theorems of projective geometry without continuity considerations exerted no lesser impact, and perhaps even a greater one, on Hilbert at the time. Moreover, the idea of changing names of the central concepts while leaving the deductive structure intact was an idea that Hilbert already knew, if not from other, earlier mathematical sources, then at least from his attentive reading of the relevant passages in Dedekind's *Was sind und was sollen die Zahlen?*,[6] where he may not have failed to see the introductory remarks on the role of continuity in geometry. If Hilbert's famous declaration was actually pronounced for the first time after this lecture, as Blumenthal reported, one can then perhaps conclude that Wiener's ideas were more than just a revelation for Hilbert, but acted as a catalyst binding together several threads that may have already been present in his mind for a while.

Roughly at the time when Hilbert's research efforts started to focus on the theory of algebraic number fields, from 1893 on, his interest regarding the foundations of geometry also became more intensive, at least at the level of teaching. In preparing a course on non-Euclidean geometry to be taught that year, Hilbert was already adopting a more axiomatic perspective. The original manuscript of the course clearly reveals that Hilbert had decided to follow more closely the model put forward by Pasch. As for the latter, using the axiomatic approach was a direct expression of a nat-

4 Cf. (Toepell 1986, 37).
5 He may have also attended Wiener's second lecture in 1893. Cf. (Rowe 1999, 556).
6 As we know from a letter to Paul du Bois-Reymond of March-April, 1888. Cf. (Dugac 1976, 203).

uralistic approach to geometry, rather than a formalistic one: the axioms of geometry—Hilbert wrote—express observations of facts of experience, which are so simple that they need no additional confirmation by physicists in the laboratory.[7] From his correspondence with Felix Klein (1849–1925),[8] however, we learn that Hilbert soon realized certain shortcomings in Pasch's treatment, and in particular, certain redundancies that affected it. Hilbert explicitly stipulated at this early stage that a successful axiomatic analysis should aim to establish the *minimal* set of presuppositions from which the whole of geometry could be deduced. Such a task had not been fully accomplished by Pasch himself, Hilbert pointed out, since his Archimedean axiom, could be derived from others in his system.

Hilbert's correspondence also reveals that he kept thinking about the correct way to implement an axiomatic analysis of geometry. In a further letter to Klein, on 15 November while criticizing Lie's approach to the foundations of geometry, he formulated additional tasks to be accomplished by such an analysis. He thus wrote:

> It seems to me that Lie always introduces into the issue a preconceived one-sidedly analytic viewpoint and forgets completely the principal task of non-Euclidean geometry, namely, that of constructing the various possible geometries by the successive introduction of elementary axioms, up until the final construction of the only remaining one, Euclidean geometry.[9]

The course on non-Euclidean geometry was not taught as planned in 1893, since only one student registered for it.[10] It did take place the following year, announced as "Foundations of Geometry." Hilbert had meanwhile considerably broadened his reading in the field, as indicated by the list of almost forty references mentioned in the notes. This list included most of the recent, relevant foundational works. A clear preference for works that followed an empiricist approach is evident, but also articles presenting the ideas of Grassmann were included.[11] It is not absolutely clear to what extent Hilbert read Italian, but none of the current Italian works were included in his list, except for a translated text of Peano (being the only one by a non-German author).[12] It seems quite certain, at any rate, that Hilbert was unaware of the recent works of Fano, Veronese, and others, works that could have been of great interest for him in the direction he was now following.

---

7 "Das Axiom entspricht einer Beobachtung, wie sich leicht durch Kugeln, Lineal und Pappdeckel zeigen lässt. Doch sind diese Erfahrungsthatsachen so einfach, von Jedem so oft beobachtet und daher so bekannt, dass der Physiker sie nicht extra im Laboratorium bestätigen darf." (Hilbert 1893–1894, 10)

8 Hilbert to Klein, 23 May 1893. Quoted in (Frei 1985, 89–90).

9 Hilbert to Klein, 15 November 1893. Quoted in (Frei 1985, 101). On 11 November, he wrote an almost identical letter to Lindemann. Cf. (Toepell 1986, 47).

10 Cf. (Toepell 1986, 51).

11 The full bibliographical list appears in (Toepell 1986, 53–55).

12 At the 1893 annual meeting of the *Deutsche Mathematiker- Vereinigung* in Lübeck (16–20 September), Frege discussed Peano's conceptual language. If not earlier than that, Hilbert certainly heard about Peano's ideas at this opportunity, when he and Minkowski also presented the plans for their expected reports on the theory of numbers. Cf. *Jahresbericht der Deutschen Mathematiker-Vereinigung,* Vol. 4 (1894–1895), p. 8.

Hilbert became acquainted with Hertz's book on the foundations of mechanics, though it was not mentioned in the list. This book seems to have provided a final, significant catalyst for the wholehearted adoption of the axiomatic perspective for geometry. Simultaneously the book established, in Hilbert's view, a direct connection between the latter and the axiomatization of physics in general. Moreover, Hilbert adopted Hertz's more specific, methodological ideas about what is actually involved in axiomatizing a theory. The very fact that Hilbert came to hear about Hertz is not surprising; he would probably have read Hertz's book sooner or later. But that he read it so early was undoubtedly due to Minkowski. During his Bonn years, Minkowski felt closer to Hertz and to his work than to anyone else, and according to Hilbert, his friend had explicitly declared that, had it not been for Hertz's untimely death, he would have dedicated himself exclusively to physics.[13]

Just as with many other aspects of Hilbert's early work, there is every reason to believe that Minkowski's enthusiasm for Hertz was transmitted to his friend. When revising the lecture notes for his course, Hilbert added the following comment:

> Nevertheless the origin [of geometrical knowledge] is in experience. The axioms are, as Hertz would say, pictures or symbols in our mind, such that consequents of the images are again images of the consequences, i.e., what we can logically deduce from the images is itself valid in nature.[14]

Hilbert defined the task to be pursued as part of the axiomatic analysis, including the need to establish the independence of the axioms of geometry. In doing so, however, he stressed once again the objective and factual character of this science. Hilbert wrote:

> The problem can be formulated as follows: What are the necessary, sufficient, and mutually independent conditions that must be postulated for a system of things, in order that any of their properties correspond to a geometrical fact and, conversely, in order that a complete description and arrangement of all the geometrical facts be possible by means of this system of things.[15]

But already at this point it is absolutely clear that, for Hilbert, such questions were not just abstract tasks. Rather, he was directly focused on important, open problems of the discipline, and in particular, on the role of the axiom of continuity in the questions of coordinatization and metrization in projective geometry, as well as in the proof of the fundamental theorems. In a passage that was eventually crossed out, Hilbert expressed his doubts about the prospects of actually proving Wiener's assertion that continuity considerations could be circumvented in projective geometry (Toepell

13 See (Hilbert 1932–1935, 3: 355). Unfortunately, there seems to be no independent confirmation of Minkowski's own statement to this effect.

14 "Dennoch der Ursprung aus der Erfahrung. Die Axiome sind, wie Herz [sic] sagen würde, Bilde[r] oder Symbole in unserem Geiste, so dass Folgen der Bilder wieder Bilder der Folgen sind d.h. was wir aus den Bildern logisch ableiten, stimmt wieder in der Natur." It is worth noting that Hilbert's quotation of Hertz, drawn from memory, was somewhat inaccurate. I am indebted to Ulrich Majer for calling my attention to this passage. (Hilbert 1893–1894, 10)

15 Quoted from the original in (Toepell 1986, 58–59).

1986, 78). Eventually, however, a main achievement of *Grundlagen der Geometrie* would be a detailed realization of this possibility and its consequences, but Hilbert probably decided to follow this direction only after hearing about the result of Friedrich Schur (1856–1932) in 1898. I return to this matter in the next section.

Concerning the validity of the parallel axiom, Hilbert adopted in 1893–1894 a thoroughly empirical approach that reminds us very much of Riemann's *Habilitationsschrift*. Hilbert referred also directly to Gauss's experimental measurement of the sum of angles of the triangle described by three Hannoverian mountain peaks.[16] Although Gauss's measurements were convincing enough for Hilbert to indicate the correctness of Euclidean geometry as a true description of physical space, he still saw an open possibility that future measurements would show it to be otherwise. Hilbert also indicated that existing astronomical observations are not decisive in this respect, and therefore the parallel axiom must be taken at least as a limiting case. In his later lectures on physics, Hilbert would return to this example very often to illustrate the use of axiomatics in physics. In the case of geometry, this particular axiom alone might be susceptible to change following possible new experimental discoveries. Thus, what makes geometry especially amenable to a full axiomatic analysis is the very advanced stage of development it has attained, rather than any other specific, essential trait concerning its nature. In all other respects, geometry is like any other natural science. Hilbert thus stated:

> Among the appearances or facts of experience manifest to us in the observation of nature, there is a peculiar type, namely, those facts concerning the outer shape of things. Geometry deals with these facts. ... Geometry is a science whose essentials are developed to such a degree, that all its facts can already be logically deduced from earlier ones. Much different is the case with the theory of electricity or with optics, in which still many new facts are being discovered. Nevertheless, with regards to its origins, geometry is a natural science.[17]

It is the very process of axiomatization that transforms the natural science of geometry, with its factual, empirical content, into a pure mathematical science. There is no apparent reason why a similar process might not be applied to any other natural science. And in fact, from very early on Hilbert made it clear that this should be done. In the manuscript of his lectures we read that "all other sciences—above all mechanics, but subsequently also optics, the theory of electricity, etc.—should be treated according to the model set forth in geometry."[18]

---

16 The view that Gauss considered his measurement as related to the question of the parallel axiom has been questioned in (Breitenberger 1984) and (Miller 1972). They have argued that this measurement came strictly as a part of Gauss's geodetic investigations. For replies to this argument, see (Scholz 1993, 642–644), and a more recent and comprehensive discussion in (Scholz 2004). Hilbert, at any rate, certainly believed that this had been Gauss's actual intention, and he repeated this opinion on many occasions.

17 Quoted in (Toepell 1986, 58).

18 Quoted in (Toepell 1986, 94).

By 1894, then, Hilbert's interest in foundational issues of geometry had increased considerably, and he had embarked more clearly in an axiomatic direction. His acquaintance with Hertz's ideas helped him conceive the axiomatic treatment of geometry as part of a larger enterprise, relevant also for other physical theories. It also offered methodological guidelines for actually implementing this analysis. However, many of the most important foundational problems remained unsettled for him, and in this sense, even the axiomatic approach did not seem to him to be of great help. At this stage he saw in the axiomatic method no more than an exercise in adding or deleting basic propositions and guessing the consequences that would follow, but certainly not a tool for achieving real new results.[19]

1.2 Geometry in Göttingen

Hilbert moved to Göttingen in 1895 and thereafter he dedicated himself almost exclusively to number theory both in his research and in his teaching. It is worth pointing out, that some of the ideas he developed in this discipline would prove to be essential some years later for his treatment of geometry as presented in *Grundlagen der Geometrie*. In particular, Hilbert's work on the representation of algebraic forms as sums of squares, which had a deep influence on the subsequent development of the theory of real fields,[20] also became essential for Hilbert's own ideas on geometrical constructivity as manifest in *Grundlagen der Geometrie*.

In the summer semester of 1899, Hilbert once again taught a course on the elements of Euclidean geometry. The elaboration of these lectures would soon turn into the famous *Grundlagen der Geometrie*. The very announcement of the course came as a surprise to many in Göttingen, since it signified, on the face of it, a sharp departure from the two fields in which he had excelled since completing his dissertation in 1885: the theory of algebraic invariants and the theory of algebraic number fields. As Blumenthal recalled many years later:

> [The announcement] aroused great excitement among the students, since even the veteran participants of the 'number theoretical walks' (*Zahlkörpersspaziergängen*) had never noticed that Hilbert occupied himself with geometrical questions. He spoke to us only about fields of numbers. (Blumenthal 1935, 402)

Also Hermann Weyl (1855–1955) repeated this view in his 1944 obituary:

> [T]here could not have been a more complete break than the one dividing Hilbert's last paper on the theory of number fields from his classical book *Grundlagen der Geometrie*. (Weyl 1944, 635)

As already suggested, however, the break may have been less sharp than it appeared in retrospect to Hilbert's two distinguished students. Not only because of the strong connections of certain, central results of *Grundlagen der Geometrie* to Hil-

19 As expressed in a letter to Hurwitz, 6 June 1894. See (Toepell 1986, 100).
20 Cf. (Sinaceur 1984, 271–274; 1991, 199–254).

bert's number-theoretical works, or because of Hilbert's earlier geometry courses in Königsberg, but also because Hilbert became actively and intensely involved in current discussions on the foundations of projective geometry starting in early 1898. In fact, at that time Hilbert had attended a lecture in Göttingen given by Schoenflies who discussed a result recently communicated by Schur to Klein, according to which Pappus's theorem could be proven starting from the axioms of congruence alone, and therefore without relying on continuity considerations.[21] Encouraged by this result, and returning to questions that had been raised when he taught the topic several years earlier, Hilbert began to elaborate on this idea in various possible alternative directions. At some point, he even thought, erroneously as it turned out, to have proved that it would suffice to assume Desargues's theorem in order to prove Pappus's theorem.[22]

Schur's result provided the definitive motivation that led Hilbert to embark on an effort to elucidate in detail the fine structure of the logical interdependence of the various fundamental theorems of projective and Euclidean geometry and, more generally, of the structure of the various kinds of geometries that can be produced under various sets of assumptions. The axiomatic method, whose tasks and basic tools Hilbert had been steadily pondering, would now emerge as a powerful and effective instrument for properly addressing these important issues.

The course of 1899 contains much of what will appear in *Grundlagen der Geometrie*. It is worth pointing out here that in the opening lecture Hilbert stated once again the main achievement he expected to obtain from an axiomatic analysis of the foundations of geometry: a complete description, by means of independent statements, of the basic facts from which all known theorems of geometry can be derived. This time he also mentioned the precise source from which this formulation had been taken: the introduction to Hertz's *Principles of Mechanics*.[23] In Hilbert's view, this kind of task was not limited to geometry, and of course also applied, above all, to mechanics. Hilbert had taught seminars on mechanics jointly with Klein in 1897–1898. In the winter semester 1898–1899, he also taught his first full course on a physical topic in Göttingen: mechanics.[24] In the introduction to this course, he explicitly stressed the essential affinity between geometry and the natural sciences, and also explained the role that axiomatization should play in the mathematization of the latter. He compared the two domains in the following terms:

> Geometry also [like mechanics] emerges from the observation of nature, from experience. To this extent, it is an *experimental science*. ... But its experimental foundations are

---

21 Later published as (Schur 1898).

22 Cf. (Toepell 1986, 114–122). Hessenberg (1905) proves that, in fact, it is Pappus's theorem that implies Desargues's, and not the other way round.

23 Cf. (Toepell 1986, 204).

24 According to the *Nachlass* David Hilbert (Niedersächsische Staats- und Universitätsbibliothek Göttingen, Abteilung Handschriften und Seltene Drucke), (Cod. Ms. D. Hilbert, 520), which contains a list of Hilbert's lectures between 1886 and 1932 (handwritten by Hilbert himself up until 1917–1918), among the earliest courses taught by Hilbert in Königsberg was one in hydrodynamics (summer semester, 1887).

> so irrefutably and so *generally acknowledged*, they have been confirmed to such a degree, that no further proof of them is deemed necessary. Moreover, all that is needed is to derive these foundations from a minimal set of *independent axioms* and thus to construct the whole edifice of geometry by *purely logical means*. In this way [i.e., by means of the axiomatic treatment] geometry is turned into a pure mathematical science. In mechanics it is also the case that all physicists recognize its most *basic facts*. But the *arrangement* of the basic concepts is still subject to a change in perception... and therefore mechanics cannot yet be described today as a *pure mathematical* discipline, at least to the same extent that geometry is. We must strive that it becomes one. We must ever stretch the limits of pure mathematics wider, on behalf not only of our mathematical interest, but rather of the interest of science in general.[25]

This is perhaps the first explicit presentation of Hilbert's program for axiomatizing natural science in general. The more definitive status of the results of geometry, as compared to the relatively uncertain one of our knowledge of mechanics, clearly recalls similar claims made by Hertz. The difference between geometry and other physical sciences—mechanics in this case—was not for Hilbert one of essence, but rather one of historical stage of development. He saw no reason in principle why an axiomatic analysis of the kind he was then developing for geometry could not eventually be applied to mechanics with similar, useful consequences. Eventually, that is to say, when mechanics would attain a degree of development equal to geometry, in terms of the quantity and certainty of known results, and in terms of an appreciation of what really are the "basic facts" on which the theory is based.

2. GRUNDLAGEN DER GEOMETRIE

When Hilbert published his 1899 *Festschrift* (Hilbert 1899) he was actually contributing a further link to a long chain of developments in the foundations of geometry that spanned several decades over the nineteenth century. His works on invariant theory and number theory can be described in similar terms, each within its own field of relevance. In these two fields, as in the foundations of geometry, Hilbert's contribution can be characterized as the "critical" phase in the development of the discipline: a phase in which the basic assumptions and their specific roles are meticulously inspected in order to revamp the whole structure of the theory on a logically sound

25 "Auch die Geometrie ist aus der Betrachtung der Natur, aus der Erfahrung hervorgegangen und insofern eine *Experimentalwissenschaft*. ... Aber diese experimentellen Grundlagen sind so unumstösslich und so *allgemein anerkannt,* haben sich so überall bewährt, dass es einer weiteren experimentellen Prüfung nicht mehr bedarf und vielmehr alles darauf ankommt diese Grundlagen auf ein geringstes Mass *unabhängiger Axiome* zurückzuführen und hierauf *rein logisch* den ganzen Bau der Geometrie aufzuführen. Also Geometrie ist dadurch eine rein *mathematische* Wiss. geworden. Auch in der Mechanik werden die *Grundthatsachen* von allen Physikern zwar anerkannt. Aber die *Anordnung* der Grundbegriffe ist dennoch dem Wechsel der Auffassungen unterworfen... so dass die Mechanik auch heute noch nicht, jedenfalls nicht in dem Masse wie die Geometrie als eine *rein mathematische* Disciplin zu bezeichnen ist. Wir müssen streben, dass sie es wird. Wir müssen die Grenzen echter Math. immer weiter ziehen nicht nur in unserem math. Interesse sondern im Interesse der Wissenschaft überhaupt." (Hilbert 1898–1899, 1–3)

basis and within a logically transparent deductive structure. This time, however, Hilbert had consolidated the critical point of view into an elaborate approach with clearly formulated aims, and affording the proper tools to achieve those aims, at least partly. This was the axiomatic approach that characterizes *Grundlagen der Geometrie* and much of his work thereafter, particularly his research on the foundations of physical theories. However, *Grundlagen der Geometrie* was innovative not only at the methodological level. It was, in fact, a seminal contribution to the discipline, based on a purely synthetic, completely new approach to arithmetizing the various kinds of geometries. And again, as in his two previous fields of research, Hilbert's in-depth acquaintance with the arithmetic of fields of algebraic numbers played a fundamental role in his achievement.

It is important to bear in mind that, in spite of the rigor required for the axiomatic analysis underlying *Grundlagen der Geometrie*, many additions, corrections and improvements—by Hilbert himself, by some of his collaborators and by other mathematicians as well— were still needed over the following years before the goals of this demanding project could be fully attained. Still most of these changes, however important, concerned only the details. The basic structure, the groups of axioms, the theorems considered, and above all, the innovative methodological approach implied by the treatment, all these remained unchanged through the many editions of *Grundlagen der Geometrie*.

The motto of the book was a quotation taken from Kant's *Critique of Pure Reason*: "All human knowledge thus begins with intuitions, proceeds thence to concepts and ends with ideas." If he had to make a choice, Kant appears an almost obvious one for Hilbert in this context. It is hard to state precisely, however, to what extent he had had the patience to become really acquainted with the details of Kant's exacting works. Beyond the well-deserved tribute to his most distinguished fellow Königsberger, this quotation does not seem to offer a reference point for better understanding Hilbert's ideas on geometry.

Hilbert described the aim of his *Festschrift* as an attempt to lay down a "simple" and "complete" system of "mutually independent" axioms, from which all known theorems of geometry might be deduced. His axioms are formulated for three systems of undefined objects named "points," "lines," and "planes," and they establish mutual relations that these objects must satisfy. The axioms are divided into five groups: axioms of incidence, of order, of congruence, of parallels, and of continuity. From a purely logical point of view, the groups have no real significance in themselves. However, from the geometrical point of view they are highly significant, for they reflect Hilbert's actual conception of the axioms as an expression of spatial intuition: each group expresses a particular way that these intuitions manifest themselves in our understanding.

2.1 Independence, Simplicity, Completeness, Consistency

Hilbert's first requirement, that the axioms be independent, is the direct manifestation of the foundational concerns that directed his research. When analyzing independence, his interest focused mainly on the axioms of congruence, continuity and of parallels, since this independence would specifically explain how the various basic theorems of Euclidean and projective geometry are logically interrelated. But as we have seen, this requirement had already appeared—albeit more vaguely formulated—in Hilbert's early lectures on geometry, as a direct echo of Hertz's demand for appropriateness. In *Grundlagen der Geometrie*, the requirement of independence not only appeared more clearly formulated, but Hilbert also provided the tools to prove systematically the mutual independence among the individual axioms within the groups and among the various groups of axioms in the system. He did so by introducing the method that has since become standard: he constructed models of geometries that fail to satisfy a given axiom of the system but satisfy all the others. However, this was not for Hilbert an exercise in analyzing abstract relations among systems of axioms and their possible models. The motivation for enquiring about the mutual independence of the axioms remained, essentially, a geometrical one. For this reason, Hilbert's original system of axioms was not the most economical one from the logical point of view. Indeed, several mathematicians noticed quite soon that Hilbert's system of axioms, seen as a single collection rather than as a collection of five groups, contained a certain degree of redundancy.[26] Hilbert's own aim was to establish the interrelations among the groups of axioms, embodying the various manifestations of special intuition, rather than among individual axioms belonging to different groups.

The second requirement, simplicity, complements that of independence. It means, roughly, that an axiom should contain "no more than a single idea." This is a requirement that Hertz also had explicitly formulated, and Hilbert seemed to be repeating it in the introduction to his own book. Nevertheless, it was neither formally defined nor otherwise realized in any clearly identifiable way within *Grundlagen der Geometrie*. The ideal of formulating "simple" axioms as part of this system was present implicitly as an aesthetic desideratum that was not transformed into a mathematically controllable feature.[27]

The "completeness" that Hilbert demanded for his system of axioms should not be confused with the later, model-theoretical notion that bears the same name, a

---

26 Cf., for instance, (Schur 1901). For a more detailed analysis of this issue, see (Schmidt 1933, 406–408). It is worth pointing out that in the first edition of *Grundlagen der Geometrie* Hilbert stated that he intended to provide an independent system of axioms for geometry. In the second edition, however, this statement no longer appeared, following a correction by E. H. Moore (1902) who showed that one of the axioms might be derived from the others. See also (Corry 2003, §3.5; Torretti 1978, 239 ff.).

27 In a series of articles published in the USA over the first decade of the twentieth century under the influence of *Grundlagen der Geometrie*, see (Corry 2003, §3.5), a workable criterion for simplicity of axioms was systematically sought after. For instance, Edward Huntington (1904, p. 290) included simplicity among his requirements for axiomatic systems, yet he warned that "the idea of a simple statement is a very elusive one which has not been satisfactorily defined, much less attained."

notion that is totally foreign to Hilbert's axiomatic approach at this early stage. Rather it is an idea that runs parallel to Hertz's demand for "correctness." Thus, Hilbert demanded from any adequate axiomatization that it should allow for a derivation of all the known theorems of the discipline in question. The axioms formulated in *Grundlagen der Geometrie*, Hilbert claimed, would indeed yield all the known results of Euclidean geometry or of the so-called absolute geometry, namely that valid independently of the parallel postulate, if the corresponding group of axioms is ignored. Thus, reconstructing the very ideas that had given rise to his own conception, Hilbert discussed in great detail the role of each of the groups of axioms in the proofs of two crucial results: the theorem of Desargues and the theorem of Pappus. Unlike independence, however, the completeness of the system of axioms is not a property that Hilbert knew how to verify formally, except to the extent that, starting from the given axioms, he could prove all the theorems he was interested in.

The question of consistency of the various kinds of geometries was an additional concern of Hilbert's analysis, though, perhaps somewhat surprisingly, one that was not even explicitly mentioned in the introduction to *Grundlagen der Geometrie*. He addressed this issue in the *Festschrift* right after introducing all the groups of axioms and after discussing their immediate consequences. Seen from the point of view of Hilbert's later metamathematical research and the developments that followed it, the question of consistency might appear as the most important one undertaken back in 1899; but in the historical context of the evolution of his ideas it certainly was not. In fact, consistency of the axioms is discussed in barely two pages, and it is not immediately obvious why Hilbert addressed it at all. It doesn't seem likely that in 1899 Hilbert would have envisaged the possibility that the body of theorems traditionally associated with Euclidean geometry might contain contradictions. After all, he conceived Euclidean geometry as an empirically motivated discipline, turned into a purely mathematical science after a long, historical process of evolution and depuration. Moreover, and more importantly, Hilbert had presented a model of Euclidean geometry over certain, special types of algebraic number fields. If with the real numbers the issue of continuity might be thought to raise difficulties that called for particular care, in this case Hilbert would have no real reason to call into question the possible consistency of these fields of numbers. Thus, to the extent that Hilbert referred here to the problem of consistency, he seems in fact to be echoing here Hertz's demand for the permissibility of images. As seen above, a main motivation leading Hertz to introduce this requirement was the concern about possible contradictions brought about over time by the gradual addition of ever new hypotheses to a given theory. Although this was not likely to be the case for the well-established discipline of geometry, it might still have happened that the particular way in which the axioms had been formulated in order to account for the theorems of this science would have led to statements that contradict each other. The recent development of non-Euclidean geometries made this possibility only more patent. Thus, Hilbert believed that, although contradictions might in principle possibly occur within his own system, he could also easily show that this was actually not the case.

The relatively minor importance conceded by Hilbert in 1899 to the problem of the consistency of his system of axioms for Euclidean geometry is manifest not only in the fact that he devoted just two pages to it. Of course, Hilbert could not have in mind a direct proof of consistency here, but rather an indirect one, namely, a proof that any contradiction existing in Euclidean geometry must manifest itself in the arithmetic system of real numbers. This would still leave open the question of the consistency of the latter, a problem difficult enough in itself. However, even an indirect proof of this kind does not appear in explicit form in *Grundlagen der Geometrie*. Hilbert only suggested that it would suffice to show that the specific kind of synthetic geometry derivable from his axioms could be translated into the standard Cartesian geometry, taking the axes as representing the whole field of real numbers.[28] More generally stated, in this first edition of *Grundlagen der Geometrie*, Hilbert preferred to bypass a systematic treatment of the questions related to the structure of the system of real numbers. Rather, he contented himself with constructing a model of his system based on a countable, proper sub-field—of whose consistency he may have been confident—and not the whole field of real numbers (Hilbert 1899, 21). It was only in the second edition of *Grundlagen der Geometrie*, published in 1903, that he added an additional axiom, the so-called "axiom of completeness" (*Vollständigkeitsaxiom*), meant to ensure that, although infinitely many incomplete models satisfy all the other axioms, there is only one complete model that satisfies this last axiom as well, namely, the usual Cartesian geometry, obtained when the whole field of real numbers is used in the model (Hilbert 1903a, 22–24). As Hilbert took pains to stress, this axiom cannot be derived from the Archimedean axiom, which was the only one included in the continuity group in the first edition.[29] It is important to notice, however, that the property referred to by this axiom bears no relation whatsoever to Hilbert's general requirement of "completeness" for any system of axioms. Thus his choice of the term "*Vollständigkeit*" in this context seems somewhat unfortunate.

3. THE 1900 LIST OF PROBLEMS

Soon after the publication of *Grundlagen der Geometrie*, Hilbert had a unique opportunity to present his views on mathematics in general and on axiomatics in particular, when he was invited to address the Second International Congress of Mathematicians

---

28 And the same is true for Hilbert's treatment of "completeness" (in his current terminology) at that time.

29 The axiom is formulated in (Hilbert 1903a, 16). Toepell (1986, 254–256) briefly describes the relationship between Hilbert's Vollständigkeitsaxiom and related works of other mathematicians. The axiom underwent several changes throughout the various later editions of the Grundlagen, but it remained central to this part of the argument. Cf. (Peckhaus 1990, 29–35). The role of this particular axiom within Hilbert's axiomatics and its importance for later developments in mathematical logic is discussed in (Moore 1987, 109–122). In 1904 Oswald Veblen introduced the term "categorical" (Veblen 1904, 346) to denote a system to which no irredundant axioms may be added. He believed that Hilbert had checked this property in his own system of axioms. See (Scanlan 1991, 994).

held in Paris in August of 1900. The invitation was a definite sign of the reputation that Hilbert had acquired by then within the international mathematics community. Following a suggestion of Minkowski, Hilbert decided to use the opportunity to provide a glimpse into what, in his view, the new century would bring for mathematics. Thus he posed a list of problems that he considered significant challenges that could lead to fruitful research and to new and illuminating ideas for mathematicians involved in solving them.

In many ways, Hilbert's talk embodied his overall vision of mathematics and science, and he built the list of problems to a large extent according to his own mathematical horizons.[30] Some of the problems belonged to number theory and the theory of invariants, the domains that his published work had placed him in among the leading world experts. Some others belonged to domains with which he was closely acquainted, even though he had not by then published anything of the same level of importance, such as variational calculus. It further included topics that Hilbert simply considered should be given a significant push within contemporary research, such as Cantorian set theory. The list reflected Hilbert's mathematical horizon also in the sense that a very significant portion of the works he cited in reference to the various problems had been published in either of the two main Göttingen mathematical venues: the *Mathematische Annalen* and the Proceedings of the Göttingen Academy of Sciences. And although Hilbert's mathematical horizons were unusually broad, they were nonetheless clearly delimited and thus, naturally, several important, contemporary fields of research were left out of the list.[31] Likewise, important contemporary Italian works on geometry, and the problems related to them, were not referred to at all in the geometrical topics that Hilbert did consider in his list. Moreover, two major contemporary open problems, Fermat's theorem and Poincaré's three-body problem, though mentioned in the introduction, were not counted among the twenty-three problems.

The talk also reflected three other important aspects of Hilbert's scientific personality. Above all is his incurable scientific optimism, embodied in the celebrated and often quoted statement that every mathematical problem can indeed be solved: "There is the problem. Seek its solution. You can find it by pure reason, for in mathematics there is no *ignorabimus*." This was meant primarily as a reaction to a well-known pronouncement of the physiologist Emil du Bois-Reymond (1818–1896) on the inherent limitations of science as a system able to provide us with knowledge about the world.[32] Second, is the centrality of challenging problems in mathematics as a main, necessary condition for the healthy development of any branch of the discipline and, more generally, of that living organism that Hilbert took mathematics to be. And third, is the central role accorded to empirical motivations as a fundamental source of nourishment for that organism, in which mathematics and the physical sci-

30 Several versions of the talk appeared in print and they were all longer and more detailed than the actual talk. Cf. (Grattan-Guinness 2000).

31 Cf. (Gray 2000, 78–88).

ences appear tightly interrelated. But stressing the empirical motivations underlying mathematical ideas should by no means be taken as opposed to rigor. On the contrary, contrasting an "opinion occasionally advocated by eminent men," Hilbert insisted that the contemporary quest for rigor in analysis and arithmetic should in fact be *extended to both geometry and the physical sciences*. He was alluding here, most probably, to Kronecker and Weierstrass, and the Berlin purist tendencies that kept geometry and applications out of their scope of interest. Rigorous methods are often simpler and easier to understand, Hilbert said, and therefore, a more rigorous treatment would only perfect our understanding of these topics, and at the same time would provide mathematics with ever new and fruitful ideas. Explaining why rigor should not be sought only within analysis, Hilbert actually implied that this rigor should actually be pursued in axiomatic terms. He thus wrote:

> Such a one-sided interpretation of the requirement of rigor would soon lead to the ignoring of all concepts arising form geometry, mechanics and physics, to a stoppage of the flow of new material from the outside world, and finally, indeed, as a last consequence, to the rejection of the ideas of the continuum and of irrational numbers. But what an important nerve, vital to mathematical science, would be cut by rooting out geometry and mathematical physics! On the contrary I think that wherever mathematical ideas come up, whether from the side of the theory of knowledge or in geometry, or from the theories of natural or physical science, the problem arises for mathematics to investigate the principles underlying these ideas and to establish them upon a simple and complete system of axioms, so that the exactness of the new ideas and their applicability to deduction shall be in no respect inferior to those of the old arithmetical concepts.[33]

Using rhetoric reminiscent of Paul Volkmann's 1900 book, Hilbert described the development of mathematical ideas as an ongoing, dialectical interplay between the two poles of thought and experience, an interplay that brings to light a "pre-established harmony" between nature and mathematics.[34] The "edifice metaphor" was invoked to help stress the importance of investigating the foundations of mathematics not as an isolated concern, but rather as an organic part of the manifold growth of the discipline in several directions. Hilbert thus said:

> Indeed, the study of the foundations of a science is always particularly attractive, and the testing of these foundations will always be among the foremost problems of the investi-

---

32 See (Du Bois-Reymond 1872). Hilbert would repeat this claim several times later in his career, notably in (Hilbert 1930). Although the basic idea behind the pronouncement was the same on all occasions, and it always reflected his optimistic approach to the capabilities of mathematics, it would nevertheless be important to consider the specific, historical framework in which the pronouncement came and the specific meaning that the situation conveys in one and the same sentence. If in 1900 it came, partly at least, as a reaction to Du Bois-Reymond's sweeping claim about the limitation of science, in 1930 it came after the intense debate against constructivist views about the foundations of arithmetic.

33 The classical locus for the English version of the talk is (Hilbert 1902a). Here I have preferred to quote, where different, from the updated translation appearing in (Gray 2000, 240–282). This passage appears there on p. 245.

34 The issue of the "pre-established harmony" between mathematics and nature was a very central one among Göttingen scientists. This point has been discussed in (Pyenson 1982).

> gator ... [But] a thorough understanding of its special theories is necessary for the successful treatment of the foundations of the science. Only that architect is in the position to lay a sure foundation for a structure who knows its purpose thoroughly and in detail.[35]

Speaking more specifically about the importance of problems for the healthy growth of mathematics, Hilbert characterized an interesting problem as one that is "difficult in order to entice us, yet not completely inaccessible, lest it mock our efforts." But perhaps more important was the criterion he formulated for the solution of one such problem: it must be possible "to establish the correctness of the solution by a finite number of steps based upon a finite number of hypotheses which are implied in the statement of the problem and which must always be exactly formulated."

3.1 Foundational Problems

This is not the place to discuss in detail the list of problems and their historical background and development.[36] Our main concern here is with the sixth problem— Hilbert's call for the axiomatization of physical sciences—and those other problems on the list more directly connected with it. The sixth problem is indeed the last of a well-defined group within the list, to which other "foundational" problems also belong. Beyond this group, the list can be said roughly to contain three other main areas of interest: number theory, algebraic-geometrical problems, and analysis (mainly variational calculus) and its applications in physics.

The first two foundational problems, appearing at the head of Hilbert's list, are Cantor's continuum hypothesis and the compatibility of the axioms of arithmetic. In formulating the second problem on his list, Hilbert stated more explicitly than ever before, that among the tasks related to investigating an axiomatic system, proving its consistency would be the most important one. Eventually this turned into a main motto of his later program for the foundations of arithmetic beginning in the 1920s, but many years and important developments still separated this early declaration, diluted among a long list of other important mathematical tasks for the new century, from an understanding of the actual implications of such an attempt and from an actual implementation of a program to pursue it. In the years to come, as we will see below, Hilbert did many things with axiomatic systems other than attempting a proof of consistency for arithmetic.

Hilbert stated that proving the consistency of geometry could be reduced to proving that of arithmetic, and that the axioms of the latter were those presented by him in "Über den Zahlbegriff" several months prior to this talk. Yet, Hilbert was still confident that this would be a rather straightforward task, easily achievable "by means of a careful study and suitable modification of the known methods of reasoning in the theory of irrational numbers" (Hilbert 1902a, 448).Hilbert did not specify the exact

35 Quoted from (Gray 2000, 258).

36 Cf. (Rowe 1996), and a more detailed, recent, discussion in (Gray 2000).

meaning of this latter statement, but its wording would seem to indicate that in the system of axioms proposed for arithmetic, the difficulty in dealing with consistency would come from the assumption of continuity. Thus the consistency of Euclidean geometry would depend on proving the consistency of arithmetic as defined by Hilbert through his system of axioms. This would, moreover, provide a proof for the very existence of the continuum of real numbers as well. Clearly Hilbert meant his remarks in this regard to serve as an argument against Kronecker's negative reactions to unrestricted use of infinite collections in mathematics, and therefore he explicitly asserted that a consistent system of axioms could prove the existence of higher Cantorian cardinals and ordinals.[37] He thus established a clear connection between the two first problems on his list through the axiomatic approach. Still, Hilbert was evidently unaware of the difficulties involved in realizing this point of view, and, more generally, he most likely had no precise idea of what an elaborate theory of systems of axioms would involve. On reading the first draft of the Paris talk, several weeks earlier, Minkowski understood at once the challenging implications of Hilbert's view, and he hastened to write to his friend:

> In any case, it is highly original to proclaim as a problem for the future, one that mathematicians would think they had already completely possessed for a long time, such as the axioms for arithmetic. What might the many laymen in the auditorium say? Will their respect for us grow? And you will also have a though fight on your hands with the philosophers.[38]

Minkowski turned out to be right to a large extent, and among the ideas that produced the strongest reactions were those related with the status of axioms as implicit definitions, such as Hilbert introduced in formulating the second problem. He thus wrote:

> When we are engaged in investigating the foundations of a science, we must set up a system of axioms which contains an exact and complete description of the relations subsisting between the elementary ideas of the science. The axioms so set up are at the same time the definitions of those elementary ideas, and no statement within the realm of the science whose foundation we are testing is held to be correct unless it can be derived from those axioms by means of a finite number of logical steps. (Hilbert 1902a,447)[39]

The next three problems in the list are directly related with geometry and, although not explicitly formulated in axiomatic terms, they address the question of finding the correct relationship between specific assumptions and specific, significant geometrical facts. Of particular interest for the present account is the fifth. The question of the foundations of geometry had evolved over the last third of the nineteenth century along two parallel paths. First was the age-old tradition of elementary synthetic

---

37 Hilbert also pointed out that no consistent set of axioms could be similarly set up for all cardinals and all alephs. Commenting on this, Ferreirós (1999, 301), has remarked: "This is actually the first published mention of the paradoxes of Cantorian set theory — without making any fuss of it." See also (Peckhaus and Kahle 2002).

38 On 17 July 1900, (Rüdenberg and Zassenhaus 1973, 129).

39 And also quoted in (Gray 2000, 250).

geometry, where the question of foundations more naturally arises in axiomatic terms. A second, alternative, path, that came to be associated with the Helmholtz-Lie problem, had derived directly from the work of Riemann and it had a more physically-grounded orientation connected with the question of spaces that admit the free mobility of rigid bodies. Whereas Helmholtz had only assumed continuity as underlying the motion of rigid bodies, in applying his theory of group of transformations to this problem, Lie was also assuming the differentiability of the functions involved. Hilbert's work on the foundations of geometry, especially in the context that led to *Grundlagen der Geometrie*, had so far been connected with the first of these two approaches, while devoting much less attention to the second one. Now in his fifth problem, he asked whether Lie's conditions, rather than assumed, could actually be deduced from the group concept together with other geometrical axioms.

As a mathematical problem, the fifth one led to interesting, subsequent developments. Not long after his talk, on 18 November 1901, Hilbert himself proved that, in the plane, the answer is positive, and he did so with the help of a then innovative, essentially topological, approach (Hilbert 1902b). That the answer is positive in the general case was satisfactorily proved only in 1952.[40] What concerns us here more directly, however, is that the inclusion of this problem in the list underscores the actual scope of Hilbert's views over the question of the foundations of geometry and over the role of axiomatics. Hilbert suggested here the pursuit of an intricate kind of conceptual clarification involving our assumptions about motion, differentiability and symmetry, such as they appear intimately interrelated in the framework of a well-elaborate mathematical theory, namely, that of Lie. This quest is typical of the spirit of Hilbert's axiomatic involvement with physical theories. At this point, it also clearly suggests that his foundational views on geometry were much broader and open-ended than an exclusive focusing on *Grundlagen der Geometrie*— with a possible overemphasizing of certain, formalist aspects—might seem to imply. In particular, the fifth problem emphasizes, once again and from a different perspective, the prominent role that Hilbert assigned to physicalist considerations in his approach to geometry. In the long run, one can also see this aspect of Hilbert's view resurfacing at the time of his involvement with general theory of relativity. In its more immediate context, however, it makes the passage from geometry to the sixth problem appear as a natural one within the list.

Indeed, if the first two problems in the list show how the ideas deployed in *Grundlagen der Geometrie* led in one direction towards foundational questions in arithmetic, then the fifth problem suggests how they also naturally led, in a different direction, to Hilbert's call for the axiomatization of physical science in the sixth problem. The problem was thus formulated as follows:

> The investigations on the foundations of geometry suggest the problem: To treat in the same manner, by means of axioms, those physical sciences in which mathematics plays

40 This was done, simultaneously, in (Gleason 1952) and (Montgomery and Zippin 1952).

> an important part; in the first rank are the theory of probabilities and mechanics. (Hilbert 1902a, 454)[41]

As examples of what he had in mind Hilbert mentioned several existing and well-known works: the fourth edition of Mach's *Die Mechanik in ihrer Entwicklung*, Hertz's *Principles*, Boltzmann's 1897 *Vorlesungen Über die Principien der Mechanik*, and also Volkmann's 1900 *Einführung in das Studium der theoretischen Physik*. Boltzmann's work offered a good example of what axiomatization would offer, as he had indicated, though only schematically, that limiting processes could be applied, starting from an atomistic model, to obtain the laws of motion of continua. Hilbert thought it convenient to go in the opposite direction also, i.e., to derive the laws of motions of rigid bodies by limiting processes, starting from a system of axioms that describe space as filled with continuous matter in varying conditions. Thus one could investigate the equivalence of different systems of axioms, an investigation that Hilbert considered to be of the highest theoretical importance.

This is one of the few places where Hilbert emphasized Boltzmann's work over Hertz's in this regard, and this may give us the clue to the most immediate trigger that was in the back of Hilbert's mind when he decided to include this problem in the list. Hilbert had met Boltzmann several months earlier in Munich, where he heard his talk on recent developments in physics. Boltzmann had not only discussed ideas connected to the task that Hilbert was now calling for, but he also adopted a rhetoric that Hilbert seems to have found very much to the point. In fact, Boltzmann had suggested that one could follow up the recent history of physics with a look at future developments. Nevertheless, he said, "I will not be so rash as to lift the veil that conceals the future" (Boltzmann 1899, 79). Hilbert, on the contrary, opened the lecture by asking precisely, "who among us would not be glad to lift the veil behind which the future lies hidden" and the whole trust of his talk implied that he, the optimistic Hilbert, was helping the mathematical community to do so.

Together with the well-known works on mechanics referred to above, Hilbert also mentioned a recent work by the Göttingen actuarial mathematician Georg Bohlmann (1869–1928) on the foundations of the calculus of probabilities.[42] The latter was important for physics, Hilbert said, for its application to the method of mean values and to the kinetic theory of gases. Hilbert's inclusion of the theory of probabilities among the main *physical* theories whose axiomatization should be pursued has often puzzled readers of this passage. It is also remarkable that Hilbert did not mention electrodynamics among the physical disciplines to be axiomatized, even though the second half of the Gauss-Weber *Festschrift*, where Hilbert's *Grundlagen der Geometrie* was published, contained a parallel essay by Emil Wiechert (1861–1956) on the foundations of electrodynamics (Wiechert 1899). At any rate, Wiechert's presentation

---

41 Quoted in (Gray 2000, 257).

42 This article reproduced a series of lectures delivered by Bohlmann in a *Ferienkurs* in Göttingen (Bohlmann 1900). In his article Bohlmann referred the readers, for more details, to the chapter he had written for the *Encyklopädie* on insurance mathematics.

was by no means axiomatic, in any sense of the term. On the other hand, the topics addressed by him would start attracting Hilbert's attention over the next years, at least since 1905.

Modelling this research on what had already been done for geometry meant that not only theories considered to be closer to "describing reality" should be investigated, but also other, logically possible ones. The mathematician undertaking the axiomatization of physical theories should obtain a complete survey of all the results derivable from the accepted premises. Moreover, echoing the concern already found in Hertz and later to appear also in Hilbert's letters to Frege, a main task of the axiomatization would be to avoid that recurrent situation in physical research, in which new axioms are added to existing theories without properly checking to what extent the former are compatible with the latter. This proof of compatibility, concluded Hilbert, is important not only in itself, but also because it compels us to search for ever more precise formulations of the axioms.

3.2 A Context for the Sixth Problem

The sixth problem of the list deals with the axiomatization of physics. It was suggested to Hilbert by his own recent research on the foundations of geometry. He thus proposed "to treat in the same manner, by means of axioms, those physical sciences in which mathematics plays an important part." This sixth problem is not really a problem in the strict sense of the word, but rather a general task for whose complete fulfilment Hilbert set no clear criteria. Thus, Hilbert's detailed account in the opening remarks of his talk as to what a meaningful problem in mathematics is, and his stress on the fact that a solution to a problem should be attained in a finite number of steps, does not apply in any sense to the sixth one. On the other hand, the sixth problem has important connections with three other problems on Hilbert's list: the nineteenth ("Are all the solutions of the Lagrangian equations that arise in the context of certain typical variational problems necessarily analytic?"), the twentieth (dealing with the existence of solutions to partial differential equations with given boundary conditions), closely related to the nineteenth and at the same time to Hilbert's long-standing interest in the Dirichlet principle,[43] and, finally, the twenty-third (an appeal to extend and refine the existing methods of variational calculus). Like the sixth problem, the latter two are general tasks rather than specific mathematical problems with a clearly identifiable, possible solution.[44] All these three problems are also strongly connected to physics, though unlike the sixth, they are also part of mainstream, tradi-

43 On 11 October 1899, Hilbert had lectured in Göttingen on the Dirichlet principle, stressing the importance of its application to the theory of surfaces and also to mathematical physics. Cf. *Jahresbericht der Deutschen Mathematiker-Vereinigung* 8 (1900), 22.

44 A similar kind of "general task" problem that Hilbert had perhaps considered adding as the twenty-fourth problem in his list is hinted at in an undated manuscript found in *Nachlass* David Hilbert (Cod. Ms. D. Hilbert, 600). It concerns the definition of criteria for finding simplest proofs in mathematics in general. Cf. a note in (Grattan-Guinness 2001, 167), and a more detailed account in (Thiele 2003).

tional research concerns in mathematics.[45] In fact, their connections to Hilbert's own interests are much more perspicuous and, in this respect, they do not raise the same kind of historical questions that Hilbert's interest in the axiomatization of physics does. Below, I will explain in greater detail how Hilbert conceived the role of variational principles in his program for axiomatizing physics.

Another central issue to be discussed below in some detail is the role the sixth problem played in subsequent developments in mathematics and in physics. At this stage, however, a general point must be stressed about the whole list in this regard. A balanced assessment of the influence of the problems on the development of mathematics throughout the century must take into account not only the intrinsic importance of the problems,[46] but also the privileged institutional role of Göttingen in the mathematical world with the direct and indirect implications of its special status. If Hilbert wished to influence the course of mathematics over the coming century with his list, then his own career was only very partially shaped by it. Part of the topics covered by the list belonged to his previous domains of research, while others belonged to domains where he never became active. On the contrary, domains that he devoted much effort to over the next years, such as the theory of integral equations, were not contemplated in the list. In spite of the enormous influence Hilbert had on his students, the list did not become a necessary point of reference of preferred topics for dissertations. To be sure, some young mathematicians, both in Göttingen and around the world, did address problems on the list and sometimes came up with important mathematical achievements that helped launch their own international careers. But this was far from the only way for talented young mathematicians to reach prominence in or around Göttingen. But, ironically, the sixth problem, although seldom counted among the most influential of the list, will be shown here to count among those that received a greater attention from Hilbert himself and from his collaborators and students over the following years.

For all its differences and similarities with other problems on the list, the important point that emerges from the above account is that the sixth problem was in no sense disconnected from the evolution of Hilbert's early axiomatic conception. Nor was it artificially added in 1900 as an afterthought about the possible extensions of an idea successfully applied in 1899 to the case of geometry. Rather, Hilbert's ideas concerning the axiomatization of physical science arose simultaneously with his increasing enthusiasm for the axiomatic method and they fitted naturally into his overall view of pure mathematics, geometry and physical science—and the relationship among them—by that time. Moreover, as will be seen in the next chapter in some detail, Hilbert's 1905 lectures on axiomatization provide a very clear and comprehensive conception of how the project suggested in the sixth problem should be realized. In fact, it is very likely that this conception was not essentially different from what Hilbert had in mind when formulating his problem in 1900.[47] Interestingly, the devel-

45 For a detailed account of the place of variational principles in Hilbert's work, see (Blum 1994).

46 As treated in (Alexandrov 1979; Browder 1976).

opment of physics from the beginning of the century, and especially *after* 1905, brought many surprises that Hilbert could not have envisaged in 1900 or even when he lectured at Göttingen on the axioms of physics; yet, over the following years Hilbert was indeed able to accommodate these new developments to the larger picture of physics afforded by his program for axiomatization. In fact, some of his later contributions to mathematical physics came by way of realizing the vision embodied in this program, as will be seen in detail in later chapters.

4. FOUNDATIONAL CONCERNS – EMPIRICIST STANDPOINT

Following the publication of *Grundlagen der Geometrie*, Hilbert was occupied for a while with research on the foundations of geometry. Several of his students, such as Max Dehn (1878–1952), Georg Hamel (1877–1954) and Anne Lucy Bosworth (1868–1907), worked in this field as well, including on problems relating to Hilbert's 1900 list. Also many meetings of the *Göttinger Mathematische Gesellschaft* during this time were devoted to discussing related topics. On the other hand, questions relating to the foundations of arithmetic and set theory also received attention in the Hilbert circle. Ernst Zermelo (1871–1953) had already arrived in Göttingen in 1897 in order to complete his *Habilitation*, and his own focus of interest changed soon from mathematical physics to set theory and logic. Around 1899–1900 he had already found an important antinomy in set theory, following an idea of Hilbert's.[48] Later on, in the winter semester of 1900–1901, Zermelo was teaching set theory in Göttingen (Peckhaus 1990, 48–49).

Interest in the foundations of arithmetic became a much more pressing issue in 1903, after Bertrand Russell (1872–1970) published his famous paradox arising from Frege's logical system. Although Hilbert hastened to indicate to Frege that similar arguments had been known in Göttingen for several years,[49] it seems that Russell's publication, coupled with the ensuing reaction by Frege,[50] did have an exceptional impact. Probably this had to do with the high esteem that Hilbert professed towards Frege's command of these topics (which Hilbert may have come to appreciate even more following the sharp criticism recently raised by the latter towards his own ideas). The simplicity of the sets involved in Russell's argument was no doubt a further factor that explains its strong impact on the Göttingen mathematicians. If Hilbert had initially expected that the difficulty in completing the full picture of his approach to the foundations of geometry would lie on dealing with more complex assumptions such as the *Vollständigkeitsaxiom*, now it turned out that the problems perhaps started with the arithmetic itself and even with logic. He soon realized that greater attention

47 Cf. (Hochkirchen 1999), especially chap. 1.
48 See (Peckhaus and Kahle 2002).
49 Hilbert to Frege, 7 November 1903. Quoted in (Gabriel et al. 1980, 51–52).
50 As published in (Frege 1903, 253). See (Ferreirós 1999, 308–311).

should be paid to these topics, and in particular to the possible use of the axiomatic method in establishing the consistency of arithmetic (Peckhaus 1990, 56–57).

Hilbert himself gradually reduced his direct involvement with all questions of this kind, and after 1905 he completely abandoned them for many years to come. Two instances of his involvement with foundational issues during this period deserve some attention here. The first is his address to the Third International Congress of Mathematicians, held in 1904 in Heidelberg. In this talk, later published under the title of "On the Foundations of Logic and Arithmetic," Hilbert presented a program for attacking the problem of consistency as currently conceived. The basic idea was to develop simultaneously the laws of logic and arithmetic, rather than reducing one to the other or to set theory. The starting point was the basic notion of thought-object that would be designated by a sign, which offered the possibility of treating mathematical proofs, in principle, as formulae. This could be seen to constitute an interesting anticipation of what later developed as part of Hilbert's proof theory, but here he only outlined the idea in a very sketchy way. Actually, Hilbert did not go much beyond the mere declaration that this approach would help achieve the desired proof. Hilbert cursorily reviewed several prior approaches to the foundations of arithmetic, only to discard them all. Instead, he declared that the solution for this problem would finally be found in the correct application of the axiomatic method (Hilbert 1905c, 131).

Upon returning to Göttingen from Heidelberg, Hilbert devoted some time to working out the ideas outlined at the International Congress of Mathematicians. The next time he presented them was in an introductory course devoted to "The Logical Principles of Mathematical Thinking," which contains the second instance of Hilbert's involvement with the foundation of arithmetic in this period. This course is extremely important for my account here because it contains the first detailed attempt to implement the program for the axiomatization of physics.[51] I will examine it in some detail below. At this point I just want to briefly describe the other parts of the course, containing some further foundational ideas for logic and arithmetic, and some further thoughts on the axiomatization of geometry.

Hilbert discussed in this course the "logical foundations" of mathematics by introducing a formalized calculus for propositional logic. This was a rather rudimentary calculus, which did not even account for quantifiers. As a strategy for proving consistency of axiomatic systems, it could only be applied to very elementary cases.[52] Prior to defining this calculus Hilbert gave an overview of the basic principles of the axiomatic method, including a more detailed account of its application to arithmetic, geometry and the natural sciences. What needs to be stressed concerning this text is that, in spite of his having devoted increased attention over the previous years to foundational questions in arithmetic, Hilbert's fundamentally empiricist

51 There are two extant sets of notes for this course: (Hilbert 1905a and 1905b). Quotations below are taken from (Hilbert 1905a). As these important manuscripts remain unpublished, I transcribe in the footnotes some relevant passages at length. Texts are underlined or crossed-out as in the original. Later additions by Hilbert appear between < > signs.

52 For a discussion of this part of the course, see (Peckhaus 1990, 61–75).

approach to issues in the foundations of geometry was by no means weakened, but rather the opposite. In fact, in his 1905 course, Hilbert actually discussed the role of an axiomatic analysis of the foundations of arithmetic in similar, empiricist terms.

Once again, Hilbert contrasted the axiomatic method with the genetic approach in mathematics, this time making explicit reference to the contributions of Kronecker and Weierstrass to the theory of functions. Yet Hilbert clearly separated the purely logical aspects of the application of the axiomatic method from the "genetic" origin of the axioms themselves: the latter is firmly grounded on empirical experience. Thus, Hilbert asserted, it is not the case that the system of numbers is given to us through the network of concepts (*Fachwerk von Begriffen*) involved in the eighteen axioms. On the contrary, it is our direct intuition of the concept of natural number and of its successive extensions, well known to us by means of the genetic method, which has guided our construction of the axioms:

> The aim of every science is, first of all, to set up a network of concepts based on axioms to whose very conception we are naturally led by *intuition and experience*. Ideally, all the phenomena of the given domain will indeed appear as part of the network and all the theorems that can be derived from the axioms will find their expression there.[53]

What this means for the axiomatization of geometry, then, is that its starting point must be given by the intuitive facts of that discipline,[54] and that the latter must be in agreement with the network of concepts created by means of the axiomatic system. The concepts involved in the network itself, Hilbert nevertheless stressed, are totally detached from experience and intuition.[55] This procedure is rather obvious in the case of arithmetic, and to a certain extent the genetic method has attained similar results for this discipline. In the case of geometry, although the need to apply the pro-

---

53 "*Uns war das Zahlensystem schließlich nichts* als ein *Fachwerk von Begriffen*, das durch 18 Axiome definiert war. Bei der Aufstellung dieser leitete uns allerdings die Anschauung, die wir von dem Begriff der Anzahl und seiner genetischen Ausdehnung haben. ... So ist in jeder Wissenschaft die Aufgabe, in den Axiomen zunächst ein Fachwerk von Begriffen zu errichten, bei dessen Aufstellung wir uns natürlich durch die *Anschauung und Erfahrung* leiten lassen; das Ideal ist dann, daß in diesem Fachwerk alle Erscheinungen des betr. Gebietes Platz finden, und daß jeder aus den Axiomen folgende Satz dabei Verwertung findet.
Wollen wir nun für die Geometrie ein Axiomensystem aufstellen, so heißt das, daß wir uns den Anlaß dazu durch die anschaulichen Thatsachen der Geometrie geben lassen, und diesen das aufzurichtende Fachwerk entsprechen lassen; die Begriffe, die wir so erhalten, sind *aber als gänzlich losgelöst von jeder Erfahrung und Anschauung zu betrachten*. Bei der Arithmetik ist diese Forderung verhältnismäßig naheliegend, sie wird in gewissem Umfange auch schon bei der genetischen Methode angestrebt. Bei der Geometrie jedoch wurde die Notwendigkeit dieses Vorgehens viel später erkannt; dann aber wurde eine axiomatische Behandlung eher versucht, als in der Arithmetik, wo noch immer die genetische Betrachtung herrschte. Doch ist die Aufstellung eines vollständigen Axiomensystemes ziemlich schwierig, noch viel schwerer wird sie in der Mechanik, Physik etc. sein, wo das Material an Erscheinungen noch viel größer ist." (Hilbert 1905a, 36–37)

54 "... den Anlaß dazu durch die anschaulischen Thatsachen der Geometrie geben lassen..." (Hilbert 1905a, 37)

55 "... die Begriffe, die wir so erhalten, sind aber als gänzlich losgelöst von jeder Erfahrung und Anschauung zu betrachten." (Hilbert 1905a, 37)

cess truly systematically was recognized much later, the axiomatic presentation has traditionally been the accepted one. And if setting up a full axiomatic system has proven to be a truly difficult task for geometry, then, Hilbert concluded, it will be much more difficult in the case of mechanics or physics, where the range of observed phenomena is even broader.[56]

Hilbert's axioms for geometry in 1905 were based on the system of *Grundlagen der Geometrie*, including all the corrections and additions introduced to it since 1900. Here too he started by choosing three basic kinds of undefined elements: points, lines and planes. This choice, he said, is somewhat "arbitrary" and it is dictated by consideration of simplicity. But the arbitrariness to which Hilbert referred here has little to do with the arbitrary choice of axioms sometimes associated with twentieth-century formalistic conceptions of mathematics; it is not an absolute arbitrariness constrained only by the requirement of consistency. On the contrary, it is limited by the need to remain close to the "intuitive facts of geometry." Thus, Hilbert said, instead of the three chosen, basic kinds of elements, one could likewise start with [no... not with "chairs, tables, and beer-mugs," but rather with] circles and spheres, and formulate the adequate axioms that are still in agreement with the usual, intuitive geometry.[57]

Hilbert plainly declared that Euclidean geometry—as defined by his systems of axioms—is the one and only geometry that fits our spatial experience,[58] though in his opinion, it would not be the role of mathematics or logic to explain why this is so. But if that is the case, then what is the status of the non-Euclidean or non-Archimedean geometries? Is it proper at all to use the term "geometry" in relation to them? Hilbert thought it unnecessary to break with accepted usage and restrict the meaning of the term to cover only the first type. It has been unproblematic, he argued, to extend the meaning of the term "number" to include also the complex numbers, although the latter certainly do not satisfy all the axioms of arithmetic. Moreover, it would be untenable from the logical point of view to apply the restriction: although it is not highly probable, it may nevertheless be the case that some changes would still be introduced in the future to the system of axioms that describes intuitive geometry. In fact, Hilbert knew very well that this "improbable" situation had repeatedly arisen in relation to the original system he had put forward in 1900 in *Grundlagen der Geometrie*. To conclude, he compared once again the respective situations in geometry and in physics: in the theory of electricity, for instance, new theories are continually formulated that transform many of the basic facts of the discipline, but no one thinks that the name of the discipline needs to be changed accordingly.

---

56 "... das Material an Erscheinungen noch viel größer ist." (Hilbert 1905a, 37)

57 "Daß wir gerade diese zu Elementardingen des begrifflichen Fachwerkes nehmen, ist willkürlich und geschieht nur wegen ihrer augenscheinlichen Einfachheit; im Princip könnte man die ersten Dinge auch Kreise und Kugeln nennen, und die Festsetzungen über sie so treffen, daß sie diesen Dingen der anschaulichen Geometrie entsprechen." (Hilbert 1905a, 39)

58 "Die Frage, wieso man in der Natur nur gerade die durch alle diese Axiome festgelegte Euklidische Geometrie braucht, bezw. warum unsere Erfahrung gerade in dieses Axiomsystem sich einfügt, gehört nicht in unsere mathematisch-logichen Untersuchungen." (Hilbert 1905a, 67)

Hilbert also referred explicitly to the status of those theories that, like non-Euclidean and non-Archimedean geometries, are created arbitrarily through the purely logical procedure of setting down a system of independent and consistent axioms. These theories, he said, can be applied to any objects that satisfy the axioms. For instance, non-Euclidean geometries are useful to describe the paths of light in the atmosphere under the influence of varying densities and diffraction coefficients. If we assume that the speed of light is proportional to the vertical distance from a horizontal plane, then one obtains light-paths that are circles orthogonal to the planes, and light-times equal to the non-Euclidean distance from them.[59] Thus, the most advantageous way to study the relations prevailing in this situation is to apply the conceptual schemes provided by non-Euclidean geometry.[60]

A further point of interest in Hilbert's discussion of the axioms of geometry in 1905 concerns his remarks about what he called the philosophical implications of the use of the axiomatic method. These implications only reinforced Hilbert's empiricist view of geometry. Geometry, Hilbert said, arises from reality through intuition and observation, but it works with idealizations: for instance, it considers very small bodies as points. The axioms in the first three groups of his system are meant to express idealizations of a series of facts that are easily recognizable as independent from one other; the assertion that a straight line is determined by two points, for instance, never gave rise to the question of whether or not it follows from other, basic axioms of geometry. But establishing the status of the assertion that the sum of the angles in a triangle equals two right angles requires a more elaborate axiomatic analysis. This analysis shows that such an assertion is a separate piece of knowledge, which—we now know for certain—cannot be deduced from earlier facts (or from their idealizations, as embodied in the three first groups of axioms). This knowledge can only be gathered from new, independent empirical observation. This was Gauss's aim, according to Hilbert, when he confirmed the theorem for the first time, by measuring the angles of the large triangle formed by the three mountain peaks.[61] The network of concepts that constitute geometry, Hilbert concluded, has been proved consistent, and therefore it exists mathematically, independently of any observation. Whether or not

---

59 As in many other places in his lectures, Hilbert gave no direct reference to the specific physical theory he had in mind here, and in this particular case I have not been able to find it.

60 "Ich schließe hier noch die Bemerkung an, daß man jedes solches Begriffschema, das wir so rein logisch aus irgend welchen Axiomen aufbauen, anwenden kann auf beliebige gegenständliche Dinge, wenn sie nur diesen Axiomen genügen. ... Ein solches Beispiel für die Anwendung des Begriffschemas der nichteuklidischen Geometrie bildet das System der Lichtwege in unserer Atmosphäre unter dem Einfluß deren variabler Dichte und Brechungsexponenten; machen wir nämlich die einfachste mögliche Annahme, daß die Lichtgeschwindigkeit proportional ist dem vertikalen Abstande y von einer Horizontalebene, so ergeben sich als Lichtwege gerade die Orthogonalkreise jener Ebene, als Lichtzeit gerade die nichteuklidiche Entfernung auf ihnen. Um die hier obwaltenden Verhältnisse also genauer zu untersuchen, können wir gerade mit Vorteil das Begriffschema der nichteuklidischen Geometrie anwenden." (Hilbert 1905a, 69–70)

61 "In diesem Sinne und zu diesem Zwecke hat zuerst Gauß durch Messung an großen Dreiecken den Satz bestätigt." (Hilbert 1905a, 98)

it corresponds to reality is a question that can be decided only by observation, and our analysis of the independence of the axioms allows determining very precisely the minimal set of observations needed in order to do so.[62] Later on, he added, the same kind of perspective must be adopted concerning physical theories, although there its application will turn out to be much more difficult than in geometry.

In concluding his treatment of geometry, and before proceeding to discuss the specific axiomatization of individual physical theories, Hilbert summarized the role of the axiomatic method in a passage which encapsulates his view of science and of mathematics as living organisms whose development involves both an expansion in scope and an ongoing clarification of the logical structure of their existing parts.[63] The axiomatic treatment of a discipline concerns the latter; it is an important part of this growth but—Hilbert emphasized—only one part of it. The passage, reads as follows:

> The edifice of science is not raised like a dwelling, in which the foundations are first firmly laid and only then one proceeds to construct and to enlarge the rooms. Science prefers to secure as soon as possible comfortable spaces to wander around and only subsequently, when signs appear here and there that the loose foundations are not able to sustain the expansion of the rooms, it sets about supporting and fortifying them. This is not a weakness, but rather the right and healthy path of development.[64]

This metaphor provides the ideal background for understanding what Hilbert went on to realize at this point in his lectures, namely, to present his first detailed account of how the general idea of axiomatization of physical theories would be actually implemented in each case. But before we can really discuss that detailed account, it is necessary to broaden its context by briefly describing some relevant developments in physics just before 1905, and how they were manifest in Göttingen.

5. HILBERT AND PHYSICS IN GÖTTINGEN CIRCA 1905

The previous section described Hilbert's foundational activities in mathematics between 1900 and 1905. Those activities constituted the natural outgrowth of the seeds planted in *Grundlagen der Geometrie* and the developments that immediately

---

62 "Das Begriffsfachwerk der Geometrie selbst ist nach Erweisung seiner Widerspruchslosigkeit natürlich auch unabhängig von jeder Beobachtung mathematisch existent; der Nachweis seiner Übereinstimmung mit der Wirklichkeit kann nur durch Beobachtungen geführt werden, und die kleinste notwendige solche wird durch die Unabhängigkeitsuntersuchungen gegeben." (Hilbert 1905a, 98)

63 Elsewhere Hilbert called these two aspects of mathematics the "progressive" and "regressive" functions of mathematics, respectively (both terms not intended as value judgments, of course). See (Hilbert 1992, 17–18).

64 "Das Gebäude der Wissenschaft wird nicht aufgerichtet wie ein Wohnhaus, wo zuerst die Grundmauern fest fundiert werden und man dann erst zum Auf- und Ausbau der Wohnräume schreitet; die Wissenschaft zieht es vor, sich möglichst schnell wohnliche Räume zu verschaffen, in denen sie schalten kann, und erst nachträglich, wenn es sich zeigt, dass hier und da die locker gefügten Fundamente den Ausbau der Wohnräume nicht zu tragen vermögen, geht sie daran, dieselben zu stützen und zu befestigen. Das ist kein Mangel, sondern die richtige und gesunde Entwicklung." (Hilbert 1905a, 102.) Other places where Hilbert uses a similar metaphor are (Hilbert 1897, 67; Hilbert 1918, 148).

followed it. My account is not meant to imply, however, that Hilbert's focus of interest was limited to, or even particularly focused around, this kind of enquiry during those years. On 18 September 1901, for instance, Hilbert's keynote address at the commemoration of the 150th anniversary of the Göttingen Scientific Society (*Gesellschaft der Wissenschaften zu Göttingen*) was devoted to analyzing the conditions of validity of the Dirichlet principle (Hilbert 1904, 1905d). Although thus far he had published very little in this field, Hilbert's best efforts from then on would be given to analysis, and in particular, the theory of linear integral equations. His first publication in this field appeared in 1902, and others followed, up until 1912. But at the same time, he sustained his interest in physics, which is directly connected with analysis and related fields to begin with, and this interest in physics became only more diverse throughout this period. His increased interest in analysis had a natural affinity with the courses on potential theory (winter semester, 1901–1902; summer semester, 1902) and on continuum mechanics (winter semester, 1902–1903; summer semester, 1903) that he taught at that time. Perhaps worthy of greater attention, however, is Hilbert's systematic involvement around 1905 with the theories of the electron, on which I will elaborate in the present section.

Still, a brief remark on Hilbert's courses on continuum mechanics: The lecture notes of these two semesters (Hilbert 1902–1903, 1903b) are remarkable for the thoroughness with which the subject was surveyed. The presentation was probably the most systematic and detailed among all physical topics taught by Hilbert so far, and it comprised detailed examinations of the various existing approaches (particularly those of Lagrange, Euler and Helmholtz). Back in 1898–1899, in the final part of a course on mechanics, Hilbert had briefly dealt with the mechanics of systems of an infinite number of mass-points while stressing that the detailed analysis of such systems would actually belong to a different part of physics. This was precisely the subject he would consider in 1902. In that earlier course Hilbert had also discussed some variational principles of mechanics, without however presenting the theory in anything like a truly axiomatic perspective. Soon thereafter, in 1900 in Paris, Hilbert publicly presented his call for the axiomatization of physics. But in 1902–1903, in spite of the high level of detail with which he systematically discussed the physical discipline of continuum mechanics, the axiomatic presentation was not yet the guiding principle. Hilbert did state that a main task to be pursued was the axiomatic description of physical theories[65] and throughout the text he specifically accorded the status of axioms to some central statements.[66] Still, the notes convey the distinct impression that Hilbert did not believe that the time was ripe for the fully axiomatic

65 The manuscript shows an interesting hesitation on how this claim was stated: "Das <Als ein wichtiges> Ziel der Vorlesung ist <denke ich mir> die mathematische Beschreibung der Axiome der Physik. Vergl. Archiv der Mathematik und Physik, meine Rede: 'Probleme der Mathematik'." However, it is not clear if this amendment of the text reflects a hesitation on the side of Hilbert, or on the side of Berkowski, who wrote down the notes. (Hilbert 1902–1903, 2)

66 Thus for instance in (Hilbert 1902–1903).

treatment of mechanics, or at least of continuum mechanics, in axiomatic terms similar to those previously deployed in full for geometry.

On the other hand, it is worth stressing that in many places Hilbert set out to develop a possible unified conception of mechanics, thermodynamics (Hilbert 1903b, 47–91) and electrodynamics (Hilbert 1903b, 91–164) by using formal analogies with the underlying ideas of his presentation of the mechanics of continua. These ideas, which were treated in greater detail from an axiomatic point of view in the 1905 lectures, are described more fully below; therefore, at this point I will not give a complete account of them. Suffice it to say that Hilbert considered the material in these courses to be original and important, and not merely a simple repetition of existing presentations. In fact, the only two talks he delivered in 1903 at the meetings of the *Göttinger Mathematische Gesellschaft* were dedicated to reporting on their contents.[67]

Still in 1903, Hilbert gave a joint seminar with Minkowski on stability theory.[68] He also presented a lecture on the same topic at the yearly meeting of the *Gesellschaft Deutscher Naturforscher und Ärzte* at Kassel,[69] sparking a lively discussion with Boltzmann.[70] In the winter semester of 1904–1905 Hilbert taught an exercise course on mechanics and later gave a seminar on the same topic. The course "Logical Principles of Mathematical Thinking," containing the lectures on axiomatization of physics, was taught in the summer semester of 1905. He then lectured again on mechanics (winter semester, 1905–1906) and two additional semesters on continuum mechanics.

The renewed encounter with Minkowski signified a major source of intellectual stimulation for these two old friends, and it particularly offered a noteworthy impulse to the expansion of Hilbert's horizon in physics. As usual, their walks were an opportunity to discuss a wide variety of mathematical topics, but now physics became a more prominent, common interest than it had been in the past. Teaching in Zürich since 1894, Minkowski had kept alive his interest in mathematical physics, and in particular in analytical mechanics and thermodynamics (Rüdenberg and Zassenhaus 1973, 110–114). Now at Göttingen, he further developed these interests. In 1906 Minkowski published an article on capillarity (Minkowski 1906), commissioned for

---

67 See the announcements in *Jahresbericht der Deutschen Mathematiker-Vereinigung* 12 (1903), 226 and 445. Earlier volumes of the *Jahresbericht der Deutschen Mathematiker-Vereinigung* do not contain announcements of the activities of the *Göttinger Mathematische Gesellschaft*, and therefore it is not known whether he also discussed his earlier courses there.

68 *Nachlass* David Hilbert, (Cod. Ms. D. Hilbert, 570/1) contains a somewhat random collection of handwritten notes related to many different courses and seminars of Hilbert. Notes of this seminar on stability theory appear on pp. 18–24. Additional, related notes appear in (Cod. Ms. D. Hilbert, 696).

69 *Nachlass* David Hilbert, (Cod. Ms. D. Hilbert, 593) contains what appear to be the handwritten notes of this talk, with the title "Vortrag über Stabilität einer Flüssigkeit in einem Gefässe," and includes some related bibliography.

70 As reported in *Naturwissenschaftliche Rundschau*, vol. 18, (1903), 553–556 (cf. Schirrmacher 2003, 318, note 63). The reporter of this meeting, however, considered that Hilbert was addressing a subtlety, rather than a truly important physical problem.

the physics volume of the *Encyklopädie*, edited by Sommerfeld. At several meetings of the *Göttinger Mathematische Gesellschaft*, Minkowski lectured on this as well as other physical issues, such as Euler's equations of hydrodynamics and recent work on thermodynamics by Walter Nernst (1864–1941), (Nernst 1906), who by that time had already left Göttingen. Minkowski also taught advanced seminars on physical topics and more basic courses on mechanics, continuum mechanics, and exercises on mechanics and heat radiation.[71] In 1905 Hilbert and Minkowski organized, together with other Göttingen professors, an advanced seminar that studied recent progress in the theories of the electron.[72] In December 1906, Minkowski reported to the *Göttinger Mathematische Gesellschaft* on recent developments in radiation theory, and discussed the works of Hendrik Antoon Lorentz (1853–1928), Max Planck (1858–1947), Wilhelm Wien (1864–1928) and Lord Rayleigh (1842–1919), (Minkowski 1907, 78). Yet again in 1907, the two conducted a joint seminar on the equations of electrodynamics, and that semester Minkowski taught a course on heat radiation, after having studied with Hilbert Planck's recent book on this topic (Planck 1906).[73] Finally, as it is well known, during the last years of his life, 1907 to 1909, Minkowski's efforts were intensively dedicated to electrodynamics and the principle of relativity.

The 1905 electron theory seminar exemplifies the kind of unique scientific event that could be staged only at Göttingen at that time, in which leading mathematicians and physicists would meet on a weekly basis in order to intensively discuss current open issues of the discipline. In fact, over the preceding few years the *Göttinger Mathematische Gesellschaft* had dedicated many of its regular meetings to discussing recent works on electron theory and related topics, so that this seminar was a natural continuation of a more sustained, general interest for the local scientific community.

---

71 Cf. *Jahresbericht der Deutschen Mathematikervereinigung* 13 (1904), 492; 16 (1907), 171; 17 (1908), 116. See also the *Vorlesungsverzeichnisse*, Universität Göttingen, winter semester, 1903–1904, 14; summer semester, 1904, 14–16. A relatively large collection of documents and manuscripts from Minkowski's *Nachlass* has recently been made available at the Jewish National Library, at the Hebrew University, Jerusalem. These documents are yet to be thoroughly studied and analyzed. They contain scattered notes of courses taught at Königsberg, Bonn, Zurich and Göttingen. The notes of a Göttingen course on mechanics, winter semester, 1903–1904, are found in Box IX (folder 4) of that collection. One noteworthy aspect of these notes is that Minkowski's recommended reading list is very similar to that of Hilbert's earlier courses and comprises mainly texts then available at the *Lesezimmer*. It included classics such as Lagrange, Kirchhoff, Helmholtz, Mach, and Thomson-Tait, together with more recent, standard items such as the textbooks by Voigt, Appell, Petersen, Budde and Routh. Like Hilbert's list it also included the lesser known (Rausenberg 1888), but it also comprised two items absent from Hilbert's list: (Duhamel 1853–1854) and (Föppl 1901). Further, it recommended Voss's *Encyklopädie* article as a good summary of the field.

72 Pyenson (1979) contains a detailed and painstaking reconstruction of the ideas discussed in this seminar and the contributions of its participants. This reconstruction is based mainly on *Nachlass* David Hilbert, (Cod. Ms. D. Hilbert, 570/9). I strongly relied on this article as a starting point for my account of the seminar in the next several paragraphs. Still, my account departs from Pyenson's views in some respects.

73 The notes of the course appear in (Minkowski 1907).

Besides Minkowski and Hilbert, the seminar was led by Wiechert and Gustav Herglotz (1881–1953). Herglotz had recently joined the Göttingen faculty and received his Habilitation for mathematics and astronomy in 1904. Alongside Wiechert, he contributed important new ideas to the electron theory and the two would later become the leading geophysicists of their time. The list of students who attended the seminar includes, in retrospect, no less impressive names: two future Nobel laureates, Max von Laue (1879–1960) and Max Born (1882–1970), as well as Paul Heinrich Blasius (1883–1970) who would later distinguish himself in fluid mechanics, and Arnold Kohlschütter (1883–1969), a student of Schwarzschild who became a leading astronomer himself. Parallel to this seminar, a second one on electrotechnology was held with the participation of Felix Klein, Carl Runge (1856–1914), Ludwig Prandtl (1875–1953) and Hermann Theodor Simon (1870–1918), then head of the Göttingen Institute for Applied Electricity.[74]

The modern theory of the electron had developed through the 1890s, primarily with the contributions of Lorentz working in Leiden, but also through the efforts of Wiechert at Göttingen and—following a somewhat different outlook—of Joseph Larmor (1857–1942) at Cambridge.[75] Lorentz had attempted to account for the interaction between aether and matter in terms of rigid, negatively charged, particles: the electrons. His article of 1895 dealing with concepts such as stationary aether and local time, while postulating the existence of electrons, became especially influential (Lorentz 1895). The views embodied in Lorentz's and Larmor's theories received further impetus from contemporary experimental work, such as that of Pieter Zeeman (1865–1943) on the effect associated with his name, work by J. J. Thomson (1856–1940) especially concerning the cathode ray phenomena and their interpretation in terms of particles, and also work by Wiechert himself, Wien and Walter Kaufmann (1871–1947). Gradually, the particles postulated by the theories and the particle-laden explanations stemming from the experiments came to be identified with one another.[76]

Lorentz's theory comprised elements from both Newtonian mechanics and Maxwell's electrodynamics. While the properties of matter are governed by Newton's laws, Maxwell's equations describe the electric and magnetic fields, conceived as states of the stationary aether. The electron postulated by the theory provided the connecting link between matter and aether. Electrons moving in the aether generate electric and magnetic fields, and the latter exert forces on material bodies through the electrons themselves. The fact that Newton's laws are invariant under Galilean transformations and Maxwell's are invariant under what came to be known as Lorentz transformations was from the outset a source of potential problems and difficulties for the theory, and in a sense, these and other difficulties would be dispelled only with the formulation of Einstein's special theory of relativity in 1905. In Lorentz's theory

74 Cf. (Pyenson 1979, 102).
75 Cf. (Warwick 1991).
76 Cf. (Arabatzis 1996).

the conflict with experimental evidence led to the introduction of the famous contraction hypothesis and in fact, of a deformable electron.[77] But in addition it turned out that, in this theory, some of the laws governing the behavior of matter would be Lorentz invariant, rather than Galilean, invariant. The question thus arose whether this formal, common underlying property does not actually indicate a more essential affinity between what seemed to be separate realms, and, in fact, whether it would not be possible to reduce all physical phenomena to electrodynamics.[78]

Initially, Lorentz himself attempted to expand the scope of his theory, as a possible foundational perspective for the whole of physics, and in particular as a way to explain molecular forces in terms of electrical ones. He very soon foresaw a major difficulty in subsuming also gravitation within this explanatory scope. Still, he believed that such a difficulty could be overcome, and in 1900 he actually published a possible account of gravitation in terms of his theory. The main difficulty in this explanation was that, according to existing astronomical data, the velocity of gravitational effects would seem to have to expand much faster than electromagnetic ones, contrary to the requirements of the theory (Lorentz 1900). This and other related difficulties are in the background of Lorentz's gradual abandonment of a more committed foundational stance in connection with electron theory and the electromagnetic worldview. But the approach he had suggested in order to address gravitational phenomena in electromagnetic terms was taken over and further developed that same year by Wilhelm Wien, who had a much wider aim. Wien explicitly declared that his goal was to unify currently "isolated areas of mechanical and electromagnetic phenomena," and in fact, to do so in terms of the theory of the electron while assuming that all mass was electromagnetic in nature, and that Newton's laws of mechanics should be reinterpreted in electromagnetic terms.[79]

One particular event that highlighted the centrality of the study of the connection and interaction between aether and matter in motion among physicists in the German-speaking world was the 1898 meeting of the *Gesellschaft Deutscher Naturforscher und Ärzte*, held at Düsseldorf jointly with the annual meeting of the *Deutsche Mathematiker- Vereinigung*. Most likely both Hilbert and Minkowski had the opportunity to attend Lorentz's talk, which was the focus of interest. Lorentz described the main problem facing current research in electrodynamics in the following terms:

> Ether, ponderable matter, and, we may add, electricity are the building stones from which we compose the material world, and if we could know whether matter, when it moves, carries the ether with it or not, then the way would be opened before us by which

77 In Larmor's theory the situation was slightly different, and so were the theoretical reasons for adopting the contraction hypothesis, due also to George FitzGerald (1851–1901). For details, see (Warwick 2003, 367–376).

78 For a more detailed explanation, cf. (Janssen 2002).

79 See (Wien 1900). This is the article to which Voss referred in his survey of 1901, and that he took to be representative of the new foundationalist trends in physics. Cf. (Jungnickel and McCormmach 1986, 2: 236–240).

> we could further penetrate into the nature of these building stones and their mutual relations. (Lorentz 1898, 101)[80]

This formulation was to surface again in Hilbert's and Minkowski's lectures and seminars on electrodynamics after 1905.

The theory of the electron itself was significantly developed in Göttingen after 1900, with contributions to both its experimental and theoretical aspects. The experimental side came up in the work of Walter Kaufmann, who had arrived from Berlin in 1899. Kaufmann experimented with Becquerel rays, which produced high-speed electrons. Lorentz's theory stipulated a dependence of the mass of the electron on its velocity v, in terms of a second order relation on v/c (c being, of course, the speed of light). In order to confirm this relation it was necessary to observe electrons moving at speeds as close as possible to c, and this was precisely what Kaufmann's experiments could afford, by measuring the deflection of electrons in electric and magnetic fields. He was confident of the possibility of a purely electromagnetic physics, including the solution of open issues such as the apparent character of mass, and the gravitation theory of the electron. In 1902 he claimed that his results, combined with the recent developments of the theory, had definitely confirmed that the mass of the electrons is of "purely electromagnetic nature."[81]

The recent developments of the theory referred to by Kaufmann were those of his colleague at Göttingen, the brilliant *Privatdozent* Max Abraham (1875–1922). In a series of publications, Abraham introduced concepts such as "transverse inertia," and "longitudinal mass," on the basis of which he explained where the dynamics of the electron differed from that of macroscopic bodies. He also developed the idea of a rigid electron, as opposed to Lorentz's deformable one. He argued that explaining the deformation of the electron as required in Lorentz's theory would imply the need to introduce inner forces of non-electromagnetic origin, thus contradicting the most fundamental idea of a purely electromagnetic worldview. In Abraham's theory, the kinematic equations of a rigid body become fundamental, and he introduced variational principles to derive them. Thus, for instance, using a Lagrangian equal to the difference between the magnetic and the electrical energy, Abraham described the translational motion of the electron and showed that the principle of least action also holds for what he called "quasi-stationary" translational motion (namely, motion in which the velocity of the electron undergoes a small variation over the time required for light to traverse its diameter). Abraham attributed special epistemological significance to the fact that the dynamics of the electron could be expressed by means of a Lagrangian (Abraham 1903, 168),[82] a point that will surface interestingly in Hilbert's 1905 lectures on axiomatization, as we will see in the next section. Beyond the technical level, Abraham was a staunch promoter of the electromagnetic worldview and his theory of the electron was explicitly conceived to "shake the foundations of the

---

80 Translation quoted from (Hirosige 1976, 35).
81 Cf. (Hon 1995; Miller 1997, 44–51, 57–62).
82 On Abraham's electron theory, see (Goldberg 1970; Miller 1997, 51–57).

mechanical view of nature." Still, in 1905 he conceded that "the electromagnetic world picture is so far only a program."[83]

Among the organizers of the 1905 electron theory seminar, it was Wiechert who had been more directly involved in research of closely related issues. Early in his career he became fascinated by the unification of optics and electromagnetism offered by Maxwell's theory, and was convinced of the centrality of the aether for explaining all physical phenomena. In the 1890s, still unaware of Lorentz's work, he published the outlines of his own theory of "atoms of electricity," a theory which he judged to be still rather hypothetical, however. This work contained interesting theoretical and experimental aspects that supported his view that cathode ray particles were indeed the electric atoms of his theory. After his arrival in Göttingen in 1897, Wiechert learnt about Lorentz's theory, and quickly acknowledged the latter's priority in developing an electrodynamics based on the concept of the "electron," the term that he now also adopted. Like Lorentz, Wiechert also adopted a less committed and more skeptical approach towards the possibility of a purely electromagnetic foundation of physics.[84] Obviously Hilbert was in close, continued contact with Wiechert and his ideas, but one rather remarkable opportunity to inspect these ideas more closely came up once again in 1899, when Wiechert published an article on the foundations of electrodynamics as the second half of the *Gauss-Weber Festschrift* (Wiechert 1899).

Not surprisingly, Abraham's works on electron theory were accorded particular attention by his Göttingen colleagues in the 1905 seminar, yet Abraham himself seems not to have attended the meetings in person. He was infamous for his extremely antagonistic and aggressive personality,[85] and this background may partly explain his absence. But one wonders if also his insistence on the foundational implications of electron theory, and a completely different attitude of the seminar leaders to this question may provide an additional, partial explanation for this odd situation. I already mentioned Wiechert's basic skepticism, or at least caution, in this regard. As we will see, also Hilbert and Minkowski were far from wholeheartedly supporting a purely electromagnetic worldview. Kaufmann was closest to Abraham in this point, and he had anyway left Göttingen in 1903. It is interesting to notice, at any rate, that Göttingen physicists and mathematicians held different, and very often conflicting, views on this as well as other basic issues, and it would be misleading to speak of a "Göttingen approach" to any specific topic. The situation around the electron theory seminar sheds interesting light on this fact.

Be that as it may, the organizers relied not on Abraham's, but on other, different works as the seminar's main texts. The texts included, in the first place, Lorentz's 1895 presentation of the theory, and also his more recently published *Encyklopädie*

---

83 Quoted in (Jungnickel and McCormmach 1986, 2: 241). For a recent summary account of the electromagnetic worldview and the fate of its program, see (Kragh 1999, 105–199).

84 Cf. (Darrigol 2000, 344–347).

85 Cf., e.g., (Born 1978, 91 and 134–137).

article (Lorentz 1904a), which was to become the standard reference in the field for many years to come. Like most other surveys published in the *Encyklopädie*, Lorentz's article presented an exhaustive and systematic examination of the known results and existing literature in the field, including the most recent. The third basic text used in the seminar was Poincaré's treatise on electricity and optics (Poincaré 1901), based on his Sorbonne lectures of 1888, 1890 and 1891. This text discussed the various existing theories of the electrodynamics of moving bodies and criticized certain aspects of Lorentz's theory, and in particular a possible violation of the reaction principle due to its separation of matter and aether.[86]

Alongside the texts of Lorentz, Poincaré and Abraham, additional relevant works by Göttingen scientists were also studied. In fact, the main ideas of Abraham's theory had been recently elaborated by Schwarzschild and by Paul Hertz (1881–1940). The latter wrote a doctoral dissertation under the effective direction of Abraham, and this dissertation was studied at the seminar together with Schwarzschild's paper (Hertz 1904; Schwarzschild 1903). So were several recent papers by Sommerfeld (1904a, 1904b, 1905) who was now at Aachen, but who kept his strong ties to Göttingen always alive. Naturally, the ideas presented in the relevant works of Herglotz and Wiechert were also studied in the seminar (Herglotz 1903; Wiechert 1901).

The participants in this seminar discussed the current state of the theory, the relevant experimental work connected with it, and some of its most pressing open problems. The latter included the nature of the mass of the electron, problems related to rotation, vibration and acceleration in electron motion and their effects on the electromagnetic field, and the problem of faster-than-light motion. More briefly, they also studied the theory of dispersion and the Zeeman effect. From the point of view of the immediate development of the theory of relativity, it is indeed puzzling, as Lewis Pyenson has rightly stressed in his study of the seminar, that the participants were nowhere close to achieving the surprising breakthrough that Albert Einstein (1879–1956) had achieved at roughly the same time, and was about to publish (Pyenson 1979, 129–131).[87] Nevertheless, from the broader point of view of the development of math-

86 Cf. (Darrigol 2000, 351–366).

87 According to Pyenson, whereas Einstein "sought above all to address the most essential properties of nature," the Göttingen seminarists "sought to subdue nature, as it were, by the use of pure mathematics. They were not much interested in calculating with experimentally observable phenomena. They avoided studying electrons in metal conductors or at very low or high temperatures, and they did not spend much time elaborating the role of electrons in atomic spectra, a field of experimental physics then attracting the interest of scores of young physicists in their doctoral dissertations." Pyenson stresses the fact that Ritz's experiment was totally ignored at the seminar and adds: "For the seminar Dozenten it did not matter that accelerating an electron to velocities greater than that of light and even to infinite velocities made little physical sense. They pursued the problem because of its intrinsic, abstract interest." Noteworthy as these points are, it seems to me that by overstressing the question of why the Göttingen group achieved less than Einstein did, the main point is obscured in Pyenson's article, namely, what and why were Hilbert, Minkowski and their friends doing what they were doing, and how is this connected to the broader picture of their individual works and of the whole Göttingen mathematical culture.

ematics and physics at the turn of the century, and particularly of the account pursued here, it is all the more surprising to notice the level of detail and close acquaintance with physical theory and also, to a lesser degree, with experiment, that mathematicians such as Hilbert and Minkowski had attained by that time. All this, of course, while they were simultaneously active and highly productive in their own main fields of current, purely mathematical investigations: analysis, number theory, foundations, etc. Hilbert's involvement in the electron theory seminar clarifies the breadth and depth of the physical background that underlie his lectures on the axiomatization of physics in 1905, and that had considerably expanded in comparison with the one that prompted him to formulate, in the first place, his sixth problem back in 1900.

6. AXIOMS FOR PHYSICAL THEORIES: HILBERT'S 1905 LECTURES

Having described in some detail the relevant background, I now proceed to examine the contents of Hilbert's 1905 lectures on the "Axiomatization of Physical Theories," which devote separate sections to the following topics:

- Mechanics
- Thermodynamics
- Probability Calculus
- Kinetic Theory of Gases
- Insurance Mathematics
- Electrodynamics
- Psychophysics

Here I shall limit myself to discussing the sections on mechanics, the kinetic theory of gases, and electrodynamics.

6.1 Mechanics

Clearly, the main source of inspiration for this section is Aurel Voss's 1901 *Encyklopädie* article (Voss 1901). This is evident in the topics discussed, the authors quoted, the characterization of the possible kinds of axioms for physics, the specific axioms discussed, and sometimes even the order of exposition. Hilbert does not copy Voss, of course, and he introduces many ideas and formulations of his own, and yet the influence is clearly recognizable.

Though at this time Hilbert considered the axiomatization of physics and of natural science in general to be a task whose realization was still very distant,[88] we can mention one particular topic for which the axiomatic treatment had been almost com-

88 "Von einer durchgeführten axiomatischen Behandlung der Physik und der Naturwissenschaften ist man noch weit entfernt; nur auf einzelnen Teilgebieten finden sich Ansätze dazu, die nur in ganz wenigen Fällen durchgeführt sind. <Die Durchführung ist ein ganzes—grosses—Arbeitsprogramm, Vgl. Dissertation von Schimmack sowie Schur>." (Hilbert 1905a, 121)

pletely attained (and only very recently, for that matter). This is the "law of the parallelogram" or, what amounts to the same thing, the laws of vector-addition. Hilbert based his own axiomatic presentation of this topic on works by Darboux, by Hamel, and by one of his own students, Rudolf Schimmack (1881–1912).[89]

Hilbert defined a force as a three-component vector, and made no additional, explicit assumptions here about the nature of the vectors themselves, but it is implicitly clear that he had in mind the collection of all ordered triples of real numbers. Thus, as in his axiomatization of geometry, Hilbert was not referring to an arbitrary collection of abstract objects, but to a very concrete mathematical entity; in this case, one that had been increasingly adopted after 1890 in the treatment of physical theories, following the work of Oliver Heaviside (1850–1925) and Josiah Willard Gibbs (1839–1903).[90] In fact, in Schimmack's article of 1903—based on his doctoral dissertation—a vector was explicitly defined as a directed, real segment of line in the Euclidean space. Moreover, Schimmack defined two vectors as equal when their lengths as well as their directions coincide (Schimmack 1903, 318).

The axioms presented here were thus meant to define the addition of two such given vectors, as the sums of the components of the given vectors. At first sight, this very formulation could be taken as the single axiom needed to define the sum. But the task of axiomatic analysis is precisely to separate this single idea into a system of several, mutually independent, simpler notions that express the basic intuitions involved in it. Otherwise, it would be like taking the linearity of the equation representing the straight line as the starting point of geometry.[91] Hilbert had shown in his previous discussion on geometry that this latter result could be derived using all his axioms of geometry.

Hilbert thus formulated six axioms to define the addition of vectors: the first three assert the existence of a well-defined sum for any two given vectors (without stating what its value is), and the commutativity and associativity of this operation. The fourth axiom connects the resultant vector with the directions of the summed vectors as follows:

> 4. Let aA denote the vector $(aAx, aAy$,aAz $)$, having the same direction as A. Then every real number a defines the sum:
>
> $$A + aA = (1 + a)A.$$
>
> i.e., the addition of two vectors having the same direction is defined as the algebraic addition of the extensions along the straight line on which both vectors lie.[92]

---

89 The works referred to by Hilbert are (Darboux 1875; Hamel 1905; Schimmack 1903). Schimmak's paper was presented to the *Königliche Gesellschaft der Wissenschaften zu Göttingen* by Hilbert himself. An additional related work, also mentioned by Hilbert in the manuscript, is (Schur 1903).

90 Cf. (Crowe 1967, 150 ff.; Yavetz 1995).

91 "... das andere wäre genau dasselbe, wie wenn man in der Geometrie die Linearität der Geraden als einziges Axiom an die Spitze stellen wollte (vgl. S. 118)." (Hilbert 1905a, 123)

92 "Addition zweier Vektoren derselben Richtung geschieht durch algebraische Addition der Strecken auf der gemeinsamen Geraden." (Hilbert 1905a, 123)

The fifth one connects addition with rotation:

> 5. If D denotes a rotation of space around the common origin of two forces A and B, then the rotation of the sum of the vectors equals the sum of the two rotated vectors:
>
> $$D(A+B) = DA+DB$$
>
> i.e., the relative position of sum and components is invariant with respect to rotation.[93]

The sixth axiom concerns continuity:

> 6. Addition is a continuous operation, i.e., given a sufficiently small domain G around the end-point of $A+B$ one can always find domains G_1 and G_2, around the endpoints of A and B respectively, such that the endpoint of the sum of any two vectors belonging to each of these domains will always fall inside G.[94]

These are all simple axioms—Hilbert continued, without having really explained what a "simple" axiom is—and if we think of the vectors as representing forces, they also seem rather plausible. The axioms thus correspond to the basic known facts of experience, i.e., that the action of two forces on a point may always be replaced by a single one; that the order and the way in which they are added do not change the result; that two forces having one and the same direction can be replaced by a single force having the same direction; and, finally, that the relative position of the components and the resultant is independent of rotations of the coordinates. Finally, the demand for continuity in this system is similar and is formulated similarly to that of geometry.

That these six axioms are in fact necessary to define the law of the parallelogram was first claimed by Darboux, and later proven by Hamel. The main difficulties for this proof arose from the sixth axiom. Schimmack proved in 1903 the independence of the six axioms (in a somewhat different formulation), using the usual technique of models that satisfy all but one of the axioms. Hilbert also mentioned some possible modifications of this system. Thus, Darboux himself had showed that the continuity axiom may be abandoned, and in its place, it may be postulated that the resultant lies on the same plane as, and within the internal angle between, the two added vectors. Hamel, on the other hand, following a conjecture of Friedrich Schur, proved that the fifth axiom is superfluous if we assume that the locations of the endpoints of the resultants, seen as functions of the two added vectors, have a continuous derivative. In fact—Hilbert concluded—if we assume that all functions appearing in the natural sciences have at least one continuous derivative, and take this assumption as an even

---

93 "Nimmt man eine Drehung D des Zahlenraumes um den gemeinsamen Anfangspunkt vor, so entsteht aus $A+B$ die Summe der aus A und B durch D entstehenden Vektoren: $D(A+B) = DA+DB$; d.h. die relative Lage von Summe und Komponenten ist gegenüber allen Drehungen invariant." (Hilbert 1905a, 124)

94 "Zu einem genügend kleinen Gebiete G um den Endpunkt von $A+B$ kann man stets um die Endpunkte von A und B solche Gebiete G_1, G_2 abgrenzen, daß der Endpunkt der Summe jedes in G_1 u. G_2 endigenden Vectorpaares nach G fällt." (Hilbert 1905a, 124)

more basic axiom, then vector addition is defined by reference to only the four first axioms in the system.[95]

The sixth axiom, the axiom of continuity, plays a very central role in Hilbert's overall conception of the axiomatization of natural science—geometry, of course, included. It is part of the essence of things—Hilbert said in his lecture—that the axiom of continuity should appear in every geometrical or physical system. Therefore it can be formulated not just with reference to a specific domain, as was the case here for vector addition, but in a much more general way. A very similar opinion had been advanced by Hertz, as we saw, who described continuity as "an experience of the most general kind," and who saw it as a very basic assumption of all physical science. Boltzmann, in his 1897 textbook, had also pointed out the continuity of motion as the first basic assumption of mechanics, which in turn should provide the basis for all of physical science.[96] Hilbert advanced in his lectures the following general formulation of the principle of continuity:

> If a sufficiently small degree of accuracy is prescribed in advance as our condition for the fulfillment of a certain statement, then an adequate domain may be determined, within which one can freely choose the arguments [of the function defining the statement], without however deviating from the statement, more than allowed by the prescribed degree.[97]

Experiment—Hilbert continued—compels us to place this axiom at the top of every natural science, since it allows us to assert the validity of our assumptions and claims.[98] In every special case, this general axiom must be given the appropriate version, as Hilbert had shown for geometry in an earlier part of the lectures and here for vector addition. Of course there are many important differences between the Archimedean axiom, and the one formulated here for physical theories, but Hilbert seems to have preferred stressing the similarity rather than sharpening these differences. In fact, he suggested that from a strictly mathematical point of view, it would be possible to conceive interesting systems of physical axioms that do without continuity, that is, axioms that define a kind of "non-Archimedean physics." He did not consider such systems here, however, since the task was to see how the ideas and methods of axiomatics could be fruitfully applied to physics.[99] Nevertheless, this is an extremely important topic in Hilbert's axiomatic treatment of physical theories. When speaking of applying axiomatic ideas and methods to these theories, Hilbert

---

95 "Nimmt man nun von vornherein als Grundaxiom aller Naturwissenschaft an, daß alle auftretenden Funktionen einmal stetig differenzierbar sind, so kommt man hier mit den ersten 4 Axiomen aus." (Hilbert 1905a, 127)

96 Quoted in (Boltzmann 1974, 228–229).

97 "Schreibt man für die Erfüllung der Behauptung einen gewissen genügend kleinen Genaugikeitsgrad vor, so läßt sich ein Bereich angeben, innerhalb dessen man die Voraussetzungen frei wählen kann, ohne daß die Abweichung der Behauptung jenen vorgeschriebenen Grad überschreitet." (Hilbert 1905a, 125)

98 "Das Experiment zwingt uns geradezu dazu, ein solches Axiom an die Spitze aller Naturwissenschaft zu setzen, denn wir können bei ihm stets nur das ~~Ein~~<Zu>treffen von Voraussetzung und Behauptung mit einer gewissen beschränkten Genauigkeit feststellen." (Hilbert 1905a, 125–126)

meant in this case existing physical theories. But the possibility suggested here, of examining models of theories that preserve the basic logical structure of classical physics, except for a particular feature, opens the way to the introduction and systematic analysis of alternative theories, close enough to the existing ones in relevant respects. Hilbert's future works on physics, and in particular his work on general relativity, would rely on the actualization of this possibility.

An additional point that should be stressed in relation to Hilbert's treatment of vector addition has to do with his disciplinary conceptions. The idea of a vector space, and the operations with vectors as part of it, has been considered an integral part of algebra at least since the 1920s.[100] This was not the case for Hilbert, who did not bother here to make any connection between his axioms for vector addition and, say, the already well-known axiomatic definition of an abstract group. For Hilbert, as for the other mathematicians he cites in this section, this topic was part of physics rather than of algebra.[101] In fact, the articles by Hamel and by Schur were published in the *Zeitschrift für Mathematik und Physik*—a journal that bore the explicit subtitle: *Organ für angewandte Mathematik*. It had been established by Oscar Xavier Schlömilch (1823–1901) and by the turn of the century its editor was Carl Runge, the leading applied mathematician at Göttingen.

After the addition of vectors, Hilbert went on to discuss a second domain related to mechanics: statics. Specifically, he considered the axioms that describe the equilibrium conditions of a rigid body. The main concept here is that of a force, which can be described as a vector with an application point. The state of equilibrium is defined by the following axioms:

> I. Forces with a common application point are equivalent to their sum.
>
> II. Given two forces, K, L, with different application points, P, Q, if they have the same direction, and the latter coincides with the straight line connecting P and Q, then these forces are equivalent.
>
> III. A rigid body is in a state of equilibrium if all the forces applied to it taken together are equivalent to 0.[102]

---

99 "Rein mathematisch werden natürlich auch <physikalische> Axiomensysteme, die auf <diese> Stetigkeit Verzicht leisten, also eine 'nicht-Archimedische Physik' in erweitertem Sinne definieren, von hohem Interesse sein können; wir werden jedoch zunächst noch von ihrer Betrachtung absehen können, da es sich vorerst überhaupt nur darum handelt, die fruchtbaren Ideen und Methoden der Axiomatik in die Physik einzuführen." (Hilbert 1905a, 126)

100 Cf. (Dorier 1995; Moore 1995).

101 This point, which helps understanding Hilbert's conception of algebra, is discussed in detail in (Corry 2003, §3.4).

102 "1., Kräfte mit demselben Angriffspunkt sind ihrer Summe (im obigen Sinne) '*aequivalent*.' 2., 2 Kräfte K, L mit verschiedenen Angriffspunkten P, Q und dem gleichen (auch gleichgerichteten) Vektor, deren Richtung in die Verbindung P, Q fällt, heißen gleichfalls *aequivalent*. … Ein starrer Körper befindet sich im *Gleichgewicht*, wenn die an ihm angreifenden Kräfte zusammengenommen der Null aequivalent sind." (Hilbert 1905a, 127–128)

From these axioms, Hilbert asserted, the known formulae of equilibrium of forces lying on the same plane (e.g., for the case of a lever and or an inclined plane) can be deduced. As in the case of vector addition, Hilbert's main aim in formulating the axioms was to uncover the basic, empirical facts that underlie our perception of the phenomenon of equilibrium.

In the following lectures Hilbert analyzed in more detail the principles of mechanics and, in particular, the laws of motion. In order to study motion, one starts by assuming space and adds time to it. Since geometry provides the axiomatic study of space, the axiomatic study of motion will call for a similar analysis of time.

According to Hilbert, two basic properties define time: (1) its uniform passage and (2) its unidimensionality.[103] A consistent application of Hilbert's axiomatic approach to this characterization immediately leads to the question: Are these two independent facts given by intuition,[104] or are they derivable the one from the other? Since this question had very seldom been pursued, he said, one could only give a brief sketch of tentative answers to it. The unidimensionality of time is manifest in the fact, that, whereas to determine a point in space one needs three parameters, for time one needs only the single parameter t. This parameter t could obviously be transformed, by changing the marks that appear on our clocks,[105] which is perhaps impractical but certainly makes no logical difference. One can even take a discontinuous function for t, provided it is invertible and one-to-one,[106] though in general one does not want to deviate from the continuity principle, desirable for all the natural sciences. Hilbert's brief characterization of time would seem to allude to Carl Neumann's (Neumann 1870), in his attempt to reduce the principle of inertia into simpler ones.

Whereas time and space are alike in that, for both, arbitrarily large values of the parameters are materially inaccessible, a further basic difference between them is that time can be experimentally investigated in only one direction, namely, that of its increase.[107] While this limitation is closely connected to the unidimensionality of time,[108] the issue of the uniform passage of time is an experimental fact, which has to be deduced, according to Hilbert, from mechanics alone.[109] This claim was elaborated into a rather obscure discussion of the uniform passage for which, as usual, Hilbert gave no direct references, but which clearly harks back to Hertz's and Larmor's

103 "... ihr *gleichmäßiger Verlauf* und ihre *Eindimensionalität*." (Hilbert 1905a, 128)

104 "... anschauliche unabhängige Tatsachen." (Hilbert 1905a, 129)

105 "Es ist ohne weiteres klar, daß dieser Parameter t durch eine beliebige Funktion von sich ersetzt werden kann; das würde etwa nur auf eine andere Benennung der Ziffern der Uhr oder einen unregelmäßigen Gang des Zeigers hinauskommen." (Hilbert 1905a, 129)

106 One is reminded here of a similar explanation, though in a more general context, found in Hilbert's letter to Frege, on 29 December 1899. See (Gabriel et al. 1980, 41).

107 "~~Der~~ <Ein> wesentlicher Unterschied von Zeit und Raum ist nur der, daß wir in der Zeit nur in einem Sinne, dem des *wachsenden Parameters* experimentieren können, während Raum und Zeit darin übereinstimmen, daß uns *beliebig große Parameterwerte* unzugänglich sind." (Hilbert 1905a, 129)

108 Here Hilbert adds with his own handwriting (p. 130): <Astronomie! Wie wichtig wäre Beobachtungen in ferner Vergangenheit u. Zukunft!>."

109 "... eine experimentelle nur aus der Mechanik zu entnehmende Tatsache." (Hilbert 1905a, 130)

discussions and referred to by Voss as well, as mentioned earlier. I try to reproduce Hilbert's account here without really claiming to understand it. In short, Hilbert argued that if the flow of time were non-uniform then an essential difference between organic and inorganic matter would be reflected in the laws of mechanics, which is not actually the case. He drew attention to the fact that the differential expression $m \cdot d^2x/dt^2$ characterizes a specific physical situation only when it vanishes, namely, in the case of inertial motion. From a logical point of view, however, there is no apparent reason why the same situation might not be represented in terms of a more complicated expression, e.g., an expression of the form

$$m_1 \frac{d^2x}{dt^2} + m_2 \frac{dx}{dt}.$$

The magnitudes m_1 and m_2 may depend not only on time, but also on the kind of matter involved,[110] e.g., on whether organic or inorganic matter is involved. By means of a suitable change of variables, $t = t(\tau)$, this latter expression could in turn be transformed into $\mu \cdot d^2x/d\tau^2$, which would also depend on the kind of matter involved. Thus different kinds of substances would yield, under a suitable change of variables, different values of "time," values that nevertheless still satisfy the standard equations of mechanics. One could then use the most common kind of matter in order to measure time,[111] and when small variations of organic matter occurred along large changes in inorganic matter, clearly distinguishable non-uniformities in the passage of time would arise.[112] However, it is an intuitive (*anschauliche*) fact, indeed a mechanical axiom, Hilbert said, that the expression $m \cdot d^2x/dt^2$, always appears in the equations with *one and the same* parameter t, independently of the kind of substance involved, contrary to what the above discussion would seem to imply. This latter fact, to which Hilbert wanted to accord the status of axiom, is then the one that establishes the uniform character of the passage of time. Whatever the meaning and the validity of this strange argument, one source where Hilbert was likely to have found it is Aurel Voss's 1901 *Encyklopädie* article, which quotes, in this regard, similar passages of Larmor and Hertz.[113]

Following this analysis of the basic ideas behind the concept of time, Hilbert repeated the same kind of reasoning he had used in an earlier lecture concerning the role of continuity in physics. He suggested the possibility of elaborating a non-Galilean mechanics, i.e., a mechanics in which the measurement of time would depend on the kind of matter involved, in contrast to the characterization presented earlier in his lecture. This mechanics would, in most respects, be in accordance with

---

110 "... die m_1, m_2 von der Zeit, vor allem aber von dem Stoffe abhängig sein können." (Hilbert 1905a, 130)

111 "... der häufigste Stoff etwa kann dann zu Zeitmessungen verwandt werden." (Hilbert 1905a, 130–131)

112 "... für uns leicht große scheinbare Unstetigkeiten der Zeit auftreten." (Hilbert 1905a, 131)

113 See (Voss 1901, 14). Voss quoted (Larmor 1900, 288) and (Hertz 1894, 165).

the usual one, and thus one would be able to recognize which parts of mechanics depend essentially on the peculiar properties of time, and which parts do not. It is only in this way that the essence of the uniform passage of time can be elucidated, he thought, and one may thus at last understand the exact scope of the connection between this property and the other axioms of mechanics.

So much for the properties of space and time. Hilbert went on to discuss the properties of motion, while concentrating on a single material point. This is clearly the simplest case and therefore it is convenient for Hilbert's axiomatic analysis. However, it must be stressed that Hilbert was thereby distancing himself from Hertz's presentation of mechanics, in which the dynamics of single points is not contemplated. One of the axioms of statics formulated earlier in the course stated that a point is in equilibrium when the forces acting on it are equivalent to the null force. From this axiom, Hilbert derived the Newtonian law of motion:

$$m.\frac{d^2x}{dt^2} = X; m.\frac{d^2y}{dt^2} = Y; m.\frac{d^2z}{dt^2} = Z.$$

Newton himself, said Hilbert, had attempted to formulate a system of axioms for his mechanics, but his system was not very sharply elaborated and several objections could be raised against it. A detailed criticism, said Hilbert, was advanced by Mach in his *Mechanik*.[114]

The above axiom of motion holds for a free particle. If there are constraints, e.g., that the point be on a plane $f(x, y, z) = 0$ then one must introduce an additional axiom, namely, Gauss's principle of minimal constraint. Gauss's principle establishes that a particle in nature moves along the path that minimizes the following magnitude:

$$\frac{1}{m}\{(mx'' - X)^2 + (my'' - Y)^2 + (mz'' - Z)^2\} = \text{Minim.}$$

Here x'', y'', and z'' denote the components of the acceleration of the particle, and X, Y, Z the components of the moving force. Clearly, although Hilbert did not say it in his manuscript, if the particle is free from constraints, the above magnitude can actually become zero and we simply obtain the Newtonian law of motion. If there are constraints, however, the magnitude can still be minimized, thus yielding the motion of the particle.[115]

114 A detailed account of the kind of criticism advanced by Mach, and before him by Carl Neumann and Ludwig Lange, appears in (Barbour 1989, chap. 12).

115 For more detail on Gauss's principle, see (Lanczos 1962, 106–110). Interestingly, Lanczos points out that "Gauss was much attached to this principle because it represents a perfect physical analogy to the 'method of least squares' (discovered by him and independently by Legendre) in the adjustment of errors." Hilbert also discussed this latter method in subsequent lectures, but did not explicitly make any connection between Gauss's two contributions.

In his lectures, Hilbert explained in some detail how the Lagrangian equations of motion could be derived from this principle. But he also stressed that the Lagrangian equations could themselves be taken as axioms and set at the top of the whole of mechanics. In this case, the Newtonian and Galilean principles would no longer be considered as necessary assumptions of mechanics. Rather, they would be logical consequences of a distinct principle. Although this is a convenient approach that is often adopted by physicists, Hilbert remarked, it has the same kinds of disadvantages as deriving the whole of geometry from the demand of linearity for the equations of the straight line: many results can be derived from it, but it does not indicate what the *simplest* assumptions underlying the considered discipline may be. All the discussion up to this point, said Hilbert, concerns the simplest and oldest systems of axioms for the mechanics of systems of points. Beside them there is a long list of other possible systems of axioms for mechanics. The first of these is connected to the principle of conservation of energy, which Hilbert associated with the law of the impossibility of a *perpetuum mobile* and formulated as follows: "If a system is at rest and no forces are applied, then the system will remain at rest."[116]

Now the interesting question arises, how far can we develop the whole of mechanics by putting this law at the top? One should follow a process similar to the one applied in earlier lectures: to take a certain result that can be logically derived from the axioms and try to find out if, and to what extent, it can simply replace the basic axioms. In this case, it turns out that the law of conservation alone, as formulated above, is sufficient, though not necessary, for the derivation of the conditions of equilibrium in mechanics.[117] In order to account for the necessary conditions as well, the following axiom must be added: "A mechanical system can *only* be in equilibrium if, in accordance with the axiom of the impossibility of a *perpetuum mobile*, it is at rest."[118] The basic idea of deriving all of mechanics from this law, said Hilbert, was first introduced by Simon Stevin, in his law of equilibrium for objects in a slanted plane, but it was not clear to Stevin that what was actually involved was the reduction of the law to simpler axioms. The axiom was so absolutely obvious to Stevin, claimed Hilbert, that he had thought that a proof of it could be found without starting from any simpler assumptions.

From Hilbert's principle of conservation of energy, one can also derive the virtual velocities of the system, by adding a new axiom, namely, the principle of d'Alembert. This is done by placing in the equilibrium conditions, instead of the components

116 "Ist ein System in Ruhe und die Kräftefunction konstant (wirken keine Kräfte), so bleibt es in Ruhe." (Hilbert 1905a, 137)

117 "Es lässt sich zeigen, daß unter allen den Bedingungen, die die Gleichgewichtssätze der Mechanik liefern, wirklich Gleichgewicht eintritt." (Hilbert 1905a, 138)

118 "Es folgt jedoch nicht, daß diese Bedingungen auch *notwendig* für das Gleichgewicht sind, daß nicht etwa auch unter andern Umständen ein mechanisches System im Gleichgewicht sein kann. Es muß also noch ein Axiom hinzugenommen werden, des Inhaltes etwa: Ein mechanisches System kann *nur* dann im Gleichgewicht sein, wenn es dem Axiom von <der Unmöglichkeit des> Perpetuum mobile gemäß in Ruhe ist." (Hilbert 1905a, 138)

X, Y, Z of a given force-field acting on every mass point, the expressions $X - mx,'' Y - my'', Z - mz''$. . In other words, the principle establishes that motion takes place in such a way that at every instant of time, equilibrium obtains between the force and the acceleration. In this case we obtain a very systematic and simple derivation of the Lagrangian equations, and therefore of the whole of mechanics, from three axioms: the two connected with the principle of conservation of energy (as sufficient and necessary conditions) and d'Alembert's principle, added now.

A third way to derive mechanics is based on the concept of impulse. Instead of seeing the force field K as a continuous function of t, we consider K as first null, or of a very small value; then, suddenly, as increasing considerably in a very short interval, from t to $t + \tau$, and finally decreasing again suddenly. If one considers this kind of process at the limit, namely, when $\tau = 0$ one then obtains an impulse, which does not directly influence the acceleration, like a force, but rather creates a sudden velocity-change. The impulse is a time-independent vector, which however acts at a given point in time: at different points in time, different impulses may take place. The law that determines the action of an impulse is expressed by Bertrand's principle. This principle imposes certain conditions on the kinetic energy, so that it directly yields the velocity. It states that:

> The kinetic energy of a system set in motion as a consequence of an impulse must be maximal, as compared to the energies produced by all motions admissible under the principle of conservation of energy.[119]

The law of conservation is invoked here in order to establish that the total energy of the system is the same before and after the action of the impulse.

Bertrand's principle, like the others, could also be deduced from the elaborated body of mechanics by applying a limiting process. To illustrate this idea, Hilbert resorted to an analogy with optics: the impulse corresponds to the discontinuous change of the refraction coefficients affecting the velocity of light when it passes through the surface of contact between two media. But, again, as with the other alternative principles of mechanics, we could also begin with the concept of impulse as the basic one, in order to derive the whole of mechanics from it. This alternative assumes the possibility of constructing mechanics without having to start from the concept of force. Such a construction is based on considering a sequence of successive small impulses in arbitrarily small time-intervals, and in recovering, by a limiting process, the continuous action of a force. This process, however, necessitates the introduction of the continuity axiom discussed above. In this way, finally, the whole of mechanics is reconstructed using only two axioms: Bertrand's principle and the said axiom of continuity. In fact, this assertion of Hilbert is somewhat misleading, since his very formulation of Bertrand's principle presupposes the acceptance of the law of conservation of energy. In any case, Hilbert believed that also in this case, a

119 "Nach einem Impuls muß die kinetische Energie des Systems bei der <wirklich> eintretenden Bewegung ein Maximum sein gegenüber allen mit dem Satze von der Erhaltung der Energie verträglichen Bewegungen." (Hilbert 1905a, 141)

completely analogous process could be found in the construction of geometric optics: first one considers the process of sudden change of optical density that takes place in the surface that separates two media; then, one goes in the opposite direction, and considers, by means of a limiting process, the passage of a light ray through a medium with continuously varying optical density, seeing it as a succession of many infinitely small, sudden changes of density.

Another standard approach to the foundations of mechanics that Hilbert discussed is the one based on the use of the Hamiltonian principle as the only axiom. Consider a force field K and a potential scalar function U such that K is the gradient of U. If T is the kinetic energy of the system, then Hamilton's principle requires that the motion of the system from a given starting point, at time t_1, and an endpoint, at time t_2, takes place along the path that makes the integral

$$\int_{t_1}^{t_2}(T-U)dt$$

an *extremum* among all possible paths between those two points. The Lagrangian equations can be derived from this principle, and the principle is valid for continuous as well as for discrete masses. The principle is also valid for the case of additional constraints, insofar as these constraints do not contain differential quotients that depend on the velocity or on the direction of motion (non-holonomic conditions). Hilbert added that Gauss's principle was valid for this exception.

Hilbert's presentation of mechanics so far focused on approaches that had specifically been criticized by Hertz: the traditional one, based on the concepts of time, space, mass and force, and the energetic one, based on the use of Hamilton's principle. To conclude this section, Hilbert proceeded to discuss the approaches to the foundations of mechanics introduced in the textbooks of Hertz and Boltzmann respectively. Hilbert claimed that both intended to simplify mechanics, but each from an opposite perspective.

Expressing once again his admiration for the perfect Euclidean structure of Hertz's construction of mechanics,[120] Hilbert explained that for Hertz, all the effects of forces were to be explained by means of rigid connections between bodies; but he added that this explanation did not make clear whether one should take into account the atomistic structure of matter or not. Hertz's only axiom, as described by Hilbert, was the principle of the straightest path (*Das Prinzip von der geradesten Bahn*), which is a special case of the Gaussian principle of minimal constraint, for the force-free case. According to Hilbert, Hertz's principle is obtained from Gauss's by substituting in the place of the parameter t, the arc lengths s of the curve. The curvature

120 "Er liefert jedenfalls von dieser Grundlage aus in abstrakter und präcisester Weise einen wunderbaren Aufbau der Mechanik, indem er ganz nach Euklidischem Ideale ein vollständiges System von Axiomen und Definitionen aufstellt." (Hilbert 1905a, 146)

$$m.\left\{\left(\frac{d^2x}{ds^2}\right)^2+\left(\frac{d^2y}{ds^2}\right)^2+\left(\frac{d^2z}{ds^2}\right)^2\right\}$$

of the path is to be minimized, in each of its points, when compared with all the other possible paths in the same direction that satisfy the constraint. On this path, the body moves uniformly if one also assumes Newton's first law.[121] In fact, this requirement had been pointed out by Hertz himself in the introduction to the Principles. As one of the advantages of his mathematical formulation, Hertz mentioned the fact that he does not need to assume, with Gauss, that nature intentionally keeps a certain quantity (the constraint) as small as possible. Hertz felt uncomfortable with such assumptions.

Boltzmann, contrary to Hertz, intended to explain the constraints and the rigid connections through the effects of forces, and in particular, of central forces between any two mass points. Boltzmann's presentation of mechanics, according to Hilbert, was less perfect and less fully elaborated than that of Hertz.

In discussing the principles of mechanics in 1905, Hilbert did not explicitly separate differential and integral principles. Nor did he comment on the fundamental differences between the two kinds. He did so, however, in the next winter semester, in a course devoted exclusively to mechanics (Hilbert 1905–6, §3.1.2).[122]

Hilbert closed his discussion on the axiomatics of mechanics with a very interesting, though rather speculative, discussion involving Newtonian astronomy and continuum mechanics, in which methodological and formal considerations led him to ponder the possibility of unifying mechanics and electrodynamics. It should be remarked that neither Einstein's nor Poincaré's 1905 articles on the electrodynamics of moving bodies is mentioned in any of Hilbert's 1905 lectures; most likely, Hilbert was not aware of these works at the time.[123] Hilbert's brief remarks here, on the other hand, strongly bring to mind the kind of argument, and even the notation, used by Minkowski in his first public lectures on these topics in 1907 in Göttingen.

Earlier presentations of mechanics, Hilbert said, considered the force—expressed in terms of a vector field—as given, and then investigated its effect on motion. In Boltzmann's and Hertz's presentations, for the first time, force and motion were con-

121 "Die Bewegung eines jeden Systemes erfolgt gleichförmig in einer '*geradesten Bahn*', d.h. für einen Punkt: die Krümmung

$$m.\left\{\left(\frac{d^2x}{ds^2}\right)^2+\left(\frac{d^2y}{ds^2}\right)^2+\left(\frac{d^2z}{ds^2}\right)^2\right\}$$

der Bahnkurve soll ein Minimum sein, in jedem Orte, verglichen mit allen andern den Zwangsbedingungen gehorchenden Bahnen derselben Richtung, und auf dieser Bahn bewegt sich der Punkt gleichförmig." (Hilbert 1905a, 146–147)

122 The contents of this course are analyzed in some detail in (Blum 1994).

123 This particular lecture of Hilbert is dated in the manuscript 26 July 1905, whereas Poincaré's article was submitted for publication on 23 July 1905, and Einstein's paper three weeks later. Poincaré had published a short announcement on 5 June 1905, in the *Comptes rendus* of the Paris Academy of Sciences.

sidered not as separate, but rather as closely interconnected and mutually interacting, concepts. Astronomy is the best domain in which to understand this interaction, since Newtonian gravitation is the only force acting on the system of celestial bodies. In this system, however, the force acting on a mass point depends not only on its own position but also on the positions and on the motions of the other points. Thus, the motions of the points and the acting forces can only be determined simultaneously. The potential energy in a Newtonian system composed of two points $(a|b|c)$ and $(x|y|z)$ equals, as it is well-known, $-\frac{1}{r_{a,b,c \atop x,y,z}}$, the denominator of this fraction being the distance between the two points. This is a symmetric function of the two points, and thus it conforms to Newton's law of the equality of action and reaction. Starting from these general remarks, Hilbert went on to discuss some ideas that, he said, came from an earlier work of Boltzmann and which might lead to interesting results. Which of Boltzmann's works Hilbert was referring to here is not stated in the manuscript. However, from the ensuing discussion it is evident that Hilbert had in mind a short article by Boltzmann concerning the application of Hertz's perspective to continuum mechanics (Boltzmann 1900).

Hertz himself had already anticipated the possibility of extending his point of view from particles to continua. In 1900 Richard Reiff (1855–1908) published an article that developed this direction (Reiff 1900), and soon Boltzmann published a reply pointing out an error. Boltzmann indicated, however, that Hertz's point of view could be correctly extended to include continua, the possibility seemed to arise of constructing a detailed account of the whole world of observable phenomena.[124] Boltzmann meant by this that one could conceivably follow an idea developed by Lord Kelvin, J.J. Thomson and others, that considered atoms as vortices or other similar stationary motion phenomena in incompressible fluids; this would offer a concrete representation of Hertz's concealed motions and could provide the basis for explaining all natural phenomena. Such a perspective, however, would require the addition of many new hypotheses which would be no less artificial than the hypothesis of action at a distance between atoms, and therefore—at least given the current state of physical knowledge—little would be gained by pursuing it.

Boltzmann's article also contained a more positive suggestion, related to the study of the mechanics of continua in the spirit of Hertz. Following a suggestion of Brill, Boltzmann proposed to modify the accepted Eulerian approach to this issue. The latter consisted in taking a fixed point in space and deriving the equations of motion of the fluid by studying the behavior of the latter at the given point. Instead of this Boltzmann suggested a Lagrangian approach, deducing the equations by looking at an element of the fluid as it moves through space. This approach seemed to Boltzmann to be the natural way to extend Hertz's point of view from particles to continua,

124 "... ein detailliertes Bild der gesamten Erscheinungswelt zu erhalten." (Boltzmann 1900, 668)

and he was confident that it would lead to the equations of motion of an incompressible fluid as well as to those of a rigid body submerged in such a fluid.[125] In 1903 Boltzmann repeated these ideas in a seminar taught in Vienna, and one of his students decided to take the problem as the topic of his doctoral dissertation of 1904: this was Paul Ehrenfest (1880–1993). Starting from Boltzmann's suggestion, Ehrenfest studied various aspects of the mechanics of continua using a Lagrangian approach. In fact, Ehrenfest in his dissertation used the terms Eulerian and Lagrangian with the meaning intended here, as Boltzmann in his 1900 article had not (Ehrenfest 1904, 4–5). The results obtained in the dissertation helped to clarify the relations between the differential and the integral variational principles for non-holonomic systems, but they offered no real contribution to an understanding of all physical phenomena in terms of concealed motions and masses, as Boltzmann and Ehrenfest may have hoped.[126]

Ehrenfest studied in Göttingen between 1901 and 1903, and returned there in 1906 for one year, before moving with his mathematician wife Tatyana to St. Petersburg. We don't know the details of Ehrenfest's attendance at Hilbert's lectures during his first stay in Göttingen. Hilbert taught courses on the mechanics of continua in the winter semester of 1902–1903 and in the following summer semester of 1903, which Ehrenfest may well have attended. Nor do we know whether Hilbert knew anything about Ehrenfest's dissertation when he taught his course in 1905. But be that as it may, at this point in his lectures, Hilbert connected his consideration of Newtonian astronomy to the equations of continuum mechanics, while referring to the dichotomy between the Lagrangian and the Eulerian approach, and using precisely those terms. Interestingly enough, the idea that Hilbert pursued in response to Boltzmann's article was not that the Lagrangian approach would be the natural one for studying mechanics of continua, but rather the opposite, namely, that a study of the continua following the Eulerian approach, and assuming an atomistic worldview, could lead to a unified explanation of all natural phenomena.

Consider a free system subject only to central forces acting between its mass-points —and in particular only forces that satisfy Newton's law, as described above. An axiomatic description of this system would include the usual axioms of mechanics, together with the Newtonian law as an additional one. We want to express this system, said Hilbert, as concisely as possible by means of differential equations. In the most general case we assume the existence of a continuous mass distribution in space, $\rho = (x, y, z, t)$. In special cases we have $\rho = 0$ within a well-delimited region; the case of astronomy, in which the planets are considered mass-points, can be derived from this special case by a process of passage to the limit. Hilbert explained what the Lagrangian approach to this problem would entail. That approach, he added, is the most appropriate one for discrete systems, but often it is also conveniently used in the mechanics of continua. Here, however, he would follow the Eulerian approach

125 For more details, cf. (Klein 1970, 64–66).
126 For details on Ehrenfest's dissertation, see (Klein 1970, 66–74).

to derive equations of the motion of a unit mass-particle in a continuum. The ideas discussed in this section, as well as in many other parts of this course, hark back to those he developed in somewhat greater technical detail in his 1902–1903 course on continuum mechanics, but here a greater conceptual clarity and a better understanding of the possible, underlying connections across disciplines is attained, thanks to the systematic use of an axiomatic approach in the discussion.

Let V denote the velocity of the particle at time t and at coordinates (x, y, z) in the continuum. V has three components $u = u(x, y, z, t), v$ and w. The acceleration vector for the unit particle is given by dV/dt, which Hilbert wrote as follows:[127]

$$\frac{dV}{dt} = \frac{\partial V}{\partial t} + u\frac{\partial V}{\partial x} + v\frac{\partial V}{\partial y} + w\frac{\partial V}{\partial z} = \frac{\partial V}{\partial t} + V \times \operatorname{curl} V - \frac{1}{2}\operatorname{grad}(V.V).$$

Since the only force acting on the system is Newtonian attraction, the potential energy at a point $(x|y|z)$ is given by

$$P = -\iiint \frac{\rho'}{r_{\substack{x',y',z'\\x,y,z}}} dx'dy'dz'$$

where ρ' is the mass density at the point $(x'|y'|z')$. The gradient of this potential equals the force acting on the particle, and therefore we obtain three equations of motion that can succinctly be expressed as follows:

$$\frac{\partial V}{\partial t} + V \times \operatorname{curl} V - \frac{1}{2}\operatorname{grad}(V.V) = \operatorname{grad} P.$$

One can add two additional equations to these three. First, the Poisson equation, which Hilbert calls "potential equation of Laplace":

$$\Delta P = 4\pi\rho$$

where Δ denotes the Laplacian operator (currently written as ∇^2). Second, the constancy of the mass in the system is established by means of the continuity equation:[128]

$$\frac{\partial\rho}{\partial t} = -\operatorname{div}(\rho \cdot V).$$

We have thus obtained five differential equations involving five functions (the components u, v, w of V, P and ρ of the four variables x, y, z, t. The equations are

---

127 In the manuscript the formula in the leftmost side of the equation appears twice, having a "-" sign in front of $V \times \operatorname{curl} V$. This is obviously a misprint, as a straightforward calculation readily shows.

128 In his article mentioned above, Reiff had tried to derive the pressure forces in a fluid starting only from the conservation of mass (Reiff 1900). Boltzmann pointed out that Reiff had obtained a correct result because of a compensation error in his mathematics. See (Klein 1970, 65).

completely determined when we know their initial values and other boundary conditions, such as the values of the functions at infinity. Hilbert called the five equations so obtained the "Newtonian world-functions," since they account in the most general way and in an axiomatic fashion for the motion of the system in question: a system that satisfies the laws of mechanics and the Newtonian gravitational law. It is interesting that Hilbert used the term "world-function" in this context, since the similar ones "world-point" and "world-postulate," were introduced in 1908 by Minkowski in the context of his work on electrodynamics and the postulate of relativity. Unlike most of the mathematical tools and terms introduced by Minkowski, this particular aspect of his work was not favorably received, and is hardly found in later sources (with the exception of "world-line"). Hilbert, however, used the term "world-function" not only in his 1905 lectures, but also again in his 1915 work on general relativity, where he again referred to the Lagrangian function used in the variational derivation of the gravitational field equations as a "world-function."

Besides the more purely physical background to the issues raised here, it is easy to detect that Hilbert was excited about the advantages and the insights afforded by the vectorial formulation of the Eulerian equations. Vectorial analysis as a systematic way of dealing with physical phenomena was a fairly recent development that had crystallized towards the turn of the century, mainly through its application by Heaviside in the context of electromagnetism and through the more mathematical discussion of the alternative systems by Gibbs.[129] The possibility of extending its use to disciplines like hydrodynamics had arisen even more recently, especially in the context of the German-speaking world. Thus, for instance, the *Encyklopädie* article on hydrodynamics, written in 1901, still used the pre-vectorial notation (Love 1901, 62–63).[130] Only one year before Hilbert's course, speaking at the International Congress of Mathematicians in Heidelberg, the Göttingen applied mathematician Ludwig Prandtl still had to explain to his audience how to write the basic equations of hydrodynamics "following Gibbs's notation" (Prandtl 1904, 489). Among German textbooks on vectorial analysis of the turn of the century,[131] formulations of the Eulerian equations like that quoted above appear in Alfred Heinrich Bucherer's textbook of 1903 (Bucherer 1903, 77–84) and in Richard Gans's book of 1905 (Gans 1905, 66–67). Whether he learnt about the usefulness of the vectorial notation in this context from his colleague Prandtl or from one of these textbooks, Hilbert was certainly impressed by the unified perspective it afforded from the formal point of view. Moreover, he seems also to have wanted to deduce far-reaching physical conclusions from this formal similarity. Hilbert pointed out in his lectures the strong analogy between this formulation of the equations and Maxwell's equations of electrodynamics, though in the latter we have two vectors E, and B, the electric and the magnetic fields, against only one here, V.

129 Cf. (Crowe 1967, 182–224).

130 The same is the case for (Lamb 1895, 7). This classical textbook, however, saw many later editions in which the vectorial formulation was indeed adopted.

131 Cf. (Crowe 1967, 226–233).

He also raised the following question: can one obtain the whole of mechanics starting from these five partial equations as a single axiom, or, if that is not the case, how far can its derivation in fact be carried? In other words: if we want to derive the whole of mechanics, to what extent can we limit ourselves to assuming only Newtonian attraction or the corresponding field equations?[132] It would also be interesting, he said, to address the question of how far the analogy of gravitation with electrodynamics can be extended. Perhaps, he said, one can expect to find a formula that simultaneously encompasses these five equations and the Maxwellian ones together. This discussion of a possible unification of mechanics and electrodynamics also echoed, of course, the current foundational discussion that I have described in the preceding sections. It also anticipates what will turn out to be one of the pillars of Hilbert's involvement with general relativity in 1915.

Hilbert's reference to Hertz and Boltzmann in this context, and his silence concerning recent works of Lorentz, Wien, and others, is the only hint he gave in his 1905 lectures as to his own position on the foundational questions of physics. In fact, throughout these lectures Hilbert showed little inclination to take a stand on physical issues of this kind. Thus, his suggestion of unifying the equations of gravitation and electrodynamics was advanced here mainly on methodological grounds, rather than expressing, at this stage at least, any specific commitment to an underlying unified vision of nature. At the same time, however, his suggestion is quite characteristic of the kind of mathematical reasoning that would allow him in later years to entertain the possibility of unification and to develop the mathematical and physical consequences that could be derived from it.

6.2 Kinetic Theory of Gases

A main application of the calculus of probabilities that Hilbert considered is in the kinetic theory of gases. He opened this section by expressing his admiration for the remarkable way this theory combined the postulation of far-reaching assumptions about the structure of matter with the use of probability calculus, a combination that had been applied in a very illuminating way, leading to new physical results. Several works that appeared by end of the nineteenth century had changed the whole field of the study of gases, thus leading to a more widespread appreciation of the value of the statistical approach. The work of Planck, Gibbs and Einstein attracted a greater interest in and contributed to an understanding of Boltzmann's statistical interpretation of entropy.[133]

132 "Es wäre nun die Frage, ob man mit ~~einem~~ diesen 5 partiellen Gleichungen als einzigem Axiom nicht auch überhaupt in der Mechanik auskommt, oder wie weit das geht, d.h. wie weit man sich auf Newtonsche Attraktion bezw. auf die entsprechenden Feldgleichungen beschränken kann." (Hilbert 1905a, 154)

133 Kuhn (1978, 21) quotes in this respect the well-known textbook, (Gibbs 1902), and an "almost forgotten" work, (Einstein 1902).

It is easy to see, then, why Hilbert would have wished to undertake an axiomatic treatment of the kinetic theory of gases: not only because it combined physical hypotheses with probabilistic reasoning in a scientifically fruitful way, as Hilbert said in these lectures, but also because the kinetic theory was a good example of a physical theory where, historically speaking, additional assumptions had been gradually added to existing knowledge without properly checking the possible logical difficulties that would arise from this addition. The question of the role of probability arguments in physics was not settled in this context. In Hilbert's view, the axiomatic treatment was the proper way to restore order to this whole system of knowledge, so crucial to the contemporary conception of physical science.

In stating the aim of the theory as the description of the macroscopic states of a gas, based on statistical considerations about the molecules that compose it, Hilbert assumed without any further comment the atomistic conception of matter. From this picture, he said, one obtains, for instance, the pressure of the gas as the number of impacts of the gas molecules against the walls of its container, and the temperature as the square of the sum of the mean velocities. In the same way, entropy becomes a magnitude with a more concrete physical meaning than is the case outside the theory. Using Maxwell's velocity distribution function, Boltzmann's logarithmic definition of entropy, and the calculus of probabilities, one obtains the law of constant increase in entropy. Hilbert immediately pointed out the difficulty of combining this latter result with the reversibility of the laws of mechanics. He characterized this difficulty as a paradox, or at least as a result not yet completely well established.[134] In fact, he stressed that the theory had not yet provided a solid justification for its assumptions, and ever new ideas and stimuli were constantly still being added.

Even if we knew the exact position and velocities of the particles of a gas— Hilbert explained—it is impossible in practice to integrate all the differential equations describing the motions of these particles and their interactions. We know nothing of the motion of individual particles, but rather consider only the average magnitudes that are dealt with by the probabilistic kinetic theory of gases. In an oblique reference to Boltzmann's replies, Hilbert stated that the combined use of probabilities and infinitesimal calculus in this context is a very original mathematical contribution, which may lead to deep and interesting consequences, but which at this stage has in no sense been fully justified. Take, for instance, one of the well-known results of the theory, namely, the equations of vis viva. In the probabilistic version of the theory, Hilbert said, the solution of the corresponding differential equation does not emerge solely from the differential calculus, and yet it is correctly determined. It might conceivably be the case, however, that the probability calculus could have contradicted well-known results of the theory, in which case, using that calculus would clearly yield what would be considered unacceptable conclusions. Hilbert explained this

134 "Hier können wir aber bereits ein paradoxes, zum mindesten nicht recht befriedigendes Resultat feststellen." (Hilbert 1905a, 176)

warning by showing how a fallacious probabilistic argument could lead to contradiction in the theory of numbers.

Take the five classes of congruence module 5 in the natural numbers, and consider how the prime numbers are distributed among these classes. For any integer x, let $A(x)$ be the number of prime numbers which are less than x, and let $A_0(x),\ldots,A_4(x)$, be the corresponding values of the same function, when only the numbers in each of the five classes are considered. Using the calculus of probabilities in a similar way to that used in the integration of the equations of motion of gas particles, one could reason as follows: The distribution of prime numbers is very irregular, but according to the laws of probability, this irregularity is compensated if we just take a large enough quantity of events. In particular, the limits at infinity of the quotients $A_i(x)/A(x)$ are all equal for $i = 0,\ldots,4$, and therefore equal to $1/5$. But it is clear, on the other hand, that in the class of numbers of the form $5m$, there are no prime numbers, and therefore $A_0(x)/A(x) = 0$. One could perhaps correct the argument by limiting its validity to the other four classes, and thus conclude that:

$$\underset{x=\infty}{L} \frac{A_i(x)}{A(x)} = \frac{1}{4}, \text{ for } i = 1,2,3,4.$$

Although this latter result is actually correct, Hilbert said, one cannot speak here of a real proof. The latter could only be obtained through deep research in the theory of numbers. Had we not used here the obvious number-theoretical fact that $5m$ can never be a prime number, we might have been misled by the probabilistic proof. Something similar happens in the kinetic theory of gases, concerning the integration of the vis viva. One assumes that Maxwell's distribution of velocities obeys a certain differential equation of mechanics, and in this way a contradiction with the known value of the integral of the vis viva is avoided. Moreover, according to the theory, because additional properties of the motion of the gas particles, which are prescribed by the differential equations, lie very deep and are only subtly distinguishable, they do not affect relatively larger values, such as the averages used in the Maxwell laws.[135] As in the case of the prime numbers, however, Hilbert did not consider this kind of reasoning to be a real proof.

All this discussion, which Hilbert elaborated in further detail, led him to formulate his view concerning the role of probabilistic arguments in mathematical and physical theories. In this view, surprisingly empiricist and straightforwardly formu-

135 "Genau so ist es nun hier in der kinetischen Gastheorie. Indem wir behaupten, daß die Maxwellsche Geschwindigkeitsverteilung den mechanischen Differentialgleichnungen genügt, vermeiden wir wohl einen Verstoß gegen das sofort bekannte Integral der lebendigen Kraft; weiterhin aber wird die Annahme gemacht, ~~daß~~ die durch die Differentialgleichungen geforderten weiteren Eigenschaften der Gaspartikelbewegung liegen soviel tiefer und sind so feine Unterscheidungen, daß sie so grobe Aussagen über mittlere Werte, wie die des Maxwellschen Gesetzes, nicht berühren." (Hilbert 1905a, 180–181)

lated, the calculus of probability is not an exact mathematical theory, but one that may appropriately be used as a first approximation, provided we are dealing with immediately apparent mathematical facts. Otherwise it may lead to significant contradictions. The use of the calculus of probabilities is justified—Hilbert concluded—insofar as it leads to results that are correct and in accordance with the facts of experience or with the accepted mathematical theories.[136]

Beginning in 1910 Hilbert taught courses on the kinetic theory of gases and on related issues, and also published original contributions to this domain. In particular, as part of his research on the theory of integral equations, which began around 1902, he solved in 1912 the so-called Boltzmann equation.[137]

6.3 Electrodynamics

The manuscript of the lecturer indicates that Hilbert did not discuss electrodynamics before 14 July 1905. By that time Hilbert must have been deeply involved with the issues studied in the electron-theory seminar. These issues must surely have appeared in the lectures as well, although the rather elementary level of discussion in the lectures differed enormously from the very advanced mathematical sophistication characteristic of the seminar. As mentioned above, at the end of his lectures on mechanics Hilbert had addressed the question of a possible unification of the equations of gravitation and electrodynamics, mainly based on methodological considerations. Now he stressed once more the similarities underlying the treatment of different physical domains. In order to provide an axiomatic treatment of electrodynamics similar to those of the domains discussed above—Hilbert opened this part of his lectures—one needs to account for the motion of an electron by describing it as a small electrified sphere and by applying a process of passage to the limit.

One starts therefore by considering a material point m in the classical presentation of mechanics. The kinetic energy of a mass-point is expressed as

$$L(v)= \frac{1}{2}mv^2.$$

The derivatives of this expression with respect to the components v_s of the velocity v define the respective components of the momentum

$$\frac{\partial L(v)}{\partial v_s}= m.v_s.$$

136 "… sie ist keine exakte mathematische Theorie, aber zu einer ersten Orientierung, wenn man nur alle unmittelbar leicht ersichtlichen mathematischen Tatsachen benutzt, häufig sehr geeignet; sonst führt sie sofort zu groben Verstößen. Am besten kann man wohl immer nachträglich sagen, daß die Anwendung der Wahrscheinlichkeitsrechnung immer dann berechtigt und erlaubt ist, wo sie zu richtigen, mit der Erfahrung bezw. der sonstigen mathematischen Theorie übereinstimmenden Resultaten führt." (Hilbert 1905a, 182–183)

137 In (Hilbert 1912a, chap. XXII).

If one equates the derivative of the latter with respect to time to the components of the forces—seen as the negative of the partial derivatives of the potential energy—one gets the equations of motion:

$$\frac{d\frac{\partial}{\partial v_s}}{dt}+\frac{\partial U}{\partial s}=0 \qquad s=(x,y,z).$$

As was seen earlier in the lectures on mechanics, an alternative way to attain these equations is to use the functions L, U and the variational equation characteristic of the Hamiltonian principle:

$$\int_{t_1}^{t_2}(L-U)dt=\text{ Minim.}$$

This principle can be applied, as Laplace did in his *Celestial Mechanics*, even without knowing anything about L, except that it is a function of the velocity. In order to determine the actual form of L, one must then introduce additional axioms. Hilbert explained that in the context of classical mechanics, Laplace had done this simply by asserting what for him was an obvious, intuitive notion concerning relative motion, namely, that we are not able to perceive any uniform motion of the whole universe.[138] From this assumption Laplace was able to derive the actual value $L(v) = (1/2)mv^2$. This was for Hilbert a classical instance of the main task of the axiomatization of a physical science, as he himself had been doing throughout his lectures for the cases of the addition of vectors, thermodynamics, insurance mathematics, etc.: namely, to formulate the specific axiom or axioms underlying a particular physical theory, from which the specific form of its central, defining function may be derived. In this case, Laplace's axiom is nothing but the expression of the Galilean invariance of the Newtonian laws of motion although Hilbert did not use this terminology here.

In the case of the electron, as Hilbert had perhaps recently learnt in the electron-theory seminar, this axiom of Galilean invariance, is no longer valid, nor is the specific form of the Lagrangian function. Yet—and this is what Hilbert stressed as a remarkable fact—the equation of motion of the electron can nevertheless be derived following considerations similar to those applied in Laplace's case. One need only find the appropriate axiom to effect the derivation. Without further explanation, Hil-

138 "Zur Festlegung von L muß man nun natürlich noch Axiome hinzunehmen, und Laplace kommt da mit einer allgemeinen, ihm unmittelbar anschaulichen Vorstellung über Relativbewegung aus, daß wir nämlich eine gleichförmige Bewegung des ganzen Weltalls nicht merken würden. Alsdann läßt sich die Form $mv^2/2$ von $L(v)$ bestimmen, und das ist wieder die ganz analoge Aufgabe zu denen, die das Fundament der Vektoraddition, der Thermodynamik, der Lebensversicherungsmathematik u.a. bildeten." (Hilbert 1905a, 187)

bert wrote down the Lagrangian that describes the motion of the electron. This may be expressed as

$$L(v) = \mu \frac{1-v^2}{v} \cdot \log \frac{1+v}{1-v}$$

where v denotes the ratio between the velocity of the electron and the speed of light, and μ is a constant, characteristic of the electron and dependent on its charge. This Lagrangian appears, for instance, in Abraham's first article on the dynamics of the electron, and a similar one appears in the article on Lorentz's *Encyklopädie* article.[139] If not earlier than that, Hilbert had studied these articles in detail in the seminar, where Lorentz's article was used as a main text.

If, as in the case of classical mechanics, one again chooses to consider the differential equation or the corresponding variational equation as the single, central axiom of electron theory, taking L as an undetermined function of v whose exact expression one seeks to derive, then—Hilbert said—in order to do so, one must introduce a specific axiom, characteristic of the theory and as simple and plausible as possible. Clearly—he said concluding this section—this theory will require more, or more complicated, axioms than the one introduced by Laplace in the case of classical mechanics.[140] The electron-theory seminar had been discussing many recent contributions, by people such as Poincaré, Lorentz, Abraham and Schwarzschild, who held conflicting views on many important issues. It was thus clear to Hilbert that, at that point in time at least, it would be too early to advance any definite opinion as to the specific axiom or axioms that should be placed at the basis of the theory. This fact, however, should not affect in principle his argument as to how the axiomatic approach should be applied to the theory.

It is noteworthy that in 1905 Hilbert did not mention the Lorentz transformations, which were to receive very much attention in his later lectures on physics. Lorentz published the transformations in an article of 1904 (Lorentz 1904b), but this article was not listed in the bibliography of the electron theory seminar,[141]and it is likely that Hilbert was not aware of it by the time of his lectures.

6.4 A post-1909 addendum

To conclude this account of the 1905 lectures, it is interesting to notice that several years after having taught the course, Hilbert returned to the manuscript and added

139 Respectively, (Abraham 1902, 37; Lorentz 1904a, 184). Lorentz's Lagrangian is somewhat different, since it contains two additional terms, involving the inverse of v^3.

140 "Nimmt man nun wieder die Differentialgleichungen bzw. das zugehörige Variationsproblem als Axiom und läßt L zunächst als noch unbestimmte Funktion von v stehen, so handelt es sich darum, dafür möglichst einfache und plausible Axiome so zu konstruiren, daß sie gerade jene Form von $L(v)$ bestimmen. Natürlich werden wir mehr oder kompliciertere Axiome brauchen, als in dem einfachen Falle der Mechanik bei Laplace." (Hilbert 1905a, 188)

141 Cf. (Pyenson 1979, 103).

some remarks on the front page in his own handwriting. He mentioned two more recent works he thought relevant to understanding the use of the axiomatic method in physics. First, he referred to a new article by Hamel on the principles of mechanics. Hamel's article, published in 1909, contained philosophical and critical remarks concerning the issues discussed in his own earlier article of 1905 (the one mentioned by Hilbert with reference to the axiomatization of vector addition). In particular, it discussed the concepts of absolute space, absolute time and force, as a priori concepts of mechanics. The contents of this article are beyond the scope of our discussion here. Hilbert's interest in it may have stemmed from a brief passage where Hamel discussed the significance of Hilbert's axiomatic method (Hamel 1909, 358). More importantly perhaps, it also contained an account of a new system of axioms for mechanics.[142]

Second, in a formulation that condenses in a very few sentences his understanding of the principles and goals of axiomatization, as they apply to geometry and to various domains of physics, Hilbert also directed attention to what he saw as Planck's application of the axiomatic method in the latter's recent research on quantum theory. Hilbert thus wrote:

> It is of special interest to notice how the axiomatic method is put to use by Planck—in a more or less consistent and in a more or less conscious manner—even in modern quantum theory, where the basic concepts have been so scantily clarified. In doing this, he sets aside electrodynamics in order to avoid contradiction, much as, in geometry, continuity is set aside in order to remove the contradiction in non-Pascalian geometry, or like, in the theory of gases, mechanics is set aside in favor of the axiom of probability (maximal entropy), thus applying only the Stossformel or the Liouville theorem, in order to avoid the objections involved in the reversibility and recurrence paradoxes.[143]

From this remark we learn not only that Hilbert was aware of the latest advances in quantum theory (though, most probably, not in great detail) but also that he had a good knowledge of recent writings of Paul and Tatyana Ehrenfest. Beginning in 1906 the Ehrenfests had made important contributions to clarifying Boltzmann's ideas in a series of publications on the conceptual foundations of statistical mechanics. The two last terms used by Hilbert in his hand-written remark (*Umkehr- oder Wiederkehreinwand)* were introduced only in 1907 by them, and were made widely known only

142 According to Clifford Truesdell (1968, 336), this article of Hamel, together with the much later (Noll 1959), are the "only two significant attempts to solve the part of Hilbert's sixth problem that concern mechanics [that] have been published." One should add to this list at least another long article (Hamel 1927) that appeared in vol. 5 of the *Handbuch der Physik*.

143 Hilbert (1905a), added "<Besonders interessant ist es zu sehen, wie die axiomatische Methode von Planck sogar bei der modernen Quantentheorie, wo die Grundbegriffe noch so wenig geklärt sind, in mehr oder weniger konsequenter und in mehr oder weniger bewusster Weise zur Anwendung gebracht werden: dabei Ausschaltung der Elektrodynamik, um Widerspruch zu vermeiden—gerade wie in der Geometrie Ausschaltung der Stetigkeit, um den Widerspruch gegen die Nichtpaskalsche Geometrie zu beseitigen, oder in der Gastheorie Ausschaltung der Mechanik (Benutzung allein der Stossformel oder des Liouvilleschen Satzes) dafür Axiom der Wahrscheinlichkeit—(Entropie Maximum), um den Widerspruch gegen den Umkehr- oder Wiederkehreinwand zu beseitigen.>"

through their *Encyklopädie* article that appeared in 1912. Hilbert may have known the term earlier from their personal contact with them, or through some other colleague.[144] Also, the *Stossformel* that Hilbert mentioned here referred probably to the *Stossanzahlansatz*, whose specific role in the kinetic theory, together with that of the Liouville theorem (that is the physicists' Liouville theorem), the Ehrenfests' article definitely contributed to clarify.[145] Moreover, the clarification of the conceptual interrelation between Planck's quantum theory and electrodynamics—alluded to by Hilbert in his added remark—was also one of Paul Ehrenfest's central contributions to contemporary physics.[146]

7. THE AXIOMATIZATION PROGRAM BY 1905 – PARTIAL SUMMARY

Hilbert's 1905 cycle of lectures on the axiomatization of physics represents the culmination of a very central thread in Hilbert's early scientific career. This thread comprises a highly visible part of his published work, namely that associated with *Grundlagen der Geometrie*, but also additional elements that, though perhaps much less evident, were nevertheless prominent within his general view of mathematics, as we have seen. Hilbert's call in 1900 for the axiomatization of physical theories was a natural outgrowth of the background from which his axiomatic approach to geometry first developed. Although in elaborating the point of view put forward in the *Grundlagen der Geometrie* Hilbert was mainly driven by the need to solve certain, open foundational questions of geometry, his attention was also attracted in this context by recent debates on the role of axioms, or first principles in physics. Hertz's textbook on mechanics provided an elaborate example of a physical theory presented in strict axiomatic terms, and—perhaps more important for Hilbert—it also discussed in detail the kind of requirements that a satisfactory system of axioms for a physical theory must fulfill. Carl Neumann's analysis of the "Galilean principle of inertia"—echoes of which we find in Hilbert's own treatment of mechanics— provided a further example of the kind of conceptual clarity that one could expect to gain from this kind of treatment. The writings of Hilbert's senior colleague at Königsberg, Paul Volkmann, show that towards the end of the century questions of this kind were also discussed in the circles he moved in. Also the works of both Boltzmann and Voss provided Hilbert with important sources of information and inspiration. From his earliest attempts to treat geometry in an axiomatic fashion in order to solve the foundational questions he wanted to address in this field, Hilbert already had in mind the axiomatization of other physical disciplines as a task that could and should be pursued in similar terms.

144 Hilbert was most likely present when, on 13 November 1906, Paul Ehrenfest gave a lecture at the *Göttinger Mathematische Gesellschaft* on Boltzmann's H-theorem and some of the objections (*Einwände*) commonly raised against it. This lecture is reported in *Jahresbericht der Deutschen Mathematiker-Vereinigung*, Vol. 15 (1906), 593.

145 Cf. (Klein 1970, 119–140).

146 Cf. (Klein 1970, 230–257).

Between 1900 and 1905 Hilbert had the opportunity to learn much new physics. The lecture notes of his course provide the earliest encompassing evidence of Hilbert's own picture of physical science in general and, in particular, of how he thought the axiomatic analysis of individual theories should be carried out. Hilbert's physical interests now covered a broad range of issues, and he seems to have been well aware of the main open questions being investigated in most of the domains addressed. His unusual mathematical abilities allowed him to gain a quick grasp of existing knowledge, and at the same time to consider the various disciplines from his own idiosyncratic perspective, suggesting new interpretations and improved mathematical treatments. However, one must exercise great care when interpreting the contents of these notes. It is difficult to determine with exactitude the extent to which he had studied thoroughly and comprehensively all the existing literature on a topic he was pursuing. The relatively long bibliographical lists that we find in the introductions to many of his early courses do not necessarily mean that he studied all the works mentioned there. Even from his repeated, enthusiastic reference to Hertz's textbook we cannot safely infer to what extent he had read that book thoroughly. Very often throughout his career he was content when some colleague or student communicated to him the main ideas of a recent book or a new piece of research. In fact, the official assignment of many of his assistants—especially in the years to come—was precisely that: to keep him abreast of recent advances by studying in detail the research literature of a specific field. Hilbert would then, if he were actually interested, study the topic more thoroughly and develop his own ideas.

It is also important to qualify properly the extent to which Hilbert carried out a full axiomatic analysis of the physical theories he discussed. As we saw in the preceding sections, there is a considerable difference between what he did for geometry and what he did for other physical theories. In these lectures, Hilbert never actually proved the independence, consistency or completeness of the axiomatic systems he introduced. In certain cases, like vector addition, he quoted works in which such proofs could be found (significantly, works of his students or collaborators). In other cases there were no such works to mention, and—as in the case of thermodynamics—Hilbert simply *stated* that his axioms are indeed independent. In still other cases, he barely mentioned anything about independence or other properties of his axioms. Also, his derivations of the basic laws of the various disciplines from the axioms are rather sketchy, when they appear at all. Often, Hilbert simply declared that such a derivation was possible. What is clear is that Hilbert considered that an axiomatization along the lines he suggested was plausible and could eventually be fully performed following the standards established in *Grundlagen der Geometrie*.

Yet for all these qualifications, the lecture notes of 1905 present an intriguing picture of Hilbert's knowledge of physics, notable both for its breadth and its incisiveness. They afford a glimpse into a much less known side of his Göttingen teaching activity, which must certainly be taken into account in trying to understand the atmosphere that dominated this world center of science, as well as its widespread influence. More specifically, these notes illustrate in detail how Hilbert envisaged that axiomatic analysis of physical theories could not only contribute to conceptual clarification but also prepare the way for the improvement of theories, in the eventuality of future experimental evidence that conflicted with current predictions. If one knew in detail the logical structure of a given theory and the specific role of each of its basic assumptions, one could clear away possible contradictions and superfluous additional premises that may have accumulated in the building of the theory. At the same time, one would be prepared to implement, in an efficient and scientifically appropriate way, the local changes necessary to readapt the theory to meet the implications of newly discovered empirical data, in the eventuality of such discoveries. Indeed, Hilbert's own future research in physics, and in particular his incursion into general relativity, will be increasingly guided by this conception.

The nature and use of axioms in physical theories was discussed by many of Hilbert's contemporaries, as we have seen. Each had his own way of classifying the various kinds of axioms that are actually used or should be used. Hilbert himself did not discuss any possible such classification in detail but in his lectures we do find three different kinds of axioms actually implemented. This *de facto* classification is reminiscent, above all, of the one previously found in the writings of Volkmann. In the first place, every theory is assumed to be governed by specific axioms that characterize it and only it. These axioms usually express mathematical properties establishing relations among the basic magnitudes involved in the theory. Secondly, there are certain general mathematical principles that Hilbert saw as being valid for all physical theories. In the lectures he stressed above all the "continuity axiom," providing both a general formulation and more specific ones for each theory. As an additional general principle of this kind he suggested the assumption that all functions appearing in the natural sciences should have at least one continuous derivative. Furthermore, the universal validity of variational principles as the key to deriving the main equations of physics was a central underlying assumption of all of Hilbert's work on physics, and that kind of reasoning appears throughout these lectures as well. In each of the theories he considered in his 1905 lectures, Hilbert attempted to show how the exact analytic expression of a particular function that condenses the contents of the theory in question could be effectively derived from the specific axioms of the theory, together with more general principles. On some occasions he elaborated this idea more thoroughly, while on others he simply declared that such a derivation should be possible.

There is yet a third type of axiom for physical theories that Hilbert, however, avoided addressing in his 1905 lectures. That type comprises claims about the ultimate nature of physical phenomena, an issue that was particularly controversial during the years preceding these lectures. Although Hilbert's sympathy for the mechanical worldview is apparent throughout the manuscript of the lectures, his axiomatic analyses of physical theories contain no direct reference to it. The logical structure of the theories is thus intended to be fully understood independently of any particular position in this debate. Hilbert himself would later adopt a different stance. His work on general relativity will be based directly on his adoption of the electromagnetic worldview and, beginning in 1913, a quite specific version of it, namely, Gustav Mie's electromagnetic theory of matter. On the other hand, Hermann Minkowski's work on electrodynamics, with its seminal reinterpretation of Einstein's special theory of relativity in terms of spacetime geometry, should be understood as an instance of the kind of axiomatic analysis that Hilbert advanced in his 1905 lectures in which, at the same time, the debate between the mechanical and the electromagnetic worldviews is avoided.

When reading the manuscript of these lectures, one cannot help speculating about the reaction of the students who attended them. This was, after all, a regular course offered in Göttingen, rather than an advanced seminar. Before the astonished students stood the great Hilbert, rapidly surveying so many different physical theories, together with arithmetic, geometry and even logic, all in the framework of a single course. Hilbert moved from one theory to the other, and from one discipline to the next, without providing motivations or explaining the historical background to the specific topics addressed, without giving explicit references to the sources, without stopping to work out any particular idea, without proving any assertion in detail, but claiming all the while to possess a unified view of all these matters. The impression must have been thrilling, but perhaps the understanding he imparted to the students did not run very deep. Hermann Weyl's account of his experience as a young student attending Hilbert's course upon his arrival in Göttingen offers direct evidence to support this impression. Thus, in his obituary of Hilbert, Weyl wrote:

> In the fullness of my innocence and ignorance I made bold to take the course Hilbert had announced for that term, on the notion of number and the quadrature of the circle. Most of it went straight over my head. But the doors of a new world swung open for me, and I had not sat long at Hilbert's feet before the resolution formed itself in my young heart that I must by all means read and study what this man had written. (Weyl 1944, 614)

But the influence of the ideas discussed in Hilbert's course went certainly beyond the kind of general inspiration described here so vividly by Weyl; they had an actual influence on later contributions to physics. Besides the works of Born and Carathéodory on thermodynamics, and of Minkowski on electrodynamics, there were many dissertations written under Hilbert, as well as the articles written under the influence of his lectures and seminars. Ehrenfest's style of conceptual clarification of existing theories, especially as manifest in the famous *Encyklopädi*e on statistical mechanics, also bears the imprint of Hilbert's approach. Still, one can safely say that little work on physical theories was actually published along the specific lines of axiomatic analysis suggested by Hilbert in *Grundlagen der Geometrie*. It seems, in fact, that such techniques were never fully applied by Hilbert or by his students and collaborators to yield detailed analyses of axiomatic systems defining physical theories. Thus, for instance, in 1927 Georg Hamel wrote a long article on the axiomatization of mechanics for the *Handbuch der Physik* (Hamel 1927). Hamel did mention Hilbert's work on geometry as the model on which any modern axiomatic analysis should be based. However, his own detailed account of the axioms needed for defining mechanics as known at that time was not followed by an analysis of the independence of the axioms, based on the construction of partial models, such as Hilbert had carried out for geometry. Similarly, the question of consistency was discussed only summarily. Nevertheless, as Hamel said, his analysis allowed for a clearer comprehension of the logical structure of all the assumptions and their interdependence.

If the 1905 lectures represent the culmination of a thread in Hilbert's early career, they likewise constitute the beginning of the next stage of his association with physics. In the next years, Hilbert himself became increasingly involved in actual research in mathematical physics and he taught many courses on various topics thus far not included within his scientific horizons.

8. LECTURES ON MECHANICS AND CONTINUUM MECHANICS

In his early courses on mechanics or continuum mechanics, Hilbert's support for the atomistic hypothesis, as the possible basis for a reductionistic, mechanical foundation of the whole of physics, was often qualified by referring to the fact that the actual attempts to provide a detailed account of how such a reduction would work in specific cases for the various physical disciplines had not been fully and successfully realized by then. Thus for instance, in his 1906 course on continuum mechanics, Hilbert described the theory of elasticity as a discipline whose subject-matter is the deformation produced on solid bodies by interaction and displacement of molecules. On first sight this would seem to be a classical case in which one might expect a direct explanation based on atomistic considerations. Nevertheless Hilbert suggested that, for lack of detailed knowledge, a different approach should be followed in this case:

> We will have to give up going here into a detailed description of these molecular processes. Rather, we will only look for those parameters on which the measurable deformation state of the body depends at each location. The form of the dependence of the Lagrangian function on these parameters will then be determined, which is actually composed by the kinetic and potential energy of the individual molecules. Similarly, in thermodynamics we will not go into the vibrations of the molecules, but we will rather introduce temperature itself as a general parameter and we will investigate the dependence of energy on it.[147]

The task of deducing the exact form of the Lagrangian under specific requirements postulated as part of the theory was the approach followed in the many examples already discussed above. This tension between reductionistic and phenomenological explanations in physics is found in Hilbert's physical ideas throughout the years and it eventually led to his abandonment of mechanical reduc-

147 "Wir werden hier auf eine eingehende Beschreibung dieser molekularen Vorgänge zu verzichten haben und dafür nur die Parameter aufsuchen, von denen der meßbare Verzerrungszustand der Körper an jeder Stelle abhängt. Alsdann wird festzustellen sein, wie die Form der Abhängigkeit der Lagranschen Funktion von diesen Parametern ist, die sich ja eigentlich aus kinetischer und potentieller Energie der einzelnen Molekel zusammensetzen wird. Ähnlich wird man in der Thermodynamik nicht auf die Schwingungen der Molekel eingehen, sondern die Temperatur selbst als allgemeinen Parameter einführen, und die Abhängigkeit der Energie von ihr untersuchen." (Hilbert 1906, 8–9)

tionism. The process becomes gradually manifest after 1910, though Hilbert still stuck to his original conceptions until around 1913.

The course on mechanics in the winter semester of 1910–1911 opened with an unambiguous statement about the essential role of mechanics as the foundation of natural science in general (Hilbert 1910–1911, 6). Hilbert praised the textbooks of Hertz and Boltzmann for their successful attempts to present in similar methodological terms, albeit starting from somewhat different premises, a fully axiomatic derivation of mechanics. This kind of presentation, Hilbert added, was currently being disputed. The course itself covered the standard topics of classical mechanics. Towards the end, however, Hilbert spoke about the "new mechanics." In this context he neither used the word "relativity" nor mentioned Einstein. Rather, he mentioned only Lorentz and spoke of invariance under the Lorentz transformations of all differential equations that describe natural phenomena as the main feature of this new mechanics. Hilbert stressed that the Newtonian equations of the "old" mechanics do not satisfy this basic principle, which, like Minkowski, he called the *Weltpostulat*. These equations must therefore be transformed, he said, so that they become Lorentz-invariant.[148] Hilbert showed that if the Lorentz transformations are used instead of the "Newton transformations," then the velocity of light is the same for every non-accelerated, moving system of reference.

Hilbert also mentioned the unsettled question of the status of gravitation in the framework of this new mechanics. He connected his presentation directly to Minkowski's sketchy treatment of this topic in 1909, and, like his friend, Hilbert does not seem to have been really bothered by the difficulties related with it. One should attempt to modify the Newtonian law in order to make it comply to the world-postulate, Hilbert said, but we must exercise special care when doing this since the Newtonian law has proved to be in the closest accordance with experience. As Hilbert knew from Minkowski's work, an adaptation of gravitation to the new mechanics would imply that its effects must propagate at the speed of light. This latter conclusion contradicts the "old theory," while in the framework of the "new mechanics," on the contrary, it finds a natural place. In order to adapt the Newtonian equations to the new mechanics, concluded Hilbert, we proceed, "as Minkowski did, via electromagnetism."[149]

The manuscript of the course does not record whether in the classroom Hilbert showed how, by proceeding "as Minkowski did, via electromagnetism," the adaptation of Newton's law should actually be realized. Perhaps at that time he still believed that Minkowski's early sketch could be further elaborated. Be that as it may, the concerns expressed here by Hilbert are not unlike those of other, contemporary physicists involved in investigating the actual place of the postulate of relativity in the general

148 "Alle grundlegenden Naturgesetzen entsprechenden Systeme von Differentialgleichungen sollen gegenüber der Lorentz-Transformation kovariant sein. ... Wir können durch Beobachtung von irgend welchen Naturvorgängen niemals entscheiden, ob wir ruhen, oder uns gleichförmig bewegen. Diesem Weltpostulate genügen die Newtonschen Gleichungen der älteren Mechanik nicht, wenn wir die Lorentz Transformation zugrunde legen: wir stehen daher vor der Aufgabe, sie dementsprechend umzugestalten." (Hilbert 1910–1911, 292)

picture of physics. It is relevant to recall at this stage, however, that Einstein himself published nothing on this topic between 1907 and June 1911.

9. KINETIC THEORY

After another standard course on continuum mechanics in the summer of 1911, Hilbert taught a course specifically devoted to kinetic theory of gases for the first time in the winter of 1911–1912. This course marked the starting point of Hilbert's definitive involvement with a broader range of physical theories. Hilbert opened the course by referring once again to three possible, alternative treatments of any physical theory. First, is the "phenomenological perspective,"[150] often applied to study the mechanics of continua. Under this perspective, the whole of physics is divided into various chapters, each of which can be approached using different, specific assumptions, from which different mathematical consequences can be derived. The main mathematical tool used in this approach is the theory of partial differential equations. In fact, much of what Hilbert had done in his 1905 lectures on the axiomatization of physics, and then in 1906 on mechanics of continua, could be said to fall within this approach.

The second approach that Hilbert mentioned assumes the validity of the "theory of atoms." In this case a "much deeper understanding is reached. ... We attempt to put forward a system of axioms which is valid for the whole of physics, and which enables all physical phenomena to be explained from a unified point of view."[151] The mathematical methods used here are obviously quite different from those of the phenomenological approach: they can be subsumed, generally speaking, under the methods of the theory of probabilities. The most salient examples of this approach are found in the theory of gases and in radiation theory. From the point of view of this approach, the phenomenological one is a palliative, indispensable as a primitive stage on the way to knowledge, which must however be abandoned "as soon as possible, in order to penetrate the real sanctuary of theoretical physics."[152] Unfortunately, Hilbert

149 "Wir können nun an die Umgestaltung des Newtonsches Gesetzes gehen, dabei müssen wir aber Vorsicht verfahren, denn das Newtonsche Gesetz ist das desjenige Naturgesetz, das durch die Erfahrung in Einklang bleiben wollen. Dieses wird uns gelingen, ja noch mehr, wir können verlangen, dass die Gravitation sich mit Lichtgeschwindigkeit fortpflanzt. Die alte Theorie kann das nicht, eine Fortpflanzung der Gravitation mit Lichtgeschwindigkeit widerspricht hier der Erfahrung: Die neue Theorie kann es, und man ist berechtigt, das als eine Vorzug derselben anzusehen, den eine momentane Fortpflanzung der Gravitation passt sehr wenig zu der modernen Physik. Um die Newtonschen Gleichungen für die neue Mechanik zu erhalten, gehen wir ähnlich vor wie Minkowski in der Elektromagnetik." (Hilbert 1910–1911, 295)

150 Boltzmann had used the term in this context in his 1899 Munich talk that Hilbert had attended. Cf. (Boltzmann 1899, 92–96).

151 "Hier ist das Bestreben, ein Axiomensystem zu schaffen, welches für die ganze Physik gilt, und aus diesem einheitlichen Gesichtspunkt alle Erscheinungen zu erklären. ... Jedenfalls gibt sie unvergleichlich tieferen Laufschuhes über Wesen und Zusammenhang der physikalischen Begriffe, ausserdem auch neue Aufklärung über physikalische Tatsachen, welche weit über die bei A) erhaltene hinausgeht." (Hilbert 1911–1912, 2)

said, mathematical analysis is not yet developed sufficiently to provide for all the demands of the second approach. One must therefore do without rigorous logical deductions and be temporarily satisfied with rather vague mathematical formulae.[153] Hilbert considered it remarkable that by using this method one nevertheless obtains ever new results that are in accordance with experience. He thus declared that the "main task of physics," embodied in the third possible approach, would be "the molecular theory of matter" itself, standing above the kinetic theory, as far as its degree of mathematical sophistication and exactitude is concerned. In the present course, Hilbert intended to concentrate on kinetic theory, yet he promised to consider the molecular theory of matter in the following semester. He did so, indeed, a year later.

Many of the important innovations implied by Hilbert's solution of the Boltzmann equation are already contained in this course of 1911–1912.[154] It was Maxwell in 1860 who first formulated an equation describing the distribution of the number of molecules of a gas, with given energy at a given point in time. Maxwell, however, was able to find only a partial solution which was valid only for a very special case.[155] In 1872 Boltzmann reformulated Maxwell's equation in terms of a single, rather complex, integro-differential equation, that has remained associated with his name ever since. The only exact solution Boltzmann had been able to find, however, was still valid for the same particular case that Maxwell had treated in his own model (Boltzmann 1872). By 1911, some progress had been made on the solution of the Boltzmann equation. The laws obtained from the partial knowledge concerning those solutions, which described the macroscopic movement and thermal processes in gases, seemed to be qualitatively correct. However, the mathematical methods used in the derivations seemed inconclusive and sometimes arbitrary. It was quite usual to rely on average magnitudes and thus the calculated values of the coefficients of heat conduction and friction appeared to be dubious. A more accurate estimation of these values remained a main concern of the theory, and the techniques developed by Hilbert apparently offered the means to deal with it.[156]

Very much as he had done with other theories in the past, Hilbert wanted to show how the whole kinetic theory could be developed starting from one basic formula, which in this case would be precisely the Boltzmann equation. His presentation would depart from the phenomenological approach by making some specific assumptions about the molecules, namely that they are spheres identical to one another in

---

152 "Wenn man auf diesem Standpunkt steht, so wird man den früheren nur als einer Notbehelf bezeichnen, der nötig ist als eine erste Stufe der Erkenntnis, über die man aber eilig hinwegschreiten muss, um in die eigentlichen Heiligtümer der theoretischen Physik einzudringen." (Hilbert 1911–1912, 2)

153 "... sich mit etwas verschwommenen mathematischen Formulierungen zufrieden geben muss." (Hilbert 1911–1912, 2)

154 In fact, in December 1911 Hilbert presented to the *Göttinger Mathematische Gesellschaft* an overview of his recent investigations on the theory, stating that he intended to publish them soon. Cf. *Jahresbericht der Deutschen Mathematiker-Vereinigung* 21 (1912), 58.

155 Cf. (Brush 1976, 432–446).

156 Cf. (Born 1922, 587–589).

size. In addition he would focus, not on the velocity of any individual such molecule, but rather on their velocity distribution φ over a small element of volume.

In the opening lectures of the course, a rather straightforward discussion of the elementary physical properties of a gas led Hilbert to formulate a quite complicated equation involving φ. Hilbert asserted that a general solution of this equation was impossible, and it was thus necessary to limit the discussion to certain specific cases (Hilbert 1911–1912, 21). In the following lectures he added some specific, physical assumptions concerning the initial and boundary conditions for the velocity distribution in order to be able to derive more directly solvable equations. These assumptions, which he formulated as axioms of the theory, restricted the generality of the problem to a certain extent, but allowed for representing the distribution function as a series of powers of a certain parameter. In a first approximation, the relations between the velocity distributions yielded the Boltzmann distribution. In a second approximation, they yielded the propagation of the average velocities in space and time. Under this representation the equation appeared as a linear symmetric equation of the second type, where the velocity distribution φ is the unknown function, thus allowing the application of Hilbert's newly developed techniques. Still, he did not prove in detail the convergence of the power series so defined, nor did he complete the evaluation of the transport coefficient appearing in the distribution formula.

Hilbert was evidently satisfied with his achievement in kinetic theory. He was very explicit in claiming that without a direct application of the techniques he had developed in the theory of integral equations, and without having formulated the physical theory in terms of such integral equations, it would be impossible to provide a solid and systematic foundation for the theory of gases as currently known (Hilbert 1912a, 268; 1912b, 562). And very much as with his more purely mathematical works, also here Hilbert was after a larger picture, searching for the underlying connections among apparently distant fields. Particularly interesting for him were the multiple connections with radiation theory, which he explicitly mentioned at the end of his 1912 article, thus opening the way for his forthcoming courses and publications. In his first publication on radiation theory he explained in greater detail and with unconcealed effusiveness the nature of this underlying connection. He thus said:

> In my treatise on the "Foundations of the kinetic theory of gases," I have shown, using the theory of linear integral equations, that starting alone from the Maxwell-Boltzmann fundamental formula —the so-called collision formula— it is possible to construct systematically the kinetic theory of gases. This construction is such, that it requires only a consistent implementation of the methods of certain mathematical operations prescribed in advance, in order to obtain the proof of the second law of thermodynamics, of Boltzmann's expression for the entropy of a gas, of the equations of motion that take into account both the internal friction and the heat conduction, and of the theory of diffusion of several gases. Likewise, by further developing the theory, we obtain the precise conditions under which the law of equipartition of energies over the intermolecular parameter is valid. Concerning the motion of compound molecules, a new law is also obtained according to which the continuity equation of hydrodynamics has a much more general meaning than the usual one. ...

> Meanwhile, there is a second physical domain whose principles have not yet been investigated at all from the mathematical point of view, and for the establishment of whose foundations—as I have recently discovered—the same mathematical tools provided by the integral equations are absolutely necessary. I mean by this the elementary theory of radiation, understanding by it the phenomenological aspect of the theory, which at the most immediate level concerns the phenomena of emission and absorption, and on top of which stand Kirchhoff's laws concerning the relations between emission and absorption. (Hilbert 1912b, 217–218)

Hilbert could boast now two powerful mathematical tools that allowed him to address the study of a broad spectrum of physical theories. On the one hand, the axiomatic method would help dispel conceptual difficulties affecting established theories—thus fostering their continued development—and also open the way for a healthy establishment of new ones. In his earlier courses he had already explored examples of the value of the method for a wide variety of disciplines, but Minkowski's contributions to electrodynamics and his analysis of the role of the principle of relativity offered perhaps, from Hilbert's point of view, the most significant example so far of the actual realization of its potential contribution. On the other hand, the theory of linear integral equations had just proven its value in the solution of such a central, open problem of physics. As far as he could see from his own, idiosyncratic perspective, the program for closing the gap between physical theories and mathematics had been more successful so far than he may have actually conceived when posing his sixth problem back in 1900. Hilbert was now prepared to attack yet another central field of physics and he would do so by combining once again the two mathematical components of his approach. The actual realization of this plan, however, was less smooth than one could guess from the above-quoted, somewhat pompous, declaration. As will be seen in the next section, although Hilbert's next incursion into the physicists' camp led to some local successes, as a whole they were less impressive in their overall significance than Hilbert would have hoped.

But even though Hilbert was satisfied with what his mastery of integral equations had allowed him to do thus far, and with what his usual optimism promised to achieve in other physical domains in the near future, there was an underlying fundamental uneasiness that he was not able to conceal behind the complex integral formulas and he preferred to explicitly share this uneasiness with his students. It concerned the possible justification of using probabilistic methods in physics in general and in kinetic theory in particular. Hilbert's qualms are worth quoting in some detail:

> If Boltzmann proves … that the Maxwell distribution … is the most probable one from among all distributions for a given amount of energy, this theorem possesses in itself a certain degree of interest, but it does not allow even a minimal inference concerning the velocity distribution that actually occurs in any given gas. In order to lay bare the core of this question, I want to recount the following example: in a raffle with one winner out of 1000 tickets, we distribute 998 tickets among 998 persons and the remaining two we give to a single person. This person thus has the greatest chance to win, compared to all other participants. His probability of winning is the greatest, and yet it is highly improbable that he will win. The probability of this is close to zero. In the same fashion, the probability of occurrence of the Maxwell velocity distribution is greater than that of any other

> distribution, but equally close to zero, and it is therefore almost absolutely certain that the Maxwell distribution will not occur.
>
> What is needed for the theory of gases is much more than that. We would like to prove that for a specified distribution, there is a probability very close to 1 that distribution is asymptotically approached as the number of molecules becomes infinitely large. And in order to achieve that, it is necessary to modify the concept of "velocity distribution" in order to obtain some margin for looseness. We should formulate the question in terms such as these: What is the probability for the occurrence of a velocity distribution that deviates from Maxwell's by no more than a given amount? And moreover: what allowed deviation must we choose in order to obtain the probability 1 in the limit?[157]

Hilbert discussed in some detail additional difficulties that arise in applying probabilistic reasoning within kinetic theory. He also gave a rough sketch of the kind of mathematical considerations that could in principle provide a way out to the dilemmas indicated. Yet he made clear that he could not give final answers in this regard.[158] This problem would continue to bother him in the near future. In any case, after this brief excursus, Hilbert continued with the discussion he had started in the first part of his lectures and went on to generalize the solutions already obtained to the cases of mixtures of gases or of polyatomic gases.

In spite of its very high level of technical sophistication of his approach to kinetic theory, it is clear that Hilbert did not want his contribution to be seen as a purely mathematical, if major, addition to the solution of just one central, open problem of

---

157 "Wenn z.B. Boltzmann beweist—übrigens auch mit einigen Vernachlässigungen—dass die Maxwellsche Verteilung (die nach dem Exponentialgesetz) unter allen Verteilungen von gegebener Gesamtenergie die wahrscheinlichste ist, so besitzt dieser Satz ja an und für sich ein gewisses Interesse, aber er gestattet auch nicht der geringsten Schluss auf die Geschwindigkeitsverteilung, welche in einem bestimmten Gase wirklich eintritt. Um den Kernpunkt der Frage klar zu legen, will ich an folgendes Beispiel erinnern: In einer Lotterie mit einem Gewinn und von 1000 Losen seien 998 Losen auf 998 Personen verteilt, die zwei übrigen Lose möge eine andere Person erhalten. Dann hat diese Person im Vergleich zu jeder einzelnen andern die grössten Gewinnchancen. Die Wahrscheinlichkeit des Gewinnen ist für sie am grössten, aber es ist immer noch höchst unwahrscheinlich, dass sie gewinnt. Denn die Wahrscheinlichkeit ist so gut wie Null.
Ganz ebenso ist die Wahrscheinlichkeit für den Eintritt der Maxwellschen Geschwindigkeitsverteilung zwar grösser als die für das Eintreten einer jeden bestimmten andern, aber doch noch so gut wie Null, und es ist daher fast mit absoluter Gewissheit sicher, dass die Maxwellsche Verteilung nicht eintritt.
Was wir für die Gastheorie brauchen, ist sehr viel mehr. Wir wünschen zu beweisen, dass für eine gewisse ausgezeichnete Verteilung eine Wahrscheinlichkeit sehr nahe an 1 besteht, derart, dass sie sich mit Unendliche wachsende Molekülzahl der 1 asymptotisch annähert. Und um das zu erreichen, müssen wir den Begriff der „Geschwindigkeitsverteilung" etwas modifizieren, indem wir einen gewissen Spielraum zulassen. Wir hätten die Frage etwa so zu formulieren: Wie gross ist die Wahrscheinlichkeit dafür, dass eine Geschwindigkeitsverteilung eintritt, welche von der Maxwellschen nur um höchstens einen bestimmten Betrag abweicht—und weiter: wie gross müssen wir die zugelassenen Abweichungen wählen, damit wir im limes die Wahrscheinlichkeit eins erhalten?" (Hilbert 1911–1912, 75–76)

158 "Ich will Ihnen nun auseinandersetzen, wie ich mir etwa die Behandlung dieser Frage denke. Es sind da sicher noch grosse Schwierigkeiten zu überwinden, aber die Idee nach wird man wohl in folgender Weise vorgehen müssen: ... " (Hilbert 1911–1912, 77)

this theory. Rather, his aim was to be directly in touch with the physical core of this and other, related domains. The actual scope of his physical interests at the time becomes more clearly evident in a seminar that he organized in collaboration with Erich Hecke (1887–1947), shortly after the publication of his article on kinetic theory.[159] The seminar was also attended by the Göttingen docents Max Born, Paul Hertz, Theodor von Kármán (1881–1963), and Erwin Madelung (1881–1972), and the issues discussed included the following:[160]

- the ergodic hypothesis and its consequences;
- on Brownian motion and its theories;
- electron theory of metals in analogy to Hilbert's theory of gases;
- report on Hilbert's theory of gases;
- on dilute gases;
- theory of dilute gases using Hilbert's theory;
- on the theory of chemical equilibrium, including a reference to the
- related work of Sackur;
- dilute solutions.

The names of the participants and younger colleagues indicate that these deep physical issues, related indeed with kinetic theory but mostly not with its purely mathematical aspects, could not have been discussed only superficially. Especially indicative of Hilbert's surprisingly broad spectrum of interests is the reference to the work of Otto Sackur (1880–1914). Sackur was a physical chemist from Breslau whose work dealt mainly with the laws of chemical equilibrium in ideal gases and on Nernst law of heat. He also wrote a widely used textbook on thermochemistry and thermodynamics (Sackur 1912). His experimental work was also of considerable significance and, more generally, his work was far from the typical kind of purely technical, formal mathematical physics that is sometimes associated with Hilbert and the Göttingen school.[161]

10. RADIATION THEORY

Already in his 1911–1912 lectures on kinetic theory, Hilbert had made clear his interest in investigating, together with this domain and following a similar approach, the

159 Hecke had also taken the notes of the 1911–1912 course.

160 References to this seminar appear in (Lorey 1916, 129). Lorey took this information from the German student's journal *Semesterberichte des Mathematischen Verereins*. The exact date of the seminar, however, is not explicitly stated.

161 See Sackur's obituary in *Physikalische Zeitschrift* 16 (1915), 113–115. According to Reid's account (1970, 129), Ewald succinctly described Hilbert's scientific program at the time of his arrival in Göttingen with the following, alleged quotation of the latter: "We have reformed mathematics, the next thing to reform is physics, and then we'll go on to chemistry." Interest in Sackur's work, as instantiated in this seminar would be an example of an intended, prospective attack on this field. There are not, however, many documented, further instances of this kind.

theory of radiation.[162] Kirchhoff's laws of emission and absorption had traditionally stood as the focus of interest of this theory. These laws, originally formulated in late 1859, describe the energetic relations of radiation in a state of thermal equilibrium.[163] They assert that in the case of purely thermal radiation (i.e., radiation produced by thermal excitation of the molecules) the ratio between the emission and absorption capacities of matter, η and α respectively, is a universal function of the temperature T and the wavelength λ,

$$\frac{\eta}{\alpha} = K(T, \lambda)$$

and is therefore independent of the substance and of any other characteristics of the body in question. One special case that Kirchhoff considered in his investigations is the case $\alpha = 1$, which defines a "black body," namely, a hypothetical entity that completely absorbs all wavelengths of thermal radiation incident in it.[164]

In the original conception of Kirchhoff's theory the study of black-body radiation may not have appeared as its most important open problem, but in retrospect it turned out to have the farthest-reaching implications for the development of physics at large. In its initial phases, several physicists attempted to determine over the last decades of the century the exact form of the spectral distribution of the radiation $K(T, \lambda)$ for a black body. Prominent among them was Wilhelm Wien, who approached the problem by treating this kind of radiation as loosely analogous to gas molecules. In 1896 he formulated a law of radiation that predicted very accurately recent existing measurements. Planck, however, was dissatisfied with the lack of a theoretical justification for what seemed to be an empirically correct law. In searching for such a justification within classical electromagnetism and thermodynamics, he modeled the atoms at the inside walls of a black-body cavity as a collection of electrical oscillators which absorbed and emitted energy at all frequencies. In 1899 he came forward with an expression for the entropy of an ideal oscillator, built on an analogy with Boltzmann's kinetic theory of gases, that provided the desired theoretical justification of Wien's law (Planck 1899). Later on, however, additional experiments produced values for the spectrum at very low temperatures and at long wavelengths that were not anymore in agreement with this law.

Another classical attempt was advanced by John William Strutt, Lord Rayleigh (1842–1919), and James Jeans (1877–1946), also at the beginning of the century.[165] Considering the radiation within the black-body cavity to be made up of a series of standing waves, they derived a law that, contrary to Wien's, approximated experimen-

162 Minkowski and Hilbert even had planned to have a seminar on the theory of heat radiation as early as 1907 (Minkowski 1907).
163 Cf., e.g., (Kirchhoff 1860).
164 Cf. (Kuhn 1978, 3–10).
165 Cf. (Kuhn 1978, 144–152).

tal data very well at long wavelengths but failed at short ones. In the latter case, it predicted that the spectrum would rise to infinity as the wavelength decreased to zero.[166]

In a seminal paper of 1900, Planck formulated an improved law that approximated Wien's formula in the case of short wavelengths and the Rayleigh-Jeans law in the case of long wavelengths. The law assumed that the resonator entropy is calculated by counting the number of distributions of a given number of finite, equal "energy elements" over a set of resonators, according to the formula:

$$E = nh\nu$$

where n is an integer, ν is the oscillators' frequency, and h is the now famous Planck constant, h = 6.55 x 10 -27 erg-sec. (Planck 1900). Based on this introduction of energy elements, assuming thermal equilibrium and applying statistical methods of kinetic theory, Planck derived the law that he had previously obtained empirically and that described the radiant energy distribution of the oscillators:

$$U_\nu = \frac{h\nu}{e^{h\nu/kT} - 1}.$$

Planck saw his assumption of energy elements as a convenient mathematical hypothesis, and not as a truly physical claim about the way in which matter and radiation actually interchange energy. In particular, he did not stress the significance of the finite energy elements that entered his calculation and he continued to think about the resonators in terms of a continuous dynamics. He considered his assumption to be very important since it led with high accuracy to a law that had been repeatedly confirmed at the experimental level, but at the same time he considered it to be a provisional one that would be removed in future formulations of the theory. In spite of its eventual revolutionary implications on the developments of physics, Planck did not realize before 1908 that his assumptions entailed any significant departure from the fundamental conceptions embodied in classical physics. As a matter of fact, he did not publish any further research on black-body radiation between 1901 and 1906.[167]

The fundamental idea of the quantum discontinuity was only slowly absorbed into physics, first through the works of younger physicists such as Einstein, Laue and Ehrenfest, then by leading ones such as Planck, Wien and Lorentz, and finally by their readers and followers. The details pertaining to this complex process are well beyond the scope of my account here. Nonetheless, it is worth mentioning that a very significant factor influencing Planck's own views in this regard was his correspondence with Lorentz in 1908. Lorentz had followed with interest since 1901 the debates around black-body radiation, and he made some effort to connect them with his own theory of the electron. At the International Congress of Mathematicians held

166 Much later Ehrenfest (1911) dubbed this phenomenon "ultraviolet catastrophe."

167 This is the main claim developed in detail in the now classic (Kuhn 1978). For a more recent, summary account of the rise of quantum theory, see (Kragh 1999, chap. 5).

in Rome in 1908, Lorentz was invited to deliver one of the plenary talks, which he devoted to this topic. This lecture was widely circulated and read thereafter and it represented one of the last attempts at interpreting cavity radiation in terms of a classical approach (Lorentz 1909). But then, following critical remarks by several colleagues, Lorentz added a note to the printed version of his talk where he acknowledged that his attempt to derive the old Rayleigh-Jeans radiation law from electron theory was impracticable unless the foundations of the latter would be deeply modified. A letter to Lorentz sent by Planck in the aftermath of the publication contains what may be the latter's first acknowledgment of the need to introduce discontinuity as a fundamental assumption. Lorentz himself, at any rate, now unambiguously adopted the idea of energy quanta and he stressed it explicitly in his lectures of early 1909 in Utrecht.[168] Later, in his 1910 Wolfskehl cycle in Göttingen, Lorentz devoted one of the lectures to explaining why the classical Hamilton principle would not work for radiation theory. An "entirely new hypothesis," he said, needed to be introduced. The new hypothesis he had in mind was "the introduction of the energy elements invented by Planck" (Lorentz 1910, 1248). Hilbert was of course in the audience and he must have attentively listened to his guest explaining the innovation implied by this fundamental assertion.

Starting in 1911 research on black-body radiation became less and less prominent and at the same time the quantum discontinuity hypothesis became a central issue in other domains such as thermodynamics, specific heats, x-rays, and atomic models. The apparent conflicts between classical physics and the consequences of the hypothesis stood at the focus of discussions in the first Solvay conference organized in Brussels in 1911.[169] These discussion prompted Poincaré, who until then was reticent to adopt the discontinuity hypothesis, to elaborate a mathematical proof that Planck's radiation law necessarily required the introduction of quanta (Poincaré 1912). His proof also succeeded in convincing Jeans in 1913, who thus became one of the latest prominent physicists to abandon the classical conception in favor of discontinuity (Jeans 1914).[170]

The notes of Hilbert's course on radiation theory in the summer semester of 1912, starting in late April, evince a clear understanding and a very broad knowledge of all the main issues of the discipline. In his previous course on kinetic theory, Hilbert had promised to address "the main task of physics," namely, the molecular theory of matter itself, a theory he described as having a greater degree of mathematical sophistication and exactitude than kinetic theory. To a certain extent, teaching this course meant fulfilling that promise. Hilbert declared that he intended to address now the "domain of physics properly said," which is based on the point of view of the atomic theory. Hilbert was clearly very much impressed by recent developments in quantum theory. "Never has there been a more proper and challenging time than now," he said, "to

168 Cf. (Kuhn 1978, 189–197).
169 Cf. (Barkan 1993).
170 Cf. (Kuhn 1978, 206–232).

undertake the research of the foundations of physics." What seems to have impressed Hilbert more than anything else were the deep interconnections recently discovered in physics, "of which formerly no one could have even dreamed, namely, that optics is nothing but a chapter within the theory of electricity, that electrodynamics and thermodynamics are one and the same, that energy also possesses inertial properties, that physical methods have been introduced into chemistry as well."[171] And above all, the "atomic theory," the "principle of discontinuity," which was not a hypothesis anymore, but rather, "like Copernicus's theory, a fact confirmed by experiment."[172]

Hilbert opened with a summary account of four-vector analysis[173] and of Special Theory of Relativity. Taking the relativity postulate to stand "on top" of physics as a whole, he then formulated the basics of electrodynamics as currently conceived, including Born's concept of a rigid body. This is perhaps Hilbert's first systematic discussion of Special Theory of Relativity in his lecture courses. As in the case of kinetic theory, Hilbert already raised here some of the ideas that he would later develop in his related, published works. But again, the course was far from being just an exercise in applying integral equations techniques to a particularly interesting, physical case. Rather, Hilbert covered most of the core, directly relevant, physical questions. Thus, among the topics discussed in the course we find the energy distribution of black-body radiation (including a discussion of Wien's and Rayleigh's laws) and Planck's theory of resonators under the effect of radiation. Hilbert particularly stressed the significance of recent works by Ehrenfest and Poincaré, as having shown the necessity of a discontinuous form of energy distribution (Hilbert 1912c, 94).[174] Hilbert also made special efforts to have Sommerfeld invited to give the last two lectures in the course, in which important, recent topics in the theory were discussed.[175]

However, as with all other physical theories, what Hilbert considered to be the main issue of the theory of radiation as a whole was the determination of the precise form of a specific law that stood at its core. In this case the law in question was Kirchhoff's law of emission and absorption, to which Hilbert devoted several lectures. Of particular interest for him was the possibility of using the techniques of the

171 "Nun kommen wir aber zu eigentlicher Physik, welche sich auf der Standpunkt der Atomistik stellt und da kann man sagen, dass keine Zeit günstiger ist und keine mehr dazu herausfordert, die Grundlagen dieser Disziplin zu untersuchen, wie die heutige. Zunächst wegen der Zusammenhänge, die man heute in der Physik entdeckt hat, wovon man sich früher nichts hätte träumen lassen, dass die Optik nur ein Kapitel der Elektrizitätslehre ist, dass Elektrodynamik und Thermodynamik dasselbe sind, dass auch die Energie Trägheit besitzt, dann dass auch in der Chemie (Metalchemie, Radioaktivität) physikalische Methoden in der Vordergrund haben." (Hilbert 1912c, 2)

172 "... wie die Lehre des Kopernikus, eine durch das Experimente bewiesene Tatsache." (Hilbert 1912c, 2)

173 A hand-written addition to the typescript (Hilbert 1912c, 4) gives here a cross-reference to Hilbert's later course, (Hilbert 1916, 45–56), where the same topic is discussed in greater detail.

174 He referred to (Ehrenfest 1911) and (Poincaré 1912). Hilbert had recently asked Poincaré for a reprint of his article. See Hilbert to Poincaré, 6 May 1912. (Hilbert 1932–1935, 546)

175 Cf. Hilbert to Sommerfeld, 5 April 1912 (*Nachlass* Arnold Sommerfeld, Deutsches Museum, Munich. HS1977–28/A, 141).

theory of integral equations for studying the foundations of the law and providing a complete mathematical justification for it. This would also become the main task pursued in his published articles on the topic, which I discuss in detail in the next four sections. In fact, just as his summer semester course was coming to a conclusion, Hilbert submitted for publication his first paper on the "Foundations of the Elementary Theory of Radiation."

11. STRUCTURE OF MATTER AND RELATIVITY: 1912–1914

After this account of Hilbert's involvement with kinetic theory and radiation theory, I return to 1912 in order to examine his courses in physics during the next two years.[176] The structure of matter was the focus of attention here, and Hilbert now finally came to adopt electromagnetism as the fundamental kind of phenomenon to which all others should be reduced. The atomistic hypothesis was a main physical assumption underlying all of Hilbert's work from very early on, and also in the period that started in 1910. This hypothesis, however, was for him secondary to more basic, mathematical considerations of simplicity and precision. A main justification for the belief in the validity of the hypothesis was the prospect that it would provide a more accurate and detailed explanation of natural phenomena once the tools were developed for a comprehensive mathematical treatment of theories based on it. Already in his 1905 lectures on the axiomatization of physics, Hilbert had stressed the problems implied by the combined application of analysis and the calculus of probabilities as the basis for the kinetic theory, an application that is not fully justified on mathematical grounds. In his physical courses after 1910, as we have seen, he again expressed similar concerns. The more Hilbert became involved with the study of kinetic theory itself, and at the same time with the deep mathematical intricacies of the theory of linear integral equations, the more these concerns increased. This situation, together with his growing mastery of specific physical issues from diverse disciplines, helps to explain Hilbert's mounting interest in questions related to the structure of matter that occupied him in the period I discuss now. The courses described below cover a wide range of interesting physical questions. In this account, for reasons of space, I will comment only on those aspects that are more directly connected with the questions of axiomatization, reductionism and the structure of matter.

11.1 Molecular Theory of Matter - 1912–1913

Hilbert's physics course in the winter semester of 1912–1913 was devoted to describing the current state of development of the molecular theory of matter (Hilbert 1912–

176 The printed version of the *Verzeichnis der Vorlesungen an der Georg-August-Universität zu Göttingen* registers several courses for which no notes or similar documents are extant, and about which I can say nothing here: summer semester, 1912 - Mathematical Foundations of Physics; winter semester, 1912–1913 - Mathematical Foundations of Physics.

1913),[177] and particularly the behavior of systems of huge quantities of particles moving in space, and affecting each other through collisions and other kinds of interacting forces.[178] The first of the course's three parts deals with the equation of state, including a section on the principles of statistical mechanics. The second part is characterized as "phenomenological" and the third part as "kinetic," in which entropy and the quantum hypothesis are discussed. This third part also includes a list of axioms for the molecular theory of matter. Hilbert was thus closing a circle initiated with the course on kinetic theory taught one year earlier.

Hilbert suggested that the correct way to come to terms with the increasingly deep mathematical difficulties implied by the atomistic hypothesis would be to adopt a "physical point of view." This means that one should make clear, through the use of the axiomatic method, those places in which physics intervenes into mathematical deduction. This would allow separating three different components in any specific physical domain considered: first, what is arbitrarily adopted as definition or taken as an assumption of experience; second, what a-priori expectations follow from these assumptions, which the current state of mathematics does not yet allow us to conclude with certainty; and third, what is truly proven from a mathematical point of view.[179] This separation interestingly brings to mind Minkowski's earlier discussion on the status of the principle of relativity. It also reflects to a large extent the various levels of discussion evident in Hilbert's articles on radiation theory, and it will resurface in his reconsideration of the view of mechanics as the ultimate explanation of physical phenomena.

In the first part of the course, Hilbert deduced the relations between pressure, volume and temperature for a completely homogenous body. He considered the body as a mechanical system composed of molecules, and applied to it the standard laws of mechanics. This is a relatively simple case, he said, that can be easily and thoroughly elucidated. However, deriving the state equation and explaining the phenomenon of condensation covers only a very reduced portion of the empirically manifest properties of matter. Thus the second part of the lectures was devoted to presenting certain, more complex physical and chemical phenomena, the kinetic significance of which would then be explained in the third part of the course.[180] The underlying approach

---

177 A second copy of the typed notes in found in *Nachlass* Max Born, Staatsbibliothek Berlin, Preussischer Kulturbesitz #1817.

178 "Das Ziel der Vorlesung ist es, die Molekulartheorie der Materie nach dem heutigen Stande unseres Wissens zu entwickeln. Diese Theorie betrachtet die physikalischen Körper und ihre Veränderungen unter dem Scheinbilde eines Systems ungeheuer vieler im Raum bewegter Massen, die durch die Stösse oder durch andere zwischen ihnen wirkenden Kräfte einander beeinflussen." (Hilbert 1912–1913, 1)

179 "Dabei werden wir aber streng axiomatisch die Stellen, in denen die Physik in die mathematische Deduktion eingreift, deutlich hervorheben, und das voneinander trennen, was erstens als logisch willkürliche Definition oder Annahme der Erfahrung entnommen wird, zweitens das, was a priori sich aus diesen Annahmen folgern liesse, aber wegen mathematischer Schwierigkeiten zur Zeit noch nicht sicher gefolgert werden kann, und dritten, das, was bewiesene mathematische Folgerung ist." (Hilbert 1912–1913, 1)

was to express the basic facts of experience in mathematical language, taking them as axioms in need of no further justification. Starting from these axioms one would then deduce as many results as possible, and the logical interdependence of these axioms would also be investigated. In this way, Hilbert declared, the axiomatic method, long applied in mathematics with great success, can also be introduced into physics.[181]

A main task that Hilbert had pursued in his 1905 lectures on axiomatization was to derive, from general physical and mathematical principles in conjunction with the specific axioms of the domain in question, an equation that stands at the center of each discipline and that accounts for the special properties of the particular system under study. Hilbert explicitly stated this as a main task for his system of axioms also in the present case.[182] A first, general axiom he introduced was the "principle of equilibrium," which reads as follows:

> In a state of equilibrium, the masses of the independent components are so distributed with respect to the individual interactions and with respect to the phases, that the characteristic function that expresses the properties of the system attains a minimum value.[183]

Hilbert declared that such an axiom had not been explicitly formulated before and claimed that its derivation from mechanical principles should be done in terms of purely kinetic considerations, such as would be addressed in the third part of the course.[184] At the same time he stated that, in principle, this axiom is equivalent to the second law of thermodynamics, which Hilbert had usually formulated in the past as

---

180 "Wir haben bisher das Problem behandelt, die Beziehung zwischen $p, v,$ und ϑ an einem chemisch völlig homogenen Körper zu ermitteln. Unser Ziel war dabei, diese Beziehung nach den Gesetzen der Mechanik aus der Vorstellung abzuleiten, dass der Körper ein mechanisches System seiner Molekele ist. In dem bisher behandelten, besonders einfache Falle, in dem wir es mit einer einzigen Molekel zu tun hatten, liess sich dies Ziel mit einer gewissen Vollständigkeit erreichen. Eine in einem bestimmten Temperaturintervall mit der Erfahrung übereinstimmende Zustandsgleichung geht nämlich aus der Kinetischen Betrachtung hervor. Mit der Kenntnis der Zustandsgleichung und der Kondensationserscheinungen ist aber nur ein sehr kleiner Teil, der sich empirisch darbietenden Eigenschaften der Stoffe erledigt. *Wir werden daher in diesem zweiten Teile diejenigen Ergebnisse der Physik und Chemie zusammenstellen, deren kinetische Deutung wir uns später zur Aufgabe machen wollen*." (Hilbert 1912–1913, 50)

181 "Die reinen Erfahrungstatsachen werden dabei in mathematischer Sprache erscheinen und als Axiome auftreten, die hier keiner weiteren Begründung bedürfen. Aus diesen Axiomen werden wir soviel als möglich, rein mathematische Folgerungen ziehen, und dabei untersuchen, welche unter den Axiomen voneinander unabhängig sind und welche zum Teil auseinander abgeleitet werden können. Wir werden also den axiomatischen Standpunkt, der in der modernen Mathematik schon zur Geltung gebracht ist, auf die Physik anwenden." (Hilbert 1912–1913, 50)

182 "Um im einzelnen Falle die charakteristische Funktion in ihrer Abhängigkeit von der eigentlichen Veränderlichen und den Massen der unabhängigen Bestandteile zu ermitteln, müssen verschiedenen neue Axiome hinzugezogen werden." (Hilbert 1912–1913, 60)

183 "Im Gleichgewicht verteilen sich die Massen der unabhängigen Bestandteile so auf die einzelnen Verbindungen und Phasen, dass die charakteristische Funktion, die den Bedingungen des Systems entspricht, ein Minimum wird." (Hilbert 1912–1913, 60)

184 "Es muss kinetischen Betrachtung überlassen bleiben, es aus den Prinzipien der Mechanik abzuleiten und wir werden im dritten Teil der Vorlesung die erste Ansätze an solchen kinetischen Theorie kennen lernen." (Hilbert 1912–1913, 61)

the impossibility of the existence of a "*perpetuum mobile*."

The topics for which Hilbert carried out an axiomatic analysis included the equation of state and the third law of thermodynamics. Hilbert's three axioms for the former allowed him a derivation of the expression for the thermodynamical potential of a mixture of gases that was followed by a discussion of the specific role of each of the axioms involved.[185] Concerning the third law of thermodynamics, Hilbert introduced five axioms meant to account for the relationship between the absolute zero temperature, specific heats and entropy. Also in this case he devoted some time to discussing the logical and physical interdependence of these axioms. Hilbert explained that the axiomatic reduction of the most important theorems into independent components (the axioms) is nevertheless not yet complete. The relevant literature, he said, is also full of mistakes, and the real reason for this lies at a much deeper layer. The basic concepts seem to be defined unclearly even in the best of books. The problematic use of the basic concepts of thermodynamics went back in some cases even as far as Helmholtz.[186]

The third part of the course contained, as promised, a "kinetic" section especially focusing on a discussion of rigid bodies. Hilbert explained that the results obtained in the previous sections had been derived from experience and then generalized by means of mathematical formulae. In order to derive them a-priori from purely mechanical considerations, however, one should have recourse to the "fundamental principle of statistical mechanics,"[187] presumably referring to the assumption that all accessible states of a system are equally probable. Hilbert thought that the task of the course would be satisfactorily achieved if those results that he had set out to derive were indeed reduced to the theorems of mechanics together with this principle.[188] At any rate, the issues he discussed in this section included entropy, thermodynamics laws and the quantum hypothesis.

---

185 "Die drei gegebenen Axiome reichen also hin, um das thermodynamische Potential der Mischung zu berechnen. Aber sind nicht in vollem Umfange dazu Notwendig. Nimmt man z.B. das dritte Axiom für eine bestimmte Temperatur gültig an, so folgt es für jede beliebige Temperatur aus den beiden ersten Axiomen. Ebensowenig ist das erste und zweite Axiom vollständig voneinander unabhängig." (Hilbert 1912–1913, 66)

186 "Die axiomatische Reduktion der vorstehenden Sätze auf ihre unabhängigen Bestandteile ist demnach noch nicht vollständig durchgeführt, und es finden sich auch in der Literatur hierüber verschiedene Ungenauigkeiten. Was den eigentlichen Kern solcher Missverständnisse anlangt, so glaube ich, dass er tiefer liegt. Die Grundbegriffe scheinen mir selbst in den besten Lehrbüchern nicht genügend klar dargestellt zu sein, ja, in einem gleich zu erörternden Punkte geht die nicht ganz einwandfreie Anwendung der thermodynamischen Grundbegriffe sogar auch Helmholtz zurück." (Hilbert 1912–1913, 80)

187 "Um die empirisch gegebenen und zu mathematischen Formeln verallgemeinerten Ergebnisse des vorigen Teiles a priori und zwar auf rein mechanischem Wege abzuleiten, greifen wir wieder auf des Grundprinzip des statistischen Mechanik zurück, von der wir bereits im ersten Teil ausgegangen waren." (Hilbert 1912–1913, 88)

188 "Auf die Kritik dieses Grundprinzipes und die Grenzen, die seiner Anwendbarkeit gesteckt sind, können wir hier nicht eingehen. Wir betrachten vielmehr unser Ziel als erreicht, wenn die Ergebnisse, die abzuleiten wir uns zur Aufgabe stellen, auf die Sätze der Mechanik und auf jenes Prinzip zurückgeführt sind." (Hilbert 1912–1913, 88)

It is noteworthy that, although in December 1912, Born himself lectured on Mie's theory of matter at the *Göttinger Mathematische Gesellschaft*,[189] and that this theory touched upon many of the issues taught by Hilbert in this course, neither Mie's name nor his theory are mentioned in the notes. Nor was the theory of relativity mentioned in any way.

11.2 Electron Theory: 1913

In April of 1913 Hilbert organized a new series of Wolfskehl lectures on the current state of research in kinetic theory, to which he invited the leading physicists of the time. Planck lectured on the significance of the quantum hypothesis for kinetic theory. Peter Debye (1884–1966), who would become professor of physics in Göttingen the next year, dealt with the equation of state, the quantum hypothesis and heat conduction. Nernst, whose work on thermodynamics Hilbert had been following with interest,[190] spoke about the kinetic theory of rigid bodies. Von Smoluchowski came from Krakow and lectured on the limits of validity of the second law of thermodynamics, a topic he had already addressed at the Münster meeting of the *Gesellschaft Deutscher Naturforscher und Ärzte*. Sommerfeld came from Munich to talk about problems of free trajectories. Lorentz was invited from Leiden; he spoke on the applications of kinetic theory to the study of the motion of the electron. Einstein was also invited, but he could not attend.[191] Evidently this was for Hilbert a major event and he took pains to announce it very prominently on the pages of the *Physikalische Zeitschrift*, including rather lengthy and detailed abstracts of the expected lectures for the convenience of those who intended to attend.[192] After the meeting Hilbert also wrote a detailed report on the lectures in the *Jahresbericht der Deutschen Mathematiker-Vereiningung*[193]as well as the introduction to the published collection (Planck et al. 1914). Hilbert expressed the hope that the meeting would stimulate further interest, especially among mathematicians, and lead to additional involvement with the exciting world of ideas created by the new physics of matter.

That semester Hilbert also taught two courses on physical issues, one on the theory of the electron and another on the principles of mathematics, quite similar to his 1905 course on the axiomatic method and including a long section on the axiomatization of physics as well. Hilbert's lectures on electron theory emphasized throughout the importance of the Lorentz transformations and of Lorentz covariance, and continually referred back to the works of Minkowski and Born. Hilbert stressed the need to

189 *Jahresbericht der Deutschen Mathematikervereinigung* 22 (1913), 27.

190 In January 1913, Hilbert had lectured on Nernst's law of heat at the Göttingen Physical Society (*Nachlass* David Hilbert, (Cod. Ms. D. Hilbert, 590). See also a remark added in Hilbert's handwriting in (Hilbert 1905a, 167).

191 Cf. Einstein to Hilbert, 4 October 1912 (CPAE 5, Doc. 417).

192 *Physikalische Zeitschrift* 14 (1913), 258–264. Cf. also (Born 1913).

193 *Jahresbericht der Deutschen Mathematiker-Vereinigung* 22 (1913), 53–68, which includes abstract of all the lectures. Cf. also *Jahresbericht der Deutschen Mathematiker-Vereinigung* 23 (1914), 41.

formulate unified theories in physics, and to explain all physical processes in terms of motion of points in space and time.[194] From this reductionistic point of view, the theory of the electron would appear as the most appropriate foundation of all of physics.[195] However, given the difficulty of explicitly describing the motion of, and the interactions among, several electrons, Hilbert indicated that the model provided by kinetic theory had to be brought to bear here. He thus underscored the formal similarities between mechanics, electrodynamics and the kinetic theory of gases. In order to describe the conduction of electricity in metals, he developed a mechanical picture derived from the theory of gases, which he then later wanted to substitute by an electrodynamical one.[196] Hilbert stressed the methodological motivation behind his quest after a unified view of nature, and the centrality of the demand for universal validity of the Lorentz covariance, in the following words:

> But if the relativity principle [i.e., invariance under Lorentz transformations] is valid, then it is so not only for electrodynamics, but for the whole of physics. We would like to consider the possibility of reconstructing the whole of physics in terms of as few basic concepts as possible. The most important concepts are the concept of force and of rigidity. From this point of view the electrodynamics would appear as the foundation of all of physics. But the attempt to develop this idea systematically must be postponed for a later opportunity. In fact, it has to start from the motion of one, of two, etc. electrons, and there are serious difficulties on the way to such an undertaking. The corresponding problem for Newtonian physics is still unsolved for more than two bodies.[197]

When looking at the kind of issues raised by Hilbert in this course, one can hardly be surprised to discover that somewhat later Gustav Mie's theory of matter eventually attracted his attention. Thus, for instance, Hilbert explained that in the existing theory of electrical conductivity in metals, only the conduction of electricity—which itself depends on the motion of electrons—has been considered, while assuming that the electron satisfies both Newton's second law, $F = ma$, and the law of collision as a perfectly elastic spherical body (as in the theory of gases).[198] This approach assumes that the magnetic and electric interactions among electrons are described correctly enough in these mechanical terms as a first approximation.[199] However, if we wish to investigate with greater exactitude the motion of the electron, while at the same time

194 "Alle physikalischen Vorgänge, die wir einer axiomatischen Behandlung zugängig machen wollen, suchen wir auf Bewegungsvorgänge an Punktsystem in Zeit und Raum zu reduzieren." (Hilbert 1913b, 1)

195 "Die Elektronentheorie würde daher von diesem Gesichtpunkt aus das Fundament der gesamten Physik sein." (Hilbert 1913b, 13)

196 "Unser nächstes Ziel ist, eine Erklärung der Elektrizitätsleitung in Metallen zu gewinnen. Zu diesem Zwecke machen wir uns von der Elektronen zunächst folgendes *der Gastheorie entnommene mechanische Bild*, das wir später durch ein elektrodynamisches ersetzen werden." (Hilbert 1913b, 14)

197 "Die wichtigsten Begriffe sind die der *Kraft* und der *Starrhei*t. Die Elektronentheorie würde daher von diesem Gesichtspunkt aus das Fundament der gesamten Physik sein. Den Versuch ihres systematischen Aufbaues verschieben wir jedoch auf später; er hätte von der Bewegung eines, zweier Elektronen u.s.w. auszugehen, und ihm stellen sich bedeutende Schwierigkeiten in der Weg, da schon die entsprechenden Probleme der Newtonschen Mechanik für mehr als zwei Körper ungelöst sind." (Hilbert 1913a, 1913c)

preserving the basic conception of the kinetic theory based on colliding spheres, then we should also take into account the field surrounding the electron and the radiation that is produced with each collision. We are thus led to investigate the influence of the motion of the electron on the distribution of energy in the free aether, or in other words, to study the theory of radiation from the point of view of the mechanism of the motion of the electron. In his 1912 lectures on the theory of radiation, Hilbert had already considered this issue, but only from a "phenomenological" point of view. This time he referred to Lorentz's work as the most relevant one.[200] From Lorentz's theory, he said, we can obtain the electrical force induced on the aether by an electron moving on the x-axis of a given coordinate system.

Later on, Hilbert returned once again to the mathematical difficulties implied by the basic assumptions of the kinetic model. When speaking of clouds of electrons, he said, one assumes the axioms of the theory of gases and of the theory of radiation. The *n*-electron problem, he said, is even more difficult than that of the *n*-bodies, and in any case, we can only speak here of averages. Hilbert thus found it more convenient to open his course by describing the motion of a single electron, and, only later on, to deal with the problem of two electrons.

In discussing the behavior of the single electron, Hilbert referred to the possibility of an electromagnetic reduction of all physical phenomena, freely associating ideas developed earlier in works by Mie and by Max Abraham. The Maxwell equations and the concept of energy, Hilbert said, do not suffice to provide a foundation of electrodynamics; the concept of rigidity has to be added to them. Electricity has to be attached to a steady scaffold, and this scaffold is what we denote as an electron. The electron, he explained to his students, embodies the concept of a rigid connection of Hertz's mechanics. In principle at least it should be possible to derive all the forces of physics, and in particular the molecular forces, from these three ideas: Maxwell's equations, the concept of energy, and rigidity. However, he stressed, one phenomenon has so far evaded every attempt at an electrodynamic explanation: the phenomenon of

---

198 "In der bisherigen Theoire der Elektricitätsleitung in Metallen haben wir nur den Elektrizitätstransport, der durch die Bewegung der Elektronen selbst bedingt wird, in Betracht gezogen; unter der Annahme, dass d*ie Elektronen erstens dem Kraftgesetz $K = mb$ gehorchen und zweitens dasselbe Stossgesetz wie vollkommen harten elastischen Kugeln befolgen* (wie in der Gastheorie)." (Hilbert 1913b, 14)

199 "Auf die elektrischen und magnetische Wirkung der Elektronen aufeinander und auf die Atome sind wir dabei nicht genauer eingegangen, vielmehr haben wir angenommen, dass die gegenseitige Beeinflussung durch das Stossgesetz in erster Annäherung hinreichend genau dargestellt würde." (Hilbert 1913b, 14)

200 "Wollte man die Wirkung der Elektronenbewegung genauer verfolgen—jedoch immer noch unter Beibehaltung des der Gastheorie entlehnten Bildes stossender Kugeln—so müsste man *das umgebende Feld der Elektronen und die Strahlung* in Rechnung setzen, die sie bei jedem Zusammenstoß aussenden. Man wird daher naturgemäß darauf geführt, den Einfluss der Elektronenbewegung auf die Energieverteilung im freien aether zu untersuchen. Ich gehe daher dazu über, die *Strahlungstheorie,* die wir früher vom phänomenologischen Standpunkt aus kennen gelernt haben, aus dem Mechanismus der Elektronenbewegung verständlich zu machen. Eine diesbezügliche Theorie hat *H. A. Lorentz* aufgestellt." (Hilbert 1913b, 14)

gravitation.[201] Still, in spite of the mathematical and physical difficulties that he considered to be associated with a conception of nature based on the model underlying kinetic theory, Hilbert did not fully abandon at this stage the mechanistic approach as a possible one, and in fact he asserted that the latter is a necessary consequence of the principle of relativity.[202]

11.3 Axiomatization of Physics: 1913

In 1913 Hilbert gave a course very similar to the one taught back in 1905, and bearing the same name: "Elements and Principles of Mathematics."[203] The opening page of the manuscript mentions three main parts that the lectures intended to cover:

A. Axiomatic Method.
B. The Problem of the Quadrature of the Circle.
C. Mathematical Logic.

In the actual manuscript, however, one finds only two pages dealing with the problem of the quadrature of the circle. Hilbert explained that, for lack of time, this section would be omitted in the course. Only a short sketch appears, indicating the stages involved in dealing with the problem. The third part of the course, "*Das mathematisch Denken und die Logik*," discussed various issues such as the paradoxes of set theory, false and deceptive reasoning, propositional calculus (*Logikkalkül*), the concept of number and its axioms, and impossibility proofs. The details of the contents of this last part, though interesting, are beyond our present concern here. In the first part Hilbert discussed in detail, like in 1905, the axiomatization of several physical theories.

Like in 1905, Hilbert divided his discussion of the axiomatic method into three parts: the axioms of algebra, the axioms of geometry, and the axioms of physics. In his first lecture Hilbert repeated the definition of the axiomatic method:

---

201 "Auf die Maxwellschen Gleichungen und den Energiebegriff allein kann man die Elektrodynamik nicht gründen. Es muss noch der Begriff der *Starrhei*t hinzukommen; die Elektrizität muss an ein festes Gerüst angeheftet sein. Dies Gerüst bezeichnen wir als Elektron. In ihm ist der Begriff der starrer Verbindung der Hertzschen Mechanik verwirklicht. Aus den Maxwellschen Gleichungen, dem Energiebegriff und dem Starrheitsbegriff lassen sich, im Prinzip wenigstens, die vollständigen Sätze der Mechanik entnehmen, auf sie lassen sich die gesamten Kräfte der Physik, im Besonderen die Molekularkräfte zurückzuführen. Nur die Gravitation hat sich bisher dem Versuch einer elektrodynamischen Erklärung widersetzt." (Hilbert 1913b, 61–62)

202 "Es sind somit die zum Aufbau der Physik unentbehrlichen starren Körper nur in den kleinsten Teilen möglich; man könnte sagen: das Relativitätsprinzip ergibt also als notwendige Folge die Atomistik." (Hilbert 1913b, 65)

203 The lecture notes of this course, (Hilbert 1913c), are not found in the Göttingen collections. Peter Damerow kindly allowed me to consult the copy of the handwritten notes in his possession. The notes do not specify who wrote them. In *Nachlass* David Hilbert, (Cod. Ms. D. Hilbert, 520, 5), Hilbert wrote that notes of the course were taken by Bernhard Baule.

> The axiomatic method consists in choosing a domain and putting certain facts on top of it; the proof of these facts does not occupy us anymore. The classical example of this is provided by geometry.[204]

Hilbert also repeated the major questions that should be addressed when studying a given system of axioms for a determined domain: Are the axioms consistent? Are they mutually independent? Are they complete?[205] The axiomatic method, Hilbert declared, is not a new one; rather it is deeply ingrained in the human way of thinking.[206]

Hilbert's treatment of the axioms of physical theories repeats much of what he presented in 1905 (the axioms of mechanics, the principle of conservation of energy, thermodynamics, calculus of probabilities, and psychophysics), but at the same time it contains some new sections: one on the axioms of radiation theory, containing Hilbert's recently published ideas on this domain, and one on space and time, containing an exposition of relativity. I comment first on one point of special interest appearing in the section on mechanics.

In his 1905 course Hilbert had considered the possibility of introducing alternative systems of mechanics defined by alternative sets of axioms. As already said, one of the intended aims of Hilbert's axiomatic analysis of physical theories was to allow for changes in the existing body of certain theories in the eventuality of new empirical discoveries that contradict the former. But if back in 1905, Hilbert saw the possibility of alternative systems of mechanics more as a mathematical exercise than as a physically interesting task, obviously the situation was considerably different in 1913. This time Hilbert seriously discussed this possibility in the framework of his presentation of the axioms of Newtonian mechanics. As in geometry, Hilbert said, one could imagine for mechanics a set of premises different from the usual ones and, from a logical point of view, one could think of developing a "non-Newtonian Mechanics."[207] More specifically, he used this point of view to stress the similarities between mechanics and electrodynamics. He had already done something similar in 1905, but now his remarks had a much more immediate significance. I quote them here in some extent:

> One can now drop or partially modify particular axioms; one would then be practicing a non-Newtonian, non-Galileian, or non-Lagrangian mechanics.

---

204 "Die axiomatische Methode besteht darin, daß man ein Gebiet herausgreift und bestimmte Tatsachen an die Spitze stellt u. nun den Beweis dieser Tatsachen sich nicht weiter besorgt. Das Musterbeispiel hierfür ist die Geometrie." (Hilbert 1913c, 1)

205 Again, Hilbert is not referring here to the model-theoretical notion of completeness. See § 2.1.

206 "Die axiomatische Methode ist nicht neu, sondern in der menschlichen Denkweise tief begründet." (Hilbert 1913c, 5)

207 "Logisch wäre es natürlich auch möglich andere Def. zu Grunde zu liegen und so eine 'Nicht-Newtonsche Mechanik' zu begründen." An elaborate formulation of a non-Newtonian mechanics had been advanced in 1909 by Gilbert N. Lewis (1875–1946) and Richard C. Tolman (1881–1948), in the framework of an attempt to develop relativistic mechanics independently of electromagnetic theory (Lewis and Tolman 1909). Hilbert did not give here a direct reference to that work but it is likely that he was aware of it, perhaps through the mediation of one of his younger colleagues. (Hilbert 1913c, 91)

> This has a very special significance: electrodynamics has compelled us to adopt the view that our mechanics is only a limiting-case of a more general one. Should anyone in the past have thought by chance of defining the kinetic energy as:
>
> $$T = \mu \frac{1-v^2}{v} \log \frac{1+v}{1-v},$$
>
> he would have then obtained the [equation of] motion of the electron, where μ is constant and depends on the electron's mass. If one ascribes to all of them [i.e., to the electrons] kinetic energy, then one obtains the theory of the electron, i.e., an essential part of electrodynamics. One can then formulate the Newtonian formula:
>
> $$ma = F$$
>
> But now the mass depends essentially on the velocity and it is therefore no more a physical constant. In the limit case, when the velocity is very small, we return to the classical physics....
>
> Lagrange's equations show how a point moves when the conditions and the forces are known. How these forces are created and what is their nature, however, this is a question which is not addressed.
>
> *Boltzmann* attempted to build the whole of physics starting from the forces; he investigated them, and formulated axioms. His idea was to reduce everything to the mere existence of central forces of repulsion or of attraction. According to Boltzmann there are only mass-points, mutually acting on each other, either attracting or repelling, over the straight line connecting them. Hertz was of precisely of the opposite opinion. For him there exist no forces at all; rigid bonds exist among the individual mass-points. Neither of these two conceptions has taken root, and this is for the simple reason that electrodynamics dominates all.
>
> The foundations of mechanics, and especially its goal, are not yet well established. Therefore it has no definitive value to construct and develop these foundations in detail, as has been done for the foundations of geometry. Nevertheless, this kind of foundational research has its value, if only because it is mathematically very interesting and of an inestimably high value.[208]

This passage illuminates Hilbert's conceptions by 1913. At the basis of his approach to physics stands, as always, the axiomatic method as the most appropriate way to examine the logical structure of a theory and to decide what are the individual assumptions from which all the main laws of the theory can be deduced. This deduction, however, as in the case of Lagrange's equation, is independent of questions concerning the ultimate nature of physical phenomena. Hilbert mentions again the mechanistic approach promoted by Hertz and Boltzmann, yet he admits explicitly, perhaps for the first time, that it is electromagnetism that pervades all physical phenomena. Finally, the introduction of Lagrangian functions from which laws of motion may be derived that are more general than the usual ones of classical mechanics was an idea that in the past might have been considered only as a pure mathematical exercise; now—Hilbert cared to stress—it has become a central issue in mechanics, given the latest advances in electrodynamics.

The last section of Hilbert's discussion of the axiomatization of physics addressed the issue of space and time, and in fact it was a discussion of the principle of relativity.[209] What Hilbert did in this section provides the most detailed evidence of his conceptions concerning the principle of relativity, mechanics and electrodynamics before his 1915 paper on the foundations of physics. His presentation did not really incorporate any major innovations, yet Hilbert attempted to make the "new mechanics" appear as organically integrated into the general picture of physics that he was so eager to put forward at every occasion, and in which all physical theories appear as in principle axiomatized (or at least axiomatizable). Back in 1905, Hilbert had suggested, among the possible ways to axiomatize classical dynamics, defining space axiomatically by means of the already established axioms of geometry, and then expanding this definition with some additional axioms that define time. He suggested that something similar should be done now for the new conception of space and time, but that the axioms defining time would clearly have to change. He thus assumed the axioms of Euclidean geometry and proceeded to redefine the concept of time using a "light pendulum." Hilbert then connected the axiomatically constructed theory with the additional empirical consideration it was meant to account for, namely, the outcome of the Michelson-Morley experiment when the values $\vartheta = 0, \pi/2, \pi$, are measured in the formula describing the velocity of the ray-light γ_ϑ in the pendulum:

---

208 "Man kann nun gewisse Teile der Axiome fallen lassen oder modifizieren; dann würde man also "*Nicht-Newtonsche*," od. "*Nicht-Galileische*", od. "*Nicht-Lagrangesche*" *Mechanik* treiben.

Das hat ganz besondere Bedeutung: Durch die Elektrodynamik sind wir zu der Auffassung gezwungen worden, daß unsere Mechanik nur eine Grenzfall einer viel allgemeineren Mechanik ist. Wäre jemand früher zufällig darauf gekommen die kinetisch Energie zu definieren als:

$$T = \mu \frac{1-v^2}{v} \log \frac{1+v}{1-v},$$

so hatte er die Bewegung eines Elektrons, wo μ eine Constante der elektr. Masse ist. Spricht man ihnen allen kinetisch Energie zu, dann hat man die Elektronentheorie d.h. einen wesentlichen Teil der Elektrodynamik. Dann kann man die Newtonschen Gleichungen aufstellen:

$$mb = K$$

Nun hängt aber die Masse ganz wesentlich von der Geschwindigkeit ab und ist keine physikalische Constante mehr. Im Grenzfall, daß die Geschwindigkeit sehr klein ist, kommt man zu der alten Mechanik zurück. (Cf. H. Stark "Experimentelle Elektrizitätslehre," S. 630).

Die Lagrangesche Gleichungen geben die Antwort wie sich ein Punkt bewegt, wenn man die Bedingungen kennt und die Kräfte. Wie diese Kräfte aber beschaffen sind und auf die Natur die Kräfte selbst gehen sie nicht ein.

Boltzmann hat versucht die Physik aufzubauen indem er von der Kräften ausging; er untersuchte diese, stellte Axiome auf u. seine Idee war, alles auf das bloße Vorhandensein von Kräften, die zentral abstoßend oder anziehend wirken sollten, zurückzuführen. Nach Boltzman gibt es nur Massenpunkte die zentral gradlinig auf einander anzieh. od. abstoßend wirkend.

Hertz hat gerade den entgegengesetzten Standpunkt. Für ihn gibt es überhaupt keine Kräfte; starre Verbindungen sind zwischen den einzelnen Massenpunkten.

Beide Auffassungen haben sich nicht eingebürgert, schon aus dem einfachen Grunde, weil die Elektrodynamik alles beherrscht.

Die Grundlagen der Mechanik und besonders die Ziele stehen noch nicht fest, so daß es auch noch nicht definitiven Wert hat die Grundlagen in den einzelnen Details so auf- und ausbauen wie die *Grundlagen der Geometrie*. Dennoch behalten die axiomatischen Untersuchungen ihren Wert, schon deshalb, weil sie mathematisch sehr interessant und von unschätzbar hohen Werte sind." (Hilbert 1913c, 105–108)

$$\gamma_{\vartheta} = \sqrt{\frac{\xi^2 + \eta^2}{t^2}} = \sqrt{\cos^2\vartheta + \sin^2\vartheta - 2v\cos\vartheta + v^2} = [1 - 2v\cos\vartheta + v^2]^{\frac{1}{2}}.$$

Hilbert stressed the similarities between the situation in this case, and in the case in geometry, when one invokes Gauss's measurement of angles in the mountain triangle for determining the validity of Euclidean geometry in reality. In his earlier lectures, Hilbert had repeatedly mentioned this experiment as embodying the empirical side of geometry. The early development of relativity theory had brought about a deep change in the conception of time, but Hilbert of course could not imagine that the really significant change was still ahead. To the empirical discovery that triggered the reformulation of the concept of time, Hilbert opposed the unchanged conception of space instantiated in Gauss's experiment. He thus said:

> Michelson set out to test the correctness of these relations, which were obtained working within the old conception of time and space. The [outcome of his] great experiment is that these formulas do not work, whereas Gauss had experimentally confirmed (i.e., by measuring the Hoher Hagen, the Brocken, and the Inselsberg) that in Euclidean geometry, the sum of the angles of a triangle equals two right angles.[210]

Although he spoke here of an old conception of space and time, Hilbert was referring to a change that actually affected only time. From the negative result of Michelson's experiment, one could conclude that the assumption implied by the old conception—according to which, the velocity of light measured in a moving system has different values in different directions—leads to contradiction. The opposite assumption was thus adopted, namely that the velocity of light behaves with respect to moving systems as it had been already postulated for stationary ones. Hilbert expresses this as a further axiom:

---

209 The following bibliographical list appears in the first page of this section (Hilbert 1913c, 119):
M. Laue Das Relativitätsprinzip 205 S.
M. Planck 8 Vorlesungen über theoretische Physik 8. Vorlesung S. 110–127
A. Brill Das Relativitätsprinzip: eine Einführung in die Theorie 28 S.
H. Minkowski Raum und Zeit XIV Seiten
Beyond this list, together with the manuscript of the course, in the same binding, we find some additions, namely, (1) a manuscript version of Minkowski's famous work (83 pages in the same handwriting as the course itself), (2) the usual preface of A. Gutzmer, appearing as an appendix, and (3) two pages containing a passage copied form Planck's *Vorlesungen*.

210 "Diese aus der alten Auffassung von Raum und Zeit entspringende Beziehung hat Michelson auf ihre Richtigkeit geprüft. Das große Experiment ist nun das, daß diese Formel nicht stimmt, während bei der Euklidischen Geometrie Gauss durch die bestimmte Messung Hoher Hagen, Brocken, Inselsberg bestätigte, daß die Winkelsumme im Dreieck 2 Rechte ist."
On p. 128 Hilbert explained the details of Michelson's calculations, namely, the comparison of velocities at different angles via the formula:

$$\frac{1}{\gamma_{\vartheta}} + \frac{1}{\gamma_{\vartheta+\pi}} = (1 - 2v\cos\vartheta + \vartheta^2)^{-\frac{1}{2}} + (1 + 2v\cos\vartheta + \vartheta^2)^{-\frac{1}{2}} = 2 + v^2(3\cos^2\vartheta - 1) + \ldots$$

where the remaining terms are of higher orders. (Hilbert 1913c, 124)

> Also in a moving system, the velocity of light is identical in all directions, and in fact, identical to that in a stationary system. The moving system has no priority over the first one.[211]

Now the question naturally arises: what is then the true relation between time as measured in the stationary system and in the moving one, t and τ respectively? Hilbert answered this question by introducing the Lorentz transformations, which he discussed in some detail, including the limiting properties of the velocity of light,[212] and the relations with a third system, moving with yet a different uniform velocity.

11.4 Electromagnetic Oscillations: 1913–1914

In the winter semester of 1913–1914, Hilbert lectured on electromagnetic oscillations. As he had done many times in the past, Hilbert opened by referring to the example of geometry as a model of an experimental science that has been transformed into a purely mathematical, and therefore a "theoretical science," thanks to our thorough knowledge of it. Foreshadowing the wording he would use later in his axiomatic formulation of the general theory of relativity, Hilbert said:

> From antiquity the discipline of geometry is a part of mathematics. The experimental grounds necessary to build it are so suggestive and generally acknowledged, that from the outset it has immediately appeared as a theoretical science. I believe that the highest glory that such a science can attain is to be assimilated by mathematics, and that theoretical physics is presently on the verge of attaining this glory. This is valid, in the first place for the relativistic mechanics, or four-dimensional electrodynamics, which belong to mathematics, as I have been already convinced for a long time.[213]

Hilbert's intensive involvement with various physical disciplines over the last years had only helped to strengthen an empirical approach to geometry rather than promoting some kind of formalist views. But as for his conceptions about physics itself, by the end of 1913 his new understanding of the foundational role of electrodynamics was becoming only more strongly established in his mind, at the expense of his old mechanistic conceptions. The manuscript of this course contains the first doc-

---

211 "Es zeigt sich also, daß unsere Folgerung der alten Auffassung, daß die Lichtgeschwindigkeit im bewegtem System nach verschiedenen Richtungen verschieden ist, auf Widerspruch führt. Wir nehmen deshalb an: *Auch im bewegtem System ist die Lichtgeschwindigkeit nach allem Seiten gleich groß, und zwar gleich der im ruhenden. Das bewegte System hat vor dem alten nicht voraus*." *(*Hilbert 1913c, 128–129)

212 "Eine größen Geschwindigkeit als die Lichtgeschwindigkeit kann nicht vorkommen." (Hilbert 1913c, 132)

213 "Seit Alters her ist die Geometrie eine Teildisziplin der Mathematik; die experimentelle Grundlagen, die sie benutzen muss, sind so naheliegend und allgemein anerkannt, dass sie von vornherein und unmittelbar als theoretische Wissenschaft auftrat. Nun glaube ich aber, dass es der höchste Ruhm einer jeden Wissenschaft ist, von der Mathematik assimiliert zu werden, und dass auch die theoretische Physik jetzt im Begriff steht, sich diesen Ruhm zu erwerben. In erster Linie gilt dies von der Relativitätsmechanik oder vierdimensionalen Elektrodynamik, von deren Zuhörigkeit zur Mathematik ich seit langem überzeugt bin." (Hilbert 1913–1914, 1)

umented instance where Hilbert seems to allude to Mie's ideas and, indeed, it is among the earliest explicit instances of a more decided adoption of electrodynamics, rather than mechanics, as the possible foundation for all physical theories. At the same time, the whole picture of mathematics was becoming ever more hierarchical and organized into an organic, comprehensive edifice, of which theoretical physics is also an essential part. Hilbert thus stated:

> In the meantime it looks as if, finally, theoretical physics completely arises from electrodynamics, to the extent that every individual question must be solved, in the last instance, by appealing to electrodynamics. According to what method each mathematical discipline more predominantly uses, one could divide mathematics (concerning contents rather than form) into one-dimensional mathematics, i.e., arithmetic; then function theory, which essentially limits itself to two dimensions; then geometry, and finally four-dimensional mechanics.[214]

In the course itself, however, Hilbert did not actually address in any concrete way the kind of electromagnetic reduction suggested in its introduction, but rather, it continued, to a certain extent, his previous course on electron theory. In the first part Hilbert dealt with the theory of dispersion of electrons, seen as a means to address the *n*-electron problem. Hilbert explained that the role of this problem in the theory of relativity is similar to that of the *n*-body problem in mechanics. In the previous course he had shown that the search for the equations of motion for a system of electrons leads to a very complicated system of integro-differential equations. A possibly fruitful way to address this complicated problem would be to integrate a certain simplified version of these equations and then work on generalizing the solutions thus obtained. In classical mechanics the parallel simplification of the *n*-body problem is embodied in the theory of small oscillations, based on the idea that bodies cannot really attain a state of complete rest. This idea offers a good example of a possible way forward in electrodynamics, and Hilbert explained that, indeed, the elementary theory of dispersion was meant as the implementation of that idea in this field. Thus, this first part of the course would deal with it.[215]

214 "Es scheint indessen, als ob die theoretische Physik schliesslich ganz und gar in der Elektrodynamik aufgeht, insofern jede einzelne noch so spezielle Frage in letzter Instanz an die Elektrodynamik appellieren muss. Nach den Methoden, die die einzelnen mathematischen Disziplinen vorwiegend benutzen, könnte man alsdann – mehr inthaltlich als formell – die Mathematik einteilen in die eindimensionale Mathematik, die Arithmetik, ferner in die Funktionentheorie, die sich im wesentlichen auf zwei Dimensionen beschränkt, in die Geometrie, und schliesslich in die vierdimensionale Mechanik." (Hilbert 1913–1914, 1)

215 "So wenig man schon mit dem *n*-Körperproblem arbeiten kann, so wäre es noch fruchtloser, auf die Behandlung des *n*-Elektronenproblemes einzugehen. Es handelt sich vielmehr für uns darum, das *n*-Elektronenproblem zu verstümmeln, die vereinfachte Gleichungen zu integrieren und von ihren Lösungen durch Korrekturen zu allgemeineren Lösungen aufzusteigen. Die gewöhnliche Mechanik liefert uns hierfür ein ausgezeichnetes Vorbild in der Theorie der kleinen Schwingungen; die Vereinfachung des *n*-Körperproblems besteht dabei darin, dass die Körper sich nur wenig aus festen Ruhelagen entfernen dürfen. In der Elektrodynamik gibt es ein entsprechendes Problem, und zwar würde ich die Theorie der Dispersion als das dem Problem der kleinen Schwingungen analoge Problem ansprechen." (Hilbert 1913–1914, 2)

In the second part of the course Hilbert dealt with the magnetized electron. He did not fail to notice the difficulties currently affecting his reductionist program. At the same time he stressed the value of an axiomatic way of thinking in dealing with such difficulties. He thus said:

> We are really still very distant from a full realization of our leading idea of reducing all physical phenomena to the *n*-electron problem. Instead of a mathematical foundation based on the equations of motion of the electrons, we still need to adopt partly arbitrary assumptions, partly temporary hypothesis, that perhaps one day in the future might be confirmed. We also must adopt, however, certain very fundamental assumptions that we later need to modify. This inconvenience will remain insurmountable for a long time. What sets our presentation apart from that of others, however, is the insistence in making truly explicit all its assumptions and never mixing the latter with the conclusions that follow from them.[216]

Hilbert did not specify what assumptions he meant to include under each of the three kinds mentioned above.Yet it would seem quite plausible to infer that the "very fundamental assumptions," that must be later modified, referred in some way or another to physical, rather than purely mathematical, assumptions, and more specifically, to the atomistic hypothesis, on which much of his own physical conceptions had hitherto been based. An axiomatic analysis of the kind he deemed necessary for physical theories could indeed compel him to modify even his most fundamental assumptions if necessary. The leading principle should remain, in any case, to separate as clearly as possible the assumptions of any particular theory from the theorems that can be derived in it. Thus, the above quotation suggests that if by this time Hilbert had not yet decided to abandon his commitment to the mechanistic reductionism and its concomitant atomistic view, he was certainly preparing the way for that possibility, should the axiomatic analysis convince him of its necessity.

In the subsequent lectures in this course, Hilbert referred more clearly to ideas of the kind developed in Mie's theory, without however explicitly mentioning his name (at least according to the record of the manuscript). Outside ponderable bodies, which are composed of molecules, Hilbert explained, the Maxwell equations are valid. He formulated them as follows:

$$\mathrm{curl} M - \frac{\partial e}{\partial t} = \rho\varpi; \mathrm{div} e = \rho$$
$$\mathrm{curl} e + \frac{\partial M}{\partial t} = 0; \mathrm{div}\ M = 0$$

216 "Von der Verwirklichung unseres leitenden Gedankens, alle physikalischen Vorgänge auf das *n*-Elektronenproblem zurückzuführen, sind wir freilich noch sehr weit entfernt. An Stelle einer mathematischen Begründung aus den Bewegungsgleichungen der Elektronen müssen vielmehr noch teils willkürliche Annahmen treten, teils vorläufige Hypothesen, die später einmal begründet werden dürften, teils aber auch Annahmen ganz prinzipieller Natur, die sicher später modifiziert werden müssen. Dieser Übelstand wird noch auf lange Zeit hinaus unvermeidlich sein. Unsere Darstellung soll sich aber gerade dadurch auszeichnen, dass die wirklich nötigen Annahmen alle ausdrücklich aufgeführt und nicht mit ihren Folgerungen vermischt werden." (Hilbert 1913–1914, 87–88)

This is also how the equations are formulated in Born's article of 1910, the text on which Hilbert was basing this presentation. But Hilbert asserted here for the first time that the equations are valid also inside the body. And he added:

> Inside the body, however, the vectors e and M are very different, since the energy density is always different from zero inside the sphere of the electron, and these spheres undergo swift oscillations. It would not help us to know the exact value of the vector fields inside the bodies, since we can only observe mean values.[217]

Hilbert thus simply stated that the Maxwell equations inside the body should be rewritten as:

$$\mathrm{curl}\overline{M}\left(-\frac{\partial\bar{e}}{\partial t}\right) = \bar{\rho}\overline{\varpi};\ \mathrm{div}e = \rho$$

$$\mathrm{curl}\bar{e} + \frac{\partial\overline{M}}{\partial t} = 0;\ \mathrm{div}\ M = 0$$

where overstrike variables indicate an average value over a space region.

Hilbert went on to discuss separately and in detail specific properties of the conduction-, polarization- and magnetization-electrons. He mentioned Lorentz as the source for the assumption that these three kinds of electrons exist. This assumption, he said, is an "assumption of principle" that should rather be substituted by a less arbitrary one.[218] By saying this, he was thus not only abiding by his self-imposed rules that every particular assumption must be explicitly formulated, but he was also implicitly stressing once again that physical assumptions about the structure of matter are of a different kind than merely mathematical axioms, that they should be avoided whenever possible, and that they should eventually be suppressed altogether.

In a later section of his lecture, dealing with diffuse radiation and molecular forces, Hilbert addressed the problem of gravitation from an interesting point of view that, once again, would seem to allude to the themes discussed by Mie, without however explicitly mentioning his name. Hilbert explained that the problem that had originally motivated the consideration of what he called "diffuse electron oscillations" (a term he did not explain) was the attempt to account for gravitation. In fact, he added, it would be highly desirable—from the point of view pursued in the course—to explain gravitation based on the assumption of the electromagnetic field and the Maxwell equations, together with some auxiliary hypotheses, such as the existence of rigid bodies. The idea of explaining gravitation in terms of "diffuse radiation of a given wavelength" was, according to Hilbert, closely related to an older idea first raised by

217 "Diese Gleichungen gelten sowohl innerhalb wie *ausserhalb des Körpers*. Im innern des Körpers werden aber die Vektoren E und M sich räumlich und zeitlich sehr stark ändern, da die Dichte der Elektrizität immer nur innerhalb der Elektronenkugeln von Null verschieden ist und diese Kugeln rasche Schwingungen ausführen. Es würde uns auch nicht helfen, wenn wir innerhalb des Körpers die genauen Werte der Feldvektoren kennen würden; denn zur Beobachtung gelangen doch nur Mittelwerte." (Hilbert 1913–1914, 89)

218 "Wir machen nur eine reihe von Annahmen, die zu den prinzipiellen gehören und später wohl durch weniger willkürlich scheinende ersetzt werden können." (Hilbert 1913–1914, 90)

Georges-Louis Le Sage (1724–1803). The latter was based on the assumption that a great number of particles move in space with a very high speed, and that their impact with ponderable bodies produces the phenomenon of weight.[219] However, Hilbert explained, more recent research has shown that an explanation of gravitation along these lines is impossible.[220] Hilbert was referring to an article published by Lorentz in 1900, showing that no force of the form $1/r^2$ is created by "diffuse radiation" between two electrical charges, if the distance between them is large enough when compared to the wavelength of the radiation in question (Lorentz 1900).[221]

And yet in 1912, Erwin Madelung had readopted Lorentz's ideas in order to calculate the force produced by radiation over short distances and, eventually, to account for the molecular forces in terms of radiation phenomena (Madelung 1912). Madelung taught physics at that time in Göttingen and, as we saw, he had attended Hilbert's 1912 advanced seminar on kinetic theory. Hilbert considered that the mathematical results obtained by him were very interesting, even though their consequences could not be completely confirmed empirically. Starting from the Maxwell equations and some simple, additional hypotheses, Madelung determined the value of an attraction force that alternatively attains positive and negative values as a function of the distance.[222]

As a second application of diffuse radiation, Hilbert mentioned the possibility of deriving Planck's radiation formula without recourse to quantum theory. Such a derivation, he indicated, could be found in two recent articles of Einstein, one of them (1910) with Ludwig Hopf (1884–1939) and the second one (1913) with Otto Stern (1888–1969).

Hilbert's last two courses on physics, before he began developing his unified theory and became involved with general relativity, were taught in the summer semester

219 LeSage's corpuscular theory of gravitation, originally formulated in 1784, was reconsidered in the late nineteenth century by J.J. Thomson. On the Le Sage-Thomson theory see (North 1965, 38–40; Roseaveare 1982, 108–112). For more recent discussions, cf. also (Edwards 2002).

220 "Das Problem, das zunächst die Betrachtung diffuser Elektronenschwingungen anregte, war die Erklärung der Gravitation. In der Tat muss es ja nach unserem leitenden Gesichtspunkte höchst wünschenswert erscheinen, die Gravitation allein aus der Annahme eines elektromagnetischen Feldes sowie er Maxwellschen Gleichungen und gewisser einfacher Zusatzhypothesen, wie z.B. die Existenz starrer Körper eine ist, zu erklären. Der Gedanke, den Grund für die Erscheinung der Gravitation in einer diffusen Strahlung von gewisser Wellelänge zu suchen, ähnelt entfernt einer Theorie von Le Sage, nach der unzählige kleine Partikel sich mit grosser Geschwindigkeit im Raume bewegen sollen und durch ihren Anprall gegen die ponderablen Körper die Schwere hervorbringen. Wie in dieser theorie ein Druck durch bewegte Partikel auf die Körper ausgeübt wird, hat man jetzt den modernen Versuch unternommen, den Strahlungsdruck für die Erklärung der Gravitation dienstbar zu machen." (Hilbert 1913–1914, 107–108)

221 On this theory, see (McCormmach 1970, 476–477).

222 "Die mathematischen Ergebnisse dieser Arbeit sind von grossem Interesse, auch wenn sich die Folgerungen nicht sämtlich bewähren sollten. Es ergibt sich nämlich allein aus den Maxwellschen Gleichungen und einfachen Zusatzhypothese eine ganz bestimmte Attraktionskraft, die als Funktion der Entfernung periodisch positiv und negativ wird." (Hilbert 1913–1914, 108)

of 1914 (statistical mechanics) and the following winter semester, 1914–1915 (lectures on the structure of matter).[223]

12. BROADENING PHYSICAL HORIZONS - CONCLUDING REMARKS

The present chapter has described Hilbert's intense and wide-ranging involvement with physical issues between 1910 and 1914. His activities comprised both published work and courses and seminars. In the published works, particular stress was laid on considerably detailed axiomatic analysis of theories, together with the application of the techniques developed by Hilbert himself in the theory of linear integral equations. The courses and seminars, however, show very clearly that Hilbert was not just looking for visible venues in which to display the applicability of these mathematical tools. Rather, they render evident the breadth and depth of his understanding of, and interest in, the actual physical problems involved.

Understanding the mixture of these two components—the mathematical and the physical—helps us to understand how the passage from mechanical to electromagnetic reductionism was also the basis of Hilbert's overall approach to physics, and particularly of his fundamental interest in the question of the structure of matter. In spite of the technical possibilities offered by the theory of integral equations in the way to solving specific, open problems in particular theories, Hilbert continued to be concerned about the possible justification of introducing probabilistic methods in physical theories at large. If the phenomenological treatment of theories was only a preliminary stage on the way to a full understanding of physical processes, it turned out that also those treatments based on the atomistic hypothesis, even where they helped reach solutions to individual problems, raised serious foundational questions that required further investigation into the theory of matter as such. Such considerations were no doubt a main cause behind Hilbert's gradual abandonment of mechanical reductionism as a basic foundational assumption.

This background should suffice to show the extent to which his unified theory of 1915 and the concomitant incursion into general theory of relativity were organically connected to the life-long evolution of his scientific horizon, and were thus anything but isolated events. In addition to this background, there are two main domains of ideas that constitute the main pillars of Hilbert's theory and the immediate catalysts for its formulation. These are the electromagnetic theory of matter developed by Gustav Mie starting in 1912, on the one hand, and the efforts of Albert Einstein to generalize the principle of relativity, starting roughly at the same time.

223 The winter semester, 1914–1915 course is registered in the printed version of the *Verzeichnis der Vorlesungen an der Georg-August-Universität zu Göttingen* (1914–1915, on p. 17) but no notes seem to be extant.

REFERENCES

Abraham, Max. 1902 "Dynamik des Elektrons." *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Mathematische-Physikalische Klasse* (1902), 20–41.

———. 1903 "Prinzipien der Dynamik des Elektrons." *Annalen der Physik* 10, 105–179.

Alexandrov, Pavel S. (ed.). 1979. *Die Hilbertsche Probleme*. (*Ostwalds Klassiker der exakten Wissenschaften*, vol. 252.) Leipzig: Akad. Verlagsgesellschaft. (German edition of the Russian original.)

Arabatzis, Theodore. 1996 "Rethinking the 'Discovery' of the Electron." *Studies in History and Philosophy of Modern Physics* 27, 405–435.

Barbour, Julian. 1989. *Absolute or Relative Motion. A Study from a Machian Point of View of the Discovery and the Structure of Dynamical Theories*. Cambridge: Cambridge University Press.

Barkan, Diana. 1993 "The Witches' Sabbath: The First International Solvay Congress in Physics." *Science in Context* 6, 59–82.

Blum, Petra. 1994. *Die Bedeutung von Variationsprinzipien in der Physik für David Hilbert*. Unpublished State Examination. Mainz: Johannes Gutenberg-Universität.

Blumenthal, Otto. 1935 "Lebensgeschichte." (Hilbert 1932–1935, vol. 3, 387–429)

Bohlmann, Georg. 1900 "Ueber Versicherungsmathematik." In F. Klein and E. Riecke (eds.), *Über angewandte Mathematik und Physik in ihrer Bedeutung für den Unterricht an der höheren Schulen*. Leipzig: Teubner, 114–145.

Boltzmann, Ludwig. 1872. "Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen." *Sitzungsberichte Akad. Wiss. Vienna* 66: 275–370. (*Wissenschaftliche Abhandlungen*, vol. 1, 316–402. English translation in (Brush 1966).

———. 1897. *Vorlesungen ueber die Principien der Mechanik*. Leipzig: Verlag von Ambrosius Barth. English translation of the 'Introduction' in (Boltzmann 1974, 223–254).

———. 1899 "Über die Entwicklung der Methoden der theoretischen Physik in neuerer Zeit." In Boltzmann *Populäre Schriften*. Leipzig: J.A. Barth (1905), 198–277). English translation in (Boltzmann 1974, 77–100.)

———. 1900 "Die Druckkräfte in der Hydrodynamik und die Hertzsche Mechanik." *Annalen der Physik* 1, 673–677. (*Wissenschaftliche Abhandlungen*, 3 vols. Leipzig (1909), vol. 3, 665–669. (Chelsea reprint, New York, 1968.)

———. 1974. *Theoretical Physics and Philosophical Problems. Selected Writings*. (Translated by Paul Foulkes, edited by Brian McGuiness, Foreword by S.R. de Groot.) Dordrecht: Reidel.

Born, Max. 1913. "Zur kinetische Theorie der Materie. Einführung zum Kongreß in Göttingen." *Die Naturwissenschaften* 1, 297–299.

———. 1922. "Hilbert und die Physik." *Die Naturwissenschaften* 10, 88–93. (Reprinted in Born 1963, Vol. 2, 584–598.)

———. 1963. *Ausgewählte Abhandlungen*. Göttingen: Vandenhoek & Ruprecht.

———. 1978. *My Life: Recollections of a Nobel Laureate*. New York: Scribner's.

Breitenberg, Ernst. 1984. "Gauss's Geodesy and the Axiom of Parallels." *Archive for History of Exact Sciences* 31, 273–289.

Browder, Felix E. 1976. *Mathematical Developments Arising from Hilbert Problems*. (*Symposia in Pure Mathematics*, vol. 28.) Providence: American Mathematical Society.

Brush, Stephen G. (ed.). 1966. *Kinetic Theory* Vol. 2, *Irreversible Processes*. Oxford: Pergamon Press, 88–175.

Brush, Stephen G. 1976. *The Kind of Motion We Call Heat: A History of the Kinetic Theory of Gases in the 19th Century*. Amsterdam/New York/Oxford: North Holland Publishing House.

Bucherer, Alfred Heinrich. 1903. *Elemente der Vektor-Analysis. Mit Beispielen aus der theroretischen Physik*. Leipzig: Teubner.

Corry, Leo. 2003. *Modern Algebra and the Rise of Mathematical Structures*. Basel/Boston: Birkhäuser. 2nd revised edition. (1st ed.: *Science Networks*, vol. 17, 1996)

———. 2004. *David Hilbert and the Axiomatization of Physics, 1898–1918: From "Grundlagen der Geometrie" to "Grundlagen der Physik"*. Dordrecht: Kluwer.

CPAE 2. 1989. John Stachel, David C. Cassidy, Jürgen Renn, and Robert Schulmann (eds.), *The Collected Papers of Albert Einstein*. Vol. 2. *The Swiss Years: Writings, 1900–1909*. Princeton: Princeton University Press.

CPAE 3. 1993. Martin J. Klein, A. J. Kox, Jürgen Renn, and Robert Schulmann (eds.), *The Collected Papers of Albert Einstein*. Vol. 3. *The Swiss Years: Writings, 1909–1911*. Princeton: Princeton University Press.

CPAE 4. 1995. Martin J. Klein, A. J. Kox, Jürgen Renn, and Robert Schulmann (eds.), *The Collected Papers of Albert Einstein*. Vol. 4. *The Swiss Years: Writings, 1912–1914*. Princeton: Princeton University Press.

CPAE 5. 1993. Martin J. Klein, A. J. Kox, and Robert Schulmann (eds.), *The Collected Papers of Albert Einstein*. Vol. 5. *The Swiss Years: Correspondence, 1902–1914*. Princeton: Princeton University Press.

Crowe, Michael J. 1967. *A History of Vector Analysis. The Evolution of the Idea of a Vectorial System*. University of Notre Dame Press.

Darboux, Gaston. 1875. "Sur la composition des forces en statique." *Bull. Sci. Math. Astr.* 18: 281–288.

Darrigol, Olivier. 2000. *Electrodynamics from Ampère to Einstein*. Chicago: The University of Chicago Press.

Dorier, Jean Luc. 1995. "A General Outline of the Genesis of Vector Space Theory." *Historia Mathematica* 22: 227–261.

Du Bois-Reymond, Emil. 1872. "Ueber die Grenzen des Naturerkennens." Vortrag in der 2. öffentlichen Sitzung der 45. *Versammlung deutscher Naturforscher und Ärzte*, Leipzig, 14 August 1872.

Dugac, Pierre. 1976. *Richard Dedekind et les fondements des mathématiques*. Paris: Vrin.

Duhamel, Jean Marie Constant. 1853–1854. *Cours de mécanique de l'École polytechnique*, 2nd.ed. Paris: Mallet-Bachelier.

Edwards, Mathew R. (ed.). 2002. *Pushing Gravity. New Perspectives on Le Sage's Theory of Gravitation*. Montreal: Apeiron.

Ehrenfest, Paul. 1904. "Die Bewegung Starrer Körper in Flüssigkeiten und die Mechanik von Hertz." In M. Klein (ed.), *Paul Ehrenfest. Collected Scientific Papers*. Amsterdam: North Holland (1959), 1–75.

———. 1911. "Welche Züge der Lichtquantenhypothese spielen in der Theorie der Wärmestrahlung eine wesentliche Rolle?" *Annalen der Physik* 36, 91–118.

Einstein, Albert. 1902. "Kinetischen Theorie der Wärmegleichgewichts und des zweiten Haupsatzes der Thermodynamik." *Annalen der Physik* 9: 417–433, (CPAE 2, Doc. 3).

Einstein, Albert and Ludwig Hopf. 1910. "Statistische Untersuchungen der Bewegung eines Resonators in einem Strahlungsfeld." *Annalen der Physik* 33, 1105–1115, (CPAE 3, Doc. 8).

Einstein, Albert and Otto Stern. 1913. "Einige Argumente für die Annahme einer molekularen Agitation beim absoluten Nullpunkt." *Annalen der Physik* 40: 551–560, (CPAE 4, Doc. 11).

Ewald, William. (ed.). 1999. *From Kant to Hilbert. A Source Book in the Foundations of Mathematics*, 2 vols. Oxford: Clarendon Press.

Ferreirós Domínguez, José. 1999. *Labyrinth of Thought. A History of Set Theory and its Role in Modern Mathematics*. Boston: Birkhäuser. (*Science Networks* Vol. 23.)

Föppl, August. 1901. *Vorlesungen über technische Mechanik*, 2nd. ed. Leipzig: Teubner.

Frege, Gottlob. 1903. *Grundgesetze der Arithmetik*, vol. 2. Jena: Pohle.

Frei, Günther. (ed.). 1985. *Der Briefwechsel David Hilbert-Felix Klein (1996–1919)*. Göttingen: Vandenhoeck & Ruprecht.

Gabriel, Gottfried. et al. (eds.). 1980. *Gottlob Frege - Philosophical and Mathematical Correspondence*. Chicago: The University of Chicago Press. (Abridged from the German edition by Brian McGuiness and translated by Hans Kaal.)

Gans, Richard. 1905. *Einführung in die Vektoranalysis. Mit Anwendungen auf die mathematische Physik*. Leipzig: Teubner.

Gibbs, Josiah Willard. 1902. *Elementary Principles of Statistical Mechanics*. New York: Scribner.

Gleason, A. 1952. "Groups without Small Subgroups." *Annals of Mathematics* 56: 193–212.

Goldberg, Stanley. 1970. "The Abraham Theory of the Electron: The Symbiosis of Experiment and Theory." *Archive for History of Exact Sciences* 7: 7–25.

Grattan-Guinness, Ivor. 2000. "A Sideways Look at Hilbert's Twenty-three Problems of 1900." *Notices American Mathematical Society* 47 (7): 752–757.

———. 2001. *Notices American Mathematical Society*, 48 (2): 167.

Gray, Jeremy J. (ed.). 2000. *The Hilbert Challenge*. New York: Oxford University Press.

Hamel, Georg. 1905. "Über die Zusammensetzung von Vektoren." *Zeitschrift für Mathematik und Physik* 49: 363–371.

———. 1909. "Über Raum, Zeit und Kraft als apriorische Formen der Mechanik." *Jahresbericht der Deutschen Mathematiker-Vereiningung* 18: 357–385.

———. 1927. "Die Axiome der Mechanik." In H. Geiger and K. Scheel (eds.), *Handbuch der Physik*, vol. 5, (*Grundlagen der Mechanik, Mechanik der Punkte und Starren Körper*). Berlin: Springer, 1–130.

Herglotz, Gustav. 1903. "Zur Elektronentheorie." *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Mathematische-Physikalische Klasse*, 357–382.

Hertz, Heinrich. 1894. *Die Prinzipien der Mechanik in neuem Zusammenhänge dargestellt*. Leipzig: Barth.

———. 1956. *The Principles of Mechanics Presented in a New Form*. New York: Dover. English translation of (Hertz 1894).

Hertz, Paul. 1904. *Untersuchungen über unstetige Bewegungen eines Elektrons*. PhD Dissertation, Universität Göttingen.

Hessenberg, Gerhard. 1905. "Beweis des Desarguesschen Satzes aus dem Pascalschen." *Mathematische Annalen* 61, 161–172.

Hilbert, David. 1891. *Projective Geometry*. Nachlass David Hilbert, (Cod. Ms. D. Hilbert, 535).

———. 1893–1894. *Die Grundlagen der Geometrie*. Nachlass David Hilbert, (Cod. Ms. D. Hilbert, 541).

———. 1897. "Die Theorie der algebraischen Zahlkörper (Zahlbericht)." *Jahresbericht der Deutschen Mathematiker-Vereiningung* 4: 175–546. (Hilbert 1932–1935, vol. 1: 63–363.)

———. 1898–1899. *Mechanik*. Nachlass David Hilbert, (Cod. Ms. D. Hilbert, 553).

———. 1899. *Grundlagen der Geometrie*. (*Festschrift zur Feier der Enthüllung des Gauss-Weber- Denkmals in Göttingen*.) Leipzig: Teubner.

———. 1901. "Mathematische Probleme." *Archiv für Mathematik und Physik* 1: 213–237, (Hilbert 1932–1935, vol. 3, 290–329).

———. 1902a. "Mathematical Problems." *Bulletin of the American Mathematical Society* 8: 437–479. English translation by M. Newson of (Hilbert 1901).

———. 1902b "Über die Grundlagen der Geometrie." *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Mathematische-Physikalische Klasse*, 233–241. Reprinted in *Mathematische Annalen* 56, added as *Supplement IV* to (Hilbert 1903a).

———. 1902–1903. *Mechanik der Continua* I. (Manuscript/Typescript of Hilbert Lecture Notes. Bibliothek des Mathematischen Instituts, Universität Göttingen, winter semester, 1902–1903, annotated by Berkovski).

———. 1903a. *Grundlagen der Geometrie,* (2nd revised edition – with five supplements). Leipzig: Teubner.

———. 1903b. *Mechanik der Continua II*. Manuscript/Typescript of Hilbert Lecture Notes. Bibliothek des Mathematisches Institut, Universität Göttingen. SS 1903, annotated by Berkovski.

———. 1904. "Über das Dirichletsche Prinzip." *Mathematische Annalen* 59: 161–186. (Hilbert 1932–1935, vol. 3: 15–37.) Reprinted from *Festschrift zur Feier des 150jährigen Bestehens der Königl. Gesellschaft der Wissenschaften zu Göttingen*, 1901.

———. 1905a. *Logische Principien des mathematischen Denkens*. (Manuscript/Typescript of Hilbert Lecture Notes. Bibliothek des Mathematischen Instituts, Universität Göttingen, summer semester, 1905, annotated by E. Hellinger.)

———. 1905b. *Logische Principien des mathematischen Denkens*. Nachlass David Hilbert, (Cod. Ms. D. Hilbert, 558a) annotated by Max Born.

———. 1905c "Über die Grundlagen der Logik und der Arithmetik." In A. Kneser (ed.), *Verhandlungen aus der Dritten Internationalen Mathematiker-Kongresses in Heidelberg*, 1904, Teubner: Leipzig, 174-185. (English translation by G.B. Halsted: "On the Foundations of Logic and Arithmetic." *The Monist* 15: 338–352. Reprinted in van Heijenoort (ed.) 1967, 129–138.

———. 1905d. "Über das Dirichletsche Prinzip." *Journal für die reine und angewandte Mathematik* 129: 63–67. (Hilbert 1932–1935, vol. 3: 10–14. Reproduced from *Jahresbericht der Deutschen Mathematiker-Vereiningung* 8 (1900): 184–188.

———. 1906. *Mechanik der Kontinua*. (Manuscript/Typescript of Hilbert Lecture Notes. Bibliothek des Mathematischen Insitituts, Universität Göttingen, summer semester, 1906.)

———. 1910–1911. *Mechanik*. (Manuscript/Typescript of Hilbert Lecture Notes. Bibliothek des Mathematischen Insitituts, Universität Göttingen winter semester, 1910–1911, annotated by F. Frankfurther.)

———. 1911–1912. *Kinetische Gastheorie*. (Manuscript/Typescript of Hilbert Lecture Notes. Bibliothek des Mathematischen Insitituts, Universität Göttingen, winter semester, 1911–1912, annotated by E. Hecke.)

———. 1912a. *Grundzüge einer allgemeinen Theorie der linearen Integralgleichungen*. Leipzig: Teubner.

———. 1912b. "Begründung der elementaren Strahlungstheorie." *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Mathematische-Physikalische Klasse*, 773–789; *Physikalische Zeitschrift* 13, 1056–1064, (Hilbert 1932–1935, vol. 3: 217–230).

———. 1912c. *Strahlungstheorie*. (Manuscript/Typescript of Hilbert Lecture Notes. Bibliothek des Mathematischen Insitituts, Universität Göttingen summer semester, 1912, annotated by E. Hecke.)

———. 1912–1913. *Molekulartheorie der Materie*. (Manuscript/Typescript of Hilbert Lecture Notes. Bibliothek des Mathematischen Insitituts, Universität Göttingen, winter semester, 1912–1913.)

———. 1913a. "Begründung der elementaren Strahlungstheorie." *Jahresbericht der Deutschen Mathematiker-Vereiningung* 22: 1–20.

———. 1913b. *Elektronentheorie*. (Manuscript/Typescript of Hilbert Lecture Notes. Bibliothek des Mathematischen Insitituts, Universität Göttingen, summer semester, 1913.)

———. 1913c. *Elemente und Prinzipien der Mathematik*. (summer semester, 1913, Private Collection, Peter Damerow, Berlin.)

———. 1913–1914. *Elektromagnetische Schwingungen*. (Manuscript/Typescript of Hilbert Lecture Notes. Bibliothek des Mathematischen Insitituts, Universität Göttingen, winter semester, 1913–1914.)

———. 1916. *Die Grundlagen der Physik I*. (Manuscript/Typescript of Hilbert Lecture Notes. Bibliothek des Mathematischen Insitituts, Universität Göttingen, summer semester, 1913.)
———. 1918. "Axiomatisches Denken." *Mathematische Annalen* 78: 405–415. (Hilbert 1932–1935, vol. 3: 146–156. Reprinted in (Ewald 1999, Vol. 2, 1107–1115.)
———. 1930. "Naturerkennen und Logik." *Die Naturwissenschaften* 9: 59–63, (Hilbert 1932–1935, vol. 3, 378–387). English translation in (Ewald 1999, 1157–1165).
———. 1932–1935. *David Hilbert – Gesammelte Abhandlungen*, 3 vols. Berlin: Springer. (2nd ed. 1970).
———. 1992. *Natur und Mathematisches Erkennen: Vorlesungen, gehalten 1919–1920 in Göttingen. Nach der Ausarbeitung von Paul Bernays*. (Edited and with an English introduction by David E. Rowe.) Basel: Birkhäuser.
Hirosige, Tetu. 1976. "The Ether Problem, the Mechanistic Worldview, and the Origins of the Theory of Relativity." *Studies in History and Philosophy of Science* 7, 3–82.
Hochkirchen, Thomas H. 1999. *Die Axiomatisierung der Wahrscheinlichkeitsrechnung und ihre Kontexte. Von Hilberts sechstem Problem zu Kolmogoroffs Grundbegriffen*. Göttingen: Vandenhoeck & Ruprecht.
Hon, Giora. 1995. "Is the Identification of an Experimental Error Contextually Dependent? The Case of Kaufmann's Experiment and its Varied Reception." In J. Buchwald (ed.), *Scientific Practice: Theories and Stories of Doing Physics*. Chicago: Chicago University Press, 170–223.
Huntington, Edward V. 1904. "Set of Independent Postulates for the Algebra of Logic." *Trans. AMS* 5: 288–390.
Janssen, Michel. 2002. "Reconsidering a Scientific Revolution: The Case of Einstein versus Lorentz." *Physics in Perspective* 4: 421–446.
Jungnickel, Christa and Russel McCormmach. 1986. *Intellectual Mastery of Nature – Theoretical Physics form Ohm to Einstein*, 2 vols. Chicago: Chicago University Press.
Kirchhoff, Gustav. 1860. "Ueber das Verhältnis zwischen dem Emissionsvermögen und dem Absorptionsvermögen der Körper für Wärme und Licht. *Annalen der Physik* 109: 275–301.
Klein, Martin J. 1970. *Paul Ehrenfest: The Making of a Theoretical Physicist*. Amsterdam: North Holland.
Kragh, Helge. 1999. *Quantum Generations. A History of Physics in the Twentieth Century.* Princeton: Princeton University Press.
Kuhn, Thomas S. 1978. *Black-Body Theory and the Quantum Discontinuity, 1994–1912*. New York. Oxford University Press.
Lamb, Horace. 1895. *Hydrodynamics* (2nd ed.). Cambridge: Cambridge University Press.
Lanczos, Cornelius. 1962. *The Variational Principles of Mechanics*, (2nd ed.). Toronto: University of Toronto Press.
Larmor, Joseph. 1900. *Aether and Matter*. Cambridge: Cambridge University Press.
Lewis, Gilbert N. and Richard C. Tolman. 1909. "The Principle of Relativity, and Non-Newtonian Mechanics." *Philosophical Magazine* 18: 510–523.
Lorentz, Hendrik Antoon. 1895. *Versuch einer Theorie der electrischen und optischen Erscheinungen in bewegten Körpern*. Leiden. In (Lorentz 1934–1939, vol. 5, 1–137).
———. 1898. "Die Fragen, welche die translatorische Bewegung des Lichtäthers betreffen." Verhandlungen GDNA 70 (2. Teil, 1. Hälfte), 56–65. In (Lorentz 1934–1939, vol. 7, 101–115).
———. 1900. "Considérations sur la Pesanteur." *Archives néerlandaises* 7 (1902), 325–338. Translated from *Versl. Kon. Akad. Wet. Amsterdam* 8: 325. In (Lorentz 1934–1939, vol. 5, 198–215).
———. 1904a. "Weiterbildung der Maxwellschen Theorie. Elektronentheorie." *Encyklopädie der mathematischen Wissenschaften mit Einschluss ihrer Anwendungen* 5, 2–14, 145–280.
———. 1904b. "Electromagnetic Phenomena in a System Moving with Velocity Smaller than that of Light." *Versl. Kon. Akad. Wet. Amsterdam* 6, 809–831. (Reprinted in A. Einstein et al. (1952) *The Principle of Relativity.* New York: Dover, 11–34.)
———. 1909. "Le partage de l'energie entre la matière pondérable et l'éther." In G. Castelnuovo (ed.) *Atti del IV congresso internazionale dei matematici* (Rome, 6–11 April 1909). Rome: Tipografia della R. Accademia dei Lincei, Vol. 1, 145–165. (Reprinted with revisions in *Nuovo Cimento* 16 (1908), 5–34.)
———. 1910. "Alte und neue Fragen der Physik." *Physikalische Zeitschrift* 11: 1234–1257. (*Collected Papers of Hendrik Anton Lorentz* 7, 205–207.)
———. 1934–1939. *Collected Papers of Hendrik Anton Lorentz* (9 vols.). The Hague: M. Nijhoff.
Lorey, Wilhelm. 1916. *Das Studium der Mathematik an den deutschen Universitäten seit Anfang des 19. Jahrhunderts*. Leipzig and Berlin: Teubner.
Love, Augustus E. H. 1901. "Hydrodynamik." In *Encyclopädie der mathematischen Wissenschaften mit Einschluss ihrer Anwendungen* 4–3, 48–149.
Madelung, Erwin. 1912. "Die ponderomotorischen Kräfte zwischen Punktladungen in einem mit diffuser electromagnetischer Strahlung erfüllten Raume und die molekularen Kräfte." *Physikalische Zeitschrift* 13: 489–495.
McCormmach, Russel. 1970. "H. A. Lorentz and the Electromagnetic View of Nature." *Isis* 61: 457–497.

Mehrtens, Herbert. 1990. *Moderne - Sprache - Mathematik*. Frankfurt: Suhrkamp.
Miller, Arthur I. 1972. "On the Myth of Gauss's Experiment on the Physical Nature of Space." *Isis* 63: 345–348.
———. 1997. *Albert Einstein's Special Theory of Relativity: Emergence (1905) and Early Interpretation, (1905–1911)*. New York: Springer.
Minkowski, Hermann. 1906. "Kapillarität." In *Encyclopädie der mathematischen Wissenschaften mit Einschluss ihrer Anwendungen* V, 558–613.
———. 1907. *Wärmestrahlung*. Nachlass David Hilbert, (Cod. Ms. D. Hilbert, 707).
Montgomery, Deane and Leo Zippin. 1952. "Small groups of finite-dimensional groups." *Annals of Mathematics* 56: 213–241.
Moore, Eliahim H. 1902. "Projective Axioms of Geometry." *Trans. American Mathematical Society* 3: 142–158.
Moore, Gregory H. 1987. "A House Divided Against Itself: the Emergence of First-Order Logic as the Basis for Mathematics." In E. R. Phillips (ed.), *Studies in the History of Mathematics*, MAA Studies in Mathematics, 98–136.
———. 1995. "The Axiomatization of Linear Algebra: 1875–1940." *Historia Mathematica* 22: 262–303.
Neumann, Carl Gottfried. 1870. *Ueber die Principien der Galilei-Newton'schen Theorie*. Leipzig: Teubner.
Nernst, Walter. 1906. "Ueber die Berechnung chemischer Gleichgewichte aus thermischen Messungen." *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen* 1: 1–39.
Noll, Walter. 1959. "The Foundations of Classical Mechanics in the Light of Recent Advances in Continuum Mechanics." In *The Axiomatic Method with Special Reference to Geometry and Physics*. Amsterdam: North Holland, 266–281. (Reprinted in W. Noll, *The Foundations of Mechanics and Thermodynamics*, New York/ Heidelberg/ Berlin: Springer (1974), 32–47.)
North, John. 1965. *The Measure of the Universe*. Oxford: Clarendon Press.
Peckhaus, Volker. 1990. *Hilbertprogramm und Kritische Philosophie. Der Göttinger Modell interdisziplinärer Zusammenarbeit zwischen Mathematik und Philosophie*. Göttingen: Vandenhoeck & Ruprecht.
Peckhaus, Volker and Reinhard Kahle. 2002. "Hilbert's Paradox." *Historia Mathematica* 29: 157–175.
Planck, Max. 1899. "Über irreversible Strahlungsvorgänge. Dritte Mitteilung (Schluss)." *Königlich Preussische Akademie der Wissenschaften (Berlin) Sitzungsberichte*, 440–480.
———. 1900. "Zur Theorie des Gesetzes der Energieverteilung im Normalspektrum." *Verhandlungen der Deutsche Physikalische Gesellschaft* 2: 237–245.
———. 1906. *Vorlesungen über die Theoire der Wärmestrahlung*, Leipzig.
Planck, Max et al. 1914. *Vorträge über die kinetische Theorie der Materie und der Elektrizität. Gehalten in Göttingen auf Einladung der Kommission der Wolfskehlstiftung*. Leipzig and Berlin: Teubner.
Poincaré, Henri. 1901. *Electricité et optique: La lumière et les théories électrodynamiques. Leçons professées à la Sorbonne en 1999, 1990, et 1990,* (eds. J. Blondin and E. Néculcéa). Paris: **VERLAG**?.
———. 1912. "Sur la théorie des quanta." *Journal Phys. Théor. et Appl.* 2: 5–34.
Prandtl, Ludwig. 1904. "Über Flüssigkeitbewegung bei sehr kleiner Reibung." In A. Kneser (ed.), *Verhandlungen aus der Dritten Internationalen Mathematiker-Kongresses in Heidelberg, 1904*. Teubner: Leipzig, 484–491.
Pyenson, Lewis R. 1979. "Physics in the Shadows of Mathematics: The Göttingen Electron-Theory Seminar of 1905." *Archive for History of Exact Sciences* 21: 55–89. (Reprinted in Pyenson 1985, 101–136.)
———. 1982. "Relativity in Late Wilhelmian Germany: The Appeal to a Pre-established Harmony Between Mathematics and Physics." *Archive for History of Exact Sciences* 138–155. (Reprinted in Pyenson 1985, 137–157.)
———. 1985. *The Young Einstein - The Advent of Relativity.* Bristol and Boston: Adam Hilger Ltd.
Rausenberg, O. 1988. *Lehrbuch der analytischen Mechanik*. Erster Band*: Mechanik der materiellen Punkte*. Zweiter Band: *Mechanik der zusammenhängenden Körper*. Leipzig, Teubner.
Reid, Constance. 1970. *Hilbert*. Berlin/New York: Springer.
Reiff, R. 1900. "Die Druckkräfte in der Hydrodynamik und die Hertzsche Mechanik." *Annalen der Physik* 1: 225–231.
Rowe, David E. 1996. "I 23 problemi de Hilbert: la matematica agli albori di un nuovo secolo." *Storia del XX Secolo: Matematica-Logica-Informatica*. Rome: Enciclopedia Italiana.
———. 1999. "Perspective on Hilbert" (Review of Mehrtens 1990, Peckhaus 1990, and Toepell 1986), *Perspectives on Science*, 5 (4): 533–570.
Rüdenberg, Lily and Hans Zassenhaus (eds.). 1973. *Hermann Minkowski - Briefe an David Hilbert*. Berlin/New York: Springer.
Sackur, Otto. 1912. *Lehrbuch der Thermochemie und Thermodynamik*. Berlin: Springer.
Scanlan, Michael. 1991. "Who were the American Postulate Theorists?" *Journal of Symbolic Logic* 56: 981–1002.

Schirrmacher, Arne. 2003. "Experimenting Theory: The Proofs of Kirchhoff's Radiation Law before and after Planck." *Historical Studies in the Physical Sciences* 33 (2): 299–335.

Schmidt, Arnold. 1933. "Zu Hilberts Grundlegung der Geometrie," (Hilbert 1932–1935, Vol. 2: 404–414).

Scholz, Erhard. 1992. "Gauss und die Begründung der 'höhere' Geodäsie." In M. Folkerts et al. (eds.), *Amphora Festschrift für Hans Wussing zu seinem 65. Geburtstag*. Berlin: Birkhäuser, 631–648.

———. 2004. "C. F. Gauß' Präzisionsmessungen terrestrischer Dreicke und seine Überlegungen zur empirischen Fundierung der Geometrie in den 1820er Jahren." In R. Seising, M. Folkerts, and U. Hashagen (eds.), *Form, Zahl, Ordnung: Studien zur Wissenschafts- und Technikgeschichte: Ivo Schneider zum 65. Geburtstag*. Wiesbaden: Franz Steiner Verlag.

Schur, Friedrich. 1898. "Über den Fundamentalsatz der projektiven Geometrie." *Mathematische Annalen* 51: 401–409.

———. 1901. "Über die Grundlagen der Geometrie." *Mathematische Annalen* 55: 265–292.

———. 1903. "Über die Zusammensetzung von Vektoren." *Zeitschrift für Mathematik und Physik* 49: 352–361.

Schwarzschild, Karl. 1903. "Zur Elektrodynamik: III. Ueber die Bewegung des Elektrons." *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Mathematische-Physikalische Klasse*, 245–278.

Schimmack, R. 1903. "Ueber die axiomatische Begründung der Vektoraddition." *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Mathematische-Physikalische Klasse,* 317–325.

Sinaceur, Hourya. 1984. "De D. Hilbert á E. Artin: les différents aspects du dix-septième problème et les filiations conceptuelles de la théorie des corps reels clos." *Archive for History of Exact Sciences* 29: 267–287.

———. 1991. *Corps et Modèles*. Paris: Vrin.

Sommerfeld, Arnold. 1904a. "Zur Elektronentheorie: I. Allgemeine Untersuchung des Feldes eines beliebig bewegten Elektrons." *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Mathematische-Physikalische Klasse*, 99–130.

———1904b. "Zur Elektronentheorie: II. Grundlagen für eine allgemeine Dynamik des Elektrons." *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Mathematische-Physikalische Klasse*, 363–439.

———. 1905. "Zur Elektronentheorie: III. Ueber Lichtgeschwindichkeits- und Ueberlichtgeschwindichkeits-Elektronen." *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Mathematische-Physikalische Klasse*, 99–130; "II. Grundlagen für eine allgemeine Dynamik des Elektrons." *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Mathematische-Physikalische Klasse*, 201–235.

Thiele, Rüdiger. 2003. "Hilbert's Twenty-Fourth Problem." *American Mathematical Monthly* 110 (1): 1–24.

Toepell, Michael M. 1986. *Über die Entstehung von David Hilberts 'Grundlagen der Geometrie'*. Göttingen: Vandenhoeck & Ruprecht.

Torretti, Roberto. 1978. *Philosophy of Geometry from Riemann to Poincaré*. Dordrecht: Reidel.

Truesdell, Clifford. 1968. *Essays in the History of Mechanics*. New York: Springer.

Veblen, Oswald. 1904. "A System of Axioms for Geometry." *Trans. American Mathematical Society* 5: 343–384.

Voss, Aurel. 1901. "Die Principien der rationellen Mechanik." *Encyclopädie der mathematischen Wissenschaften mit Einschluss ihrer Anwendungen* IV–1, 3–121.

Warwick, Andrew C. 1991. "On the Role of the FizGerald-Lorentz Contraction Hypothesis in the Development of Joseph Larmo's Theory of Matter." *Archive for History of Exact Sciences* 43: 29–91.

———. 2003. *Masters of Theory. Cambridge and the Rise of Mathematical Physics*. Chicago: The University of Chicago Press.

Weyl, Hermann. 1944. "David Hilbert and his Mathematical Work." *Bull. American Mathematical Society* 50: 612–654.

Wiechert, Emil. 1899. *Grundlagen der Elektrodynamik. Festschrift zur Feier der Enthüllung des Gauß-Weber-Denkmals in Göttingen*. Leipzig: Teubner.

———. 1901. "Elektrodymanische Elementargesetze. "*Annalen der Physik* 4: 667–689.

Wien, Wilhelm. 1900. "Ueber die Möglichkeit einer elektromagnetischen Begründung der Mechanik." Archives néerlandaises 5: 96–104. (Reprinted in *Phys. Chem. Ann.* 5 (1901), 501–513.)

Yavetz, Ido. 1995. *From Obscurity to Enigma. The Work of Oliver Heaviside, 1872–1889*. Boston: Birkhäuser. (*Science Networks*, Vol. 16.)

JÜRGEN RENN AND JOHN STACHEL

HILBERT'S FOUNDATION OF PHYSICS: FROM A THEORY OF EVERYTHING TO A CONSTITUENT OF GENERAL RELATIVITY

1. ON THE COMING INTO BEING AND FADING AWAY OF AN ALTERNATIVE POINT OF VIEW

1.1 The Legend of a Royal Road to General Relativity

Hilbert is commonly seen as having publicly presented the derivation of the field equations of general relativity on 20 November 1915, five days before Einstein and after only half a year's work on the subject in contrast to Einstein's eight years of hardship from 1907 to 1915.[1] We thus read in Kip Thorne's fascinating account of recent developments in general relativity (Thorne 1994, 117):

> Remarkably, Einstein was not the first to discover the correct form of the law of warpage [of space-time, i.e. the gravitational field equations], the form that obeys his relativity principle. Recognition for the first discovery must go to Hilbert. In autumn 1915, even as Einstein was struggling toward the right law, making mathematical mistake after mistake, Hilbert was mulling over the things he had learned from Einstein's summer visit to Göttingen. While he was on an autumn vacation on the island of Rugen in the Baltic the key idea came to him, and within a few weeks he had the right law–derived not by the arduous trial-and-error path of Einstein, but by an elegant, succinct mathematical route. Hilbert presented his derivation and the resulting law at a meeting of the Royal Academy of Sciences in Göttingen on 20 November 1915, just five days before Einstein's presentation of the same law at the Prussian Academy meeting in Berlin.[2]

Hilbert himself emphasized that he had two separate starting points for his approach: Mie's electromagnetic theory of matter as well as Einstein's attempt to base a theory of gravitation on the metric tensor. Hilbert's superior mastery of mathematics apparently allowed him to arrive quickly and independently at combined field equa-

1 For discussions of Einstein's path to general relativity see (Norton 1984; Renn and Sauer 1999; Stachel 2002), "The First Two Acts", "Pathways out of Classical Physics …", and "Untying the Knot …", (in vols. 1 and 2 of this series). For historical reviews of Hilbert's contribution, see (Guth 1970; Mehra 1974; Earman and Glymour 1978; Pais 1982, 257–261; Corry 1997; 1999a; 1999b; 1999c; Corry, Renn, and Stachel 1997; Stachel 1989; 2002; Sauer 1999; 2002), "The Origin of Hilbert's Axiomatic Method …" and "Einstein Equations and Hilbert Action" (both in this volume).

2 For a similar account see (Fölsing 1997, 375–376).

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 4
Gravitation in the Twilight of Classical Physics: The Promise of Mathematics.

tions for the electromagnetic and gravitational fields. Although his use of Mie's ideas initially led Hilbert to a theory that was, from the point of view of the subsequent general theory of relativity, restricted to a particular source for the gravitational field—the electromagnetic field—he is nevertheless regarded by many historians of science and physicists as the first to have established a mathematical framework for general relativity that provides both essential results of the theory, such as the field equations, and a clarification of previously obscure conceptual issues, such as the nature of causality in generally-covariant field theories.[3] His contributions to general relativity, although initially inspired by Mie and Einstein, hence appear as a unique and independent achievement. In addition, Hilbert is seen by some historians of science as initiating the subsequent search for unified field theories of gravitation and electromagnetism.[4] In view of all these results, established within a very short time, it appears that Hilbert indeed had found an independent "royal road" to general relativity and beyond.

In a recent paper with Leo Corry, we have shown that Hilbert actually did not anticipate Einstein in presenting the field equations (Corry, Renn, and Stachel 1997).[5] Our argument is based on the analysis of a set of proofs of Hilbert's first paper,[6] hereafter referred to as the "Proofs". These Proofs not only do not include the explicit form of the field equations of general relativity, but they also show the original version of Hilbert's theory to be in many ways closer to the earlier, non-covariant versions of Einstein's theory of gravitation than to general relativity. It was only *after* the publication on 2 December 1915 of Einstein's definitive paper that Hilbert modified his theory in such a way that his results were in accord with those of Einstein.[7] The final version of his first paper, which was not published until March 1916, now includes the explicit field equations and has no restriction on general covariance (Hilbert 1916).[8] Hilbert's second paper, a sequel to his first communication, in which he first discussed causality, apparently also underwent a major revision before eventually being published in 1917 (Hilbert 1917).[9]

3 See (Howard and Norton 1993).

4 See, for example, (Vizgin 1989), who refers to "Hilbert's 1915 unified field theory, in which the attempt was first made to unite gravitation and electromagnetism on the basis of the general theory of relativity" (see p. 301).

5 See also (Stachel 1999), reprinted in (Stachel 2002).

6 A copy of the proofs of Hilbert's first paper is preserved at Göttingen, in SUB Cod. Ms. 634. They comprise 13 pages and are virtually complete, apart from the fact that roughly the upper quarter of two pages (7 and 8) is cut off. The Proofs are dated "submitted on 20 November 1915." The Göttingen copy bears a printer's stamp dated 6 December 1915 and is marked in Hilbert's own hand "First proofs of my first note." In addition, they carry several marginal notes in Hilbert's hand, which are discussed below. A complete translation of the Proofs is given in this volume.

7 The conclusive paper is (Einstein 1915e), which Hilbert lists in the references in (Hilbert 1916).

8 In the following referred to as Paper 1.

9 In the following referred to as Paper 2.

1.2 The Transformation of the Meaning of Hilbert's Work

Hilbert presented his contribution as emerging from a research program that was entirely his own—the search for an axiomatization of physics as a whole—creating a synthesis of electromagnetism and gravitation. This view of his achievement was shared by Felix Klein, who took the distinctiveness of Hilbert's approach as an argument against seeing it from the perspective of a priority competition with Einstein:

> There can be no talk of a priority question in this connection, since both authors are pursuing quite different trains of thought (and indeed, so that initially the compatibility of their results did not even seem certain). Einstein proceeds *inductively* and immediately considers arbitrary material systems. Hilbert *deduces* from previously postulated basic variational principles, while he additionally allows the restriction to electrodynamics. In this connection, Hilbert was particularly close to Mie.[10]

It is clear that, even if one disregards the non-covariant version of his theory as presented in the proofs version of his first paper, both Hilbert's original programmatic aims as well as the interpretation he gave of his own results do not fit into the framework of general relativity as we understand it today. To give one example, which we shall discuss in detail below: In the context of Hilbert's attempt at a synthesis of electromagnetism and gravitation theory, he interpreted the contracted Bianchi identities as a substitute for the fundamental equations of electromagnetism, an interpretation that was soon recognized to be problematic by Hilbert himself.

With hindsight, however, there can be little doubt that a number of important contributions to the development of general relativity do have roots in Hilbert's work: For instance, not so much the variational formulation of the gravitational field equations, an idea which had already been introduced by Einstein[11]; but the choice of the Ricci scalar as the gravitational term in this Lagrangian; and the first hints of Noether's theorem.

The intrinsic plausibility of each of these two perspectives: viewing Hilbert's work as either aiming at a theory differing from general relativity, or as a contribution to general relativity, represents a puzzle. How can Hilbert's contributions be interpreted as making sense only within an independent research program, different in essence from that of Einstein, if ultimately they came to be seen, at least by most physicists, as constituents of general relativity? This puzzle raises a profound historical question concerning the nature of scientific development: how were Hilbert's results, produced within a research program originally aiming at an electrodynamic

---

10 "Von einer Prioritätsfrage kann dabei keine Rede sein, weil beide Autoren ganz verschiedene Gedankengänge verfolgen (und zwar so, daß die Verträglichkeit der Resultate zunächst nicht einmal sicher schien). Einstein geht *induktiv* vor und denkt gleich an beliebige materielle Systeme. Hilbert *deduziert*, indem er übrigens die [...] Beschränkung auf Elektrodynamik eintreten läßt, aus voraufgestellten obersten Variationsprinzipien. Hilbert hat dabei insbesondere auch an Mie angeknüpft." (Klein 1921, 566). The text was originally published in 1917; see (Klein 1917). The quote is from a footnote to remarks added to the 1921 republication. For a recent reconstruction of Hilbert's perspective, see (Sauer 1999).

11 See "Untying the Knot …" (in vol. 2 of this series).

foundation for *all* of physics, eventually transformed into constituents of general relativity, a theory of gravitation? The pursuit of this question promises insights into the processes by which scientific results acquire and change their meaning and, in particular, into the process by which a viewpoint that is different from the one eventually accepted as mainstream emerges and eventually fades away.[12]

Hilbert's work on the foundations of physics turns out to be especially suited for such an analysis, not only because the proofs version of his first paper provides us with a previously unknown point of departure for following his development, but also because he came back time and again to these papers, rewriting them in terms of the insights he had meanwhile acquired and in the light of the developments of Einstein's "mainstream" program. In this paper we shall interpret Hilbert's revisions as indications of the conceptual transformation that his original approach underwent as a consequence of the establishment and further development of general relativity by Einstein, Schwarzschild, Klein, Weyl, and others, including Hilbert himself. We will also show that Hilbert's own understanding of scientific progress induced him to perceive this transformation as merely an elimination of errors and the introduction of improvements and elaborations of a program he had been following from the beginning.

1.3 Structure of the Paper

In the *second section* of this paper ("The origins of Hilbert's program in the 'nostrification' of two speculative physical theories"), we shall analyze the emergence of Hilbert's program for the foundations of physics from his attempt to synthesize, in the form of an axiomatic system, techniques and results of Einstein's 1913/14 non-covariant theory of gravitation and Mie's electromagnetic theory of matter. It will become clear that Hilbert's research agenda was shaped in large part by his understanding of the axiomatic formulation of physical theories, by the technical problems of achieving the synthesis of these two theories, and by open problems in Einstein's theory.

In the *third section* ("Hilbert's attempt at a theory of everything: the proofs of his first paper"), we shall interpret the proofs version of Hilbert's first paper as an attempt to realize the research program reconstructed in the second section. In particular, we shall show that, in the course of pursuing this program, he abandoned his original goal of founding all of physics on electrodynamics, now treating the gravitational field as more fundamental. We shall argue that this reversal was induced by mathematical results, to which Hilbert gave a problematic physical interpretation suggested by his research program; and that the mathematical result at the core of Hilbert's attempt to establish a connection between gravitation and electromagnetism originated in Einstein's claim of 1913/14 that generally-covariant field equations are not compatible with physical causality, a claim supported by Einstein's well-known "hole-argument." Hilbert thus turned Einstein's argument against general covariance into support for Hilbert's own attempt at a unified theory of gravitation and electro-

12 Cf. (Stachel 1994).

magnetism. Hilbert also followed Einstein's 1913/14 attempt to relate the existence of a preferred class of coordinate systems to the requirement of energy conservation. Hilbert's definition of energy, however, was not guided by Einstein's but rather by the goal of establishing a link with Mie's theory. Hilbert's unified theory thus emerges as an extension of Einstein's non-covariant theory of gravitation, in which Mie's speculative theory of matter plays the role of a touchstone, a role played for Einstein by the principle of energy-momentum conservation in classical and special relativistic physics and in Newton's theory of gravitation.

In the *fourth section* ("Hilbert's physics and Einstein's mathematics: the exchange of late 1915") we shall examine Hilbert's and Einstein's exchange of letters at the end of 1915, focussing on the ways in which they mutually influenced each other. We show that Hilbert's attempt at combining a theory of gravitation with a theory of matter had an important impact on the final phase of Einstein's work. Hilbert's vision, which Einstein temporarily adopted, provided the latter with a rather exotic perspective but allowed him to obtain a crucial result, the calculation of Mercury's perihelion precession. This, in turn, guided his completion of the general theory of relativity, but at the same time rendered obsolete its grounding in a specific theory of matter. For Hilbert's theory, on the other hand, Einstein's conclusive paper on general relativity represented a major challenge. It undermined the entire architecture; in particular, the connections Hilbert saw between energy conservation, causality, and the need for a restriction of general covariance.

In the *fifth section* ("Hilbert's adaptation of his theory to Einstein's results: the published versions of his first paper") we shall first discuss how, under the impact of Einstein's results in November 1915, Hilbert modified essential elements of his theory before its publication in March 1916. He abandoned the attempt to develop a non-covariant theory, without as yet having found a satisfactory solution to the causality problem that Einstein had previously raised for generally-covariant theories. He replaced his original, non-covariant notion of energy by a new formulation, still differing from that of Einstein and mainly intended to strengthen the link between his own theory and Mie's electrodynamics. In fact, Hilbert did not abandon his aim of providing a foundation for all of physics. He still hoped to construct a field-theoretical model of the electron and derive its laws of motion in the atom, without, however, getting far enough to include any results in his paper. His first paper was republished twice, in 1924 and 1933, each time with significant revisions. We shall show that Hilbert eventually adopted the understanding of energy-momentum conservation developed in general relativity, thus transforming his ambitious program into an application of general relativity to a special kind of source, matter as described by Mie's theory.

In the *sixth section* ("Hilbert's adoption of Einstein's program: the second paper and its revisions") we shall show that Hilbert's second paper, published in 1917, is the outcome of his attempt to tackle the unsolved problems of his theory in the light of Einstein's results, in particular the causality problem; and at the same time to keep up with the rapid progress of general relativity. In fact, instead of pursuing the consequences of his approach for microphysics, as he originally intended, he now turned to solutions of the gravitational field equations, relating them to the mathematical tradition inaugurated by Gauss and Riemann of exploring the applicability of Euclidean geometry to the physical world. In this way, he effectively worked within the program of general relativity and contributed to solving such problems as the uniqueness of the Minkowski solution and the derivation of the Schwarzschild solution; but he was less successful in dealing with the problem of causality in a generally-covariant theory. Although he followed Einstein in focussing on the invariant features of such a theory, he attempted to develop his own solution to the causality problem, different from that of Einstein. Whereas Einstein resolved the ambiguities he had earlier encountered in the hole argument by the insight that in general relativity coordinate systems have no physical significance apart from the metric, Hilbert attempted to find a purely "mathematical response" to this problem, formulating the causality condition in terms of the Cauchy or initial-value problem for the generally-covariant field equations. While it initiated an important line of research in general relativity, this first attempt not only failed to incorporate Einstein's insights into the physical interpretation of general relativity but also suffered from Hilbert's inadequate treatment of the Cauchy problem for such a theory, a treatment that was finally corrected by the editors of the revised version published in 1933.

In the *seventh section* ("The fading away of Hilbert's point of view in the physics and mathematics communities") we shall analyze the reception of Hilbert's work in contemporary literature on general relativity and unified field theories, as well as its later fate in the textbook tradition. We show that, in spite of Hilbert's emphasis on the distinctiveness of his approach, his work was perceived almost exclusively as a contribution to general relativity. It will become clear that this reception was shaped largely by the treatment of Hilbert's work in the publications of Einstein and Weyl, although, by revising his own contributions in the light of the progress of general relativity, Hilbert was not far behind in contributing to the complete disappearance of his original, distinctive point of view. This disappearance had two remarkable consequences: First, deviations of Hilbert's theory from general relativity, such as his interpretation of the contracted Bianchi identities as the coupling between gravitation and electromagnetism, went practically unremarked. Second, in spite of his attempt to depict himself as the founding father of unified field theories, the early workers in this field tended to ignore his contribution, denying him a prominent place in their intellectual ancestry. Instead, Hilbert was assigned a prominent place in the history of general relativity, even ascribing to him achievements that were not his, such as the first formulation of the field equations or the complete clarification of the question of causality. The ease with which his work could be assimilated to general relativity provides further evidence of a different kind for the tenuous and unstable character of his own framework.

In the *eighth and final section* ("At the end of a royal road") we shall compare Hilbert's and Einstein's approaches in an effort to understand Hilbert's gradual rapprochement with general relativity. Einstein had followed a double strategy in creating general relativity: trying to explore the mathematical consequences of physical principles on the one hand; and systematically checking the physical interpretation of mathematical results, on the other. Hilbert's initial approach encompassed a much narrower physical basis. Starting from a few problematic physical assumptions, Hilbert elaborated a mathematically complex framework, but never succeeded in finding any concrete physical consequences of this framework other than those that had been or could be found within Einstein's theory of general relativity. Nevertheless, Hilbert's assimilation of specific results from the mainstream tradition of general relativity into his framework eventually changed the character of this framework, transforming his results into contributions to general relativity. Thus, in a sense, Hilbert's assimilation of insights from general relativity served as a substitute for the physical component of Einstein's double strategy that was originally lacking in Hilbert's own approach. So this double strategy emerges not only as a successful heuristic characterizing Einstein's individual pathway, but as a particular aspect of the more general process by which additional knowledge was integrated into the further development of general relativity.

2. THE ORIGINS OF HILBERT'S PROGRAM IN THE "NOSTRIFICATION" OF TWO SPECULATIVE PHYSICAL THEORIES

Leo Corry has explored in depth the roots and the history of Hilbert's program of axiomatization of physics and, in particular, its impact on his 1916 paper *Foundations of Physics*.[13] We can therefore limit ourselves to recapitulating briefly some essential elements of this program. Hilbert conceived of the axiomation of physics not as a definite foundation that has to *precede* empirical research and theory formation, but as a *post-hoc* reflection on the results of such investigations with the aim of clarifying the logical and epistemological structure of the assumptions, definitions, etc., on which they are built.[14] Nevertheless, Hilbert expected that a proper axiomatic foundation of physics would not be shaken every time a new empirical fact is discovered; but rather that new, significant facts could be incorporated into the existing body of knowledge without changing its logical structure. Furthermore, Hilbert expected that, rather than emerging from the reorganization of the existing body of knowledge, the concepts used in an axiomatic foundation of physics should be those already familiar from the history of physics. Finally, Hilbert was convinced that one can distinguish sharply between the particular, empirical and the universal ingredients of a physical theory.

Accordingly, the task that Hilbert set for himself was not to find new concepts serving to integrate the existing body of physical knowledge into a coherent conceptual whole, but rather to formulate appropriate axioms involving the already-existing

13 See (Corry 1997; 1999a; 1999b; 1999c; see also Sauer 1999, section 1) and "The Origin of Hilbert's Axiomatic Method …" (in this volume).

14 For evidence of the following claims, see, in particular, Hilbert's lecture notes (Hilbert 1905; 1913), extensively discussed in Corry's papers.

physical concepts; axioms which allow the reconstruction of available physical knowledge by deduction from these axioms. Consequently, his interest in the axiomatization of physics was oriented toward the reductionist attempts to found all of physics on the basis of either mechanics or electrodynamics (the mechanical or electromagnetic worldview). Indeed, in his discussions of the foundations of physics before 1905, the axiomatization of mechanics was central; while, at some point after the advent of the special theory of relativity, Hilbert now placed his hopes in an axiomatization of all physics based on electrodynamics.[15] In spite of the conceptual revolution brought about by special relativity, involving not only the revision of the concepts of space and time but also the autonomy of the field concept from that of the aether, Hilbert nevertheless continued to rely on traditional concepts such as force and rigidity as the building blocks for his axiomatization program.[16]

An axiomatic synthesis of existing knowledge such as that pursued by Hilbert in physics apparently also had a strategic significance for Göttingen mathematicians making it possible for them to leave their distinctive mark on a broad array of domains, which were thus "appropriated," not only intellectually but also in the sense of professional responsibility for them. Minkowski's attempt to present his work on special relativity as a decisive mathematical synthesis of the work of his predecessors may serve as an example.[17] Discussing an accusation that Emmy Noether had neglected to acknowledge her intellectual debt to British and American algebraists, Garrett Birkhoff wrote:

> This seems like an example of German 'nostrification:' reformulating other people's best ideas with increased sharpness and generality, and from then on citing the local reformulation.[18]

2.1 Mie's Theory of Matter

By 1913, Hilbert expected that the electron theory of matter would provide the foundation for all of physics. It is therefore not surprising to find him shortly afterwards attracted to Mie's theory of matter, a non-linear generalization of Maxwell's electrodynamics that aimed at the overcoming of the dualism between "aether" and "ponderable matter." Indeed, Mie had introduced a generalized Hamiltonian formalism for electrodynamics, allowing for non-linear couplings between the field variables, in the hope of deriving the electromagnetic properties of the "aether" as well as the particulate structure of matter from one and the same variational principle.[19] Mie's theory thus not only corresponded to Hilbert's hope to found all of physics on the concepts

---

15 For a discussion of Hilbert's turn from mechanical to electromagnetic monism, see (Corry 1999a, 511–517).

16 See (Hilbert 1913, 13).

17 This attempt is extensively discussed in (Walter 1999). See also (Rowe 1989).

18 Garrett Birkhoff to Bartel Leendert van der Waerden, 1 November 1973 (Eidgenössische Technische Hochschule Zürich, Handschriftenabteilung, Hs 652:1056); quoted from (Siegmund-Schultze 1998, 270). We thank Leo Corry for drawing our attention to this letter.

of electrodynamics; but it must also have been attractive to him because it was based upon the variational calculus, a tool, with the usefulness of which for the axiomatization of physical theories Hilbert was quite familiar.[20] However, Mie's theory was far from able to provide specific results concerning the electromagnetic properties of matter, results which could be confronted with empirical data. Rather, the theory provides only a framework; a suitable "world function" (Lagrangian) must still be found, from which such concrete predictions may then be derived. Mie gave examples of such world functions that, however, were meant to be no more than illustrations of certain features of his framework. In fact, Mie could not have considered these examples as the basis of a specific physical theory since they are not even compatible with basic features of physical reality such as the existence of an elementary quantum of electricity. Concerning his principal example, later taken up by Hilbert, Mie himself remarked:

> A world that is governed by the world function
>
> $$\phi = -\frac{1}{2}\eta^2 + \frac{1}{6}a \cdot \chi^6 \tag{1}$$
>
> must ultimately agglomerate into two large lumps of electric charges, one positive and one negative, and both these lumps must continually tend to separate further and further from each other.[21]

Mie drew the obvious conclusion that the unknown world function he eventually hoped to find must be more complicated than this and the other examples he had considered.[22]

Hilbert based his work on a formulation of Mie's framework actually due to Max Born.[23] In a paper of 1914, Born showed that Mie's variational principle can be considered as a special case of a four-dimensional variational principle for the deformation of a four-dimensional continuum involving the integral:[24]

$$\int \phi(a_{11}, a_{12}, a_{13}, a_{14}; a_{21} \ldots ; u_1, \ldots u_4) dx_1 dx_2 dx_3 dx_4 \ . \tag{2}$$

19 Mie's theory was published in three installments: (Mie 1912a; 1912b; 1913). For a concise account of Mie's theory, see (Corry 1999b), see also the Editorial Note in this volume. In the recent literature on Mie's theory, the problematic physical content of this theory (and hence of its adaptation by Hilbert) plays only a minor role; see the discussion below.

20 See, in particular, (Hilbert 1905).

21 "Eine Welt, die durch die Weltfunktion (1) regiert würde, müßte sich also schließlich zu zwei großen Klumpen elektrischer Ladungen zusammenballen, einem positiven und einem negativen, und diese beiden Klumpen müßten immer weiter und weiter voneinander wegstreben." (Mie 1912b, 38) For the meaning of Mie's formula and its ingredients in Hilbert's version, see (33) below.

22 See (Mie 1912b, 40).

23 For a discussion of Born's role as Hilbert's informant about both Mie's and Einstein's theories, see (Sauer 1999, 538–539).

24 See (Born 1914).

Here ϕ is a Lorentz scalar, and:

$$u_\alpha = u_\alpha(x_1, x_2, x_3, x_4) \qquad \alpha = 1, \ldots 4 \tag{3}$$

are the projections onto four orthogonal axes of the displacements of the points of the four-dimensional continuum from their equilibrium positions regarded as functions of the quasi-Cartesian coordinates x_1, x_2, x_3, x_4 along these axes, and

$$a_{\alpha\beta} = \frac{\partial u_\alpha}{\partial x_\beta} \tag{4}$$

are their derivatives. Furthermore, Born showed that the characteristic feature of Mie's theory lies in the ansatz that the function ϕ depends only on the antisymmetric part of $a_{\alpha\beta}$:

$$a_{\alpha\beta} - a_{\beta\alpha} = \frac{\partial u_\alpha}{\partial x_\beta} - \frac{\partial u_\beta}{\partial x_\alpha}. \tag{5}$$

Mie's four-dimensional continuum could thus be regarded as a four-dimensional spacetime generalization of MacCullagh's three-dimensional aether. MacCullagh had derived equations corresponding to Maxwell's equations for stationary electrodynamic processes from the assumption that the vortices of the aether, rather than its deformations, store its energy.[25]

What role does gravitation play in Mie's theory? Mie opened the series of papers on his theory with a programmatic formulation of his goals, among them to establish a link between the existence of matter and gravitation:

> The immediate goals that I set myself are: to explain the existence of the indivisible electron and: to view the actuality of gravitation as in a necessary connection with the existence of matter. I believe one must start with this, for electric and gravitational effects are surely the most direct expression of those forces upon which rests the very existence of matter. It would be senseless to imagine matter whose smallest parts did not possess electric charges, equally senseless however matter without gravitation.[26]

Initially Mie hoped that he could explain gravitation on the basis of his non-linear electrodynamics alone, without introducing further variables. His search for a new theory of gravitation was guided by a simple model, according to which gravitation is a kind of "atmosphere," arising from the electromagnetic interactions inside the atom:

> An atom is an agglomeration of a larger number of electrons glued together by a relatively dilute charge of opposite sign. Atoms are probably surrounded by more substantial

25 See (Whittaker 1951, 142–145, Schaffner 1972, 59–68).

26 "Die nächsten Ziele, die ich mir gesteckt habe, sind: die Existenz des unteilbaren Elektrons zu erklären und: die Tatsache der Gravitation mit der Existenz der Materie in einem notwendigen Zusammenhang zu sehen. Ich glaube, daß man hiermit beginnen muß, denn die elektrischen und die Gravitationswirkungen sind sicher die unmittelbarsten Äußerungen der Kräfte, auf denen die Existenz der Materie überhaupt beruht. Es wäre sinnlos, Materie zu denken, deren kleinste Teilchen nicht elektrische Ladungen haben, ebenso sinnlos aber Materie ohne Gravitation." See (Mie 1912a, 511–512).

> atmospheres, which however are still so dilute that they do not cause noticeable electric fields, but which presumably are asserted in gravitational effects.[27]

In his third and conclusive paper, however, he explicitly withdrew this model and was forced to introduce the gravitational potential as an additional variable.[28] There is thus no intrinsic connection between gravitation and the other fields in Mie's theory. By representing gravitation as an additional term in his Lagrangian giving rise to a four-vector representation of the gravitational field, he effectively returned to Abraham's gravitation theory which he had earlier rejected.[29] As a consequence, his treatment of gravitation suffers from the same objections that were raised in contemporary discussions of Abraham's theory. In summary, Mie's theory of gravitation was far from reaching the goals he had earlier set for it.

2.2 Einstein's Non-Covariant "Entwurf" Theory of Gravitation

In 1915, Hilbert became interested in Einstein's theory of gravitation after a series of talks on this topic by Einstein between 28 June and 5 July of that year in Göttingen.[30] Hilbert's attraction to Einstein's approach may have stemmed from his dissatisfaction with the contrast between Mie's programmatic statements about the need for a unification of gravitation and electromagnetism and the unsatisfactory treatment of gravitation in Mie's actual theory. This may well have motivated Hilbert to look at other theories of gravitation and perhaps even to invite Einstein. But apart from the shortcomings of Mie's theory, Hilbert's fascination with Einstein's approach to gravitation probably is rooted in the remarkable relations that Hilbert must have perceived between the structure of Mie's theory of electromagnetism and Einstein's theory of gravitation, as the latter was presented in his 1913/1914 publications and (presumably) also in the Göttingen lectures.

Like Mie's theory, Einstein's *Entwurf* theory was based on a variational principle for a Lagrangian H, here considered to be a function of the gravitational potentials (represented by the components of the metric tensor field $g_{\alpha\beta}$) and their first derivatives. In contrast to Mie, however, Einstein had specified a particular Lagrangian, from which he then derived the gravitational field equations:[31]

---

27 "Ein Atom ist eine Zusammenballung einer größeren Zahl von Elektronen, die durch eine verhältnismäßig dünne Ladung von entgegengesetztem Vorzeichen verkittet sind. Die Atome sind wahrscheinlich von kräftigeren Atmosphären umgeben, die allerdings immer noch so dünn sind, daß sie keine bemerkbaren elektrischen Felder veranlassen, die sich aber vermutlich in den Gravitationswirkungen geltend machen." See (Mie 1912a, 512–513).

28 See (Mie 1913, 5).

29 Compare (Mie 1912a, 534) with (Mie 1913, 29).

30 For notes on a part of Einstein's lectures, see "Nachschrift of Einstein's Wolfskehl Lectures" in (CPAE 6, 586–590). For a discussion of Einstein's Göttingen visit and its possible impact on Hilbert, see (Corry 1999a, 514–517).

31 Our presentation follows Einstein's major review paper, (Einstein 1914b).

$$H = \frac{1}{4}\sum_{\alpha\beta\tau\rho} g^{\alpha\beta}\frac{\partial g_{\tau\rho}}{\partial x_\alpha}\frac{\partial g^{\tau\rho}}{\partial x_\beta}. \tag{6}$$

To be more precise, Einstein was able to derive the empty-space field equations from this Lagrangian. The left-hand side of the gravitational field equations is given by the Lagrangian derivative of (6):[32]

$$\mathfrak{E}_{\mu\nu} = \frac{\partial H\sqrt{-g}}{\partial g^{\mu\nu}} - \sum_\sigma \frac{\partial}{\partial x_\sigma}\left(\frac{\partial H\sqrt{-g}}{\partial g_\sigma^{\mu\nu}}\right) \tag{7}$$

where $g_\sigma^{\mu\nu} \equiv \frac{\partial}{\partial x_\sigma} g^{\mu\nu}$. In the presence of matter, the right-hand side of the field equations is given by the energy-momentum tensor $\mathfrak{T}_{\alpha\beta}$ of matter, so that Einstein's field equations become:

$$\mathfrak{E}_{\sigma\tau} = \kappa\mathfrak{T}_{\sigma\tau}, \tag{8}$$

with the universal gravitational constant κ. In Einstein's *Entwurf* theory, the role of matter as an external source of the gravitational field is not determined by the theory, but rather to be prescribed independently. In the Lagrangian, matter thus appears simply "black-boxed," in the form of a term involving its energy-momentum tensor, rather than as an expression explicitly involving some set of variables describing the constitution of matter:

$$\int\left(\delta H - \kappa\sum_{\mu\nu}\mathfrak{T}_{\mu\nu}\delta g^{\mu\nu}\right)d\tau = 0. \tag{9}$$

Here was a possible point of contact between Mie's and Einstein's theories: Was it possible to conceive of Mie's electromagnetic matter as the source of Einstein's gravitational field? In order to answer this question, evidently one had to study how the energy-momentum tensor $\mathfrak{T}_{\alpha\beta}$ can be derived from terms of Mie's Lagrangian; in particular, what happens if Mie's matter is placed in a four-dimensional spacetime described by an arbitrary metric tensor $g_{\mu\nu}$? This naturally presupposed a reformulation of Mie's theory in generally-covariant form, with an arbitrary metric tensor $g_{\mu\nu}$ replacing the flat one of Minkowski spacetime.

Although most other expressions in his theory are generally-covariant, such as the geodesic equations of motion for a particle in the $g_{\mu\nu}$-field and the expression of energy-momentum conservation in the form of the vanishing covariant divergence of the energy tensor of matter, the field equations of Einstein's 1913/14 theory of gravitation are not. While this lack of general covariance had initially seemed to him to be a blemish on his theory, in late 1913 Einstein convinced himself that he could

32 Magnitudes in Gothic script represent tensor densities with respect to linear transformations.

even demonstrate—by means of the well-known "hole-argument"—that generally-covariant field equations are physically inadmissible because they cannot provide a unique solution for the metric tensor $g_{\mu\nu}$ describing the gravitational field produced by a given matter distribution. The hole argument involves a specific boundary value problem (whether this problem is well posed mathematically is a question that Einstein never considered) for a set of generally-covariant field equations with given sources outside of and boundary values on a "hole" (i.e. a region of spacetime without any sources in it), Einstein showed how to construct infinitely many apparently inequivalent solutions starting from any given solution. From the perspective of the hole argument, as Hilbert realized, if one considers generally-covariant field equations, then in order to pick out a unique solution these equations must be supplemented by four additional non-covariant equations. From the perspective of the 1915 theory of general relativity, however, the hole argument no longer represents an objection against generally-covariant field equations because the class of mathematically distinct solutions generated from an initial solution are not regarded as physically distinct, but merely as different mathematical representations of a single physical situation.[33]

Even in 1913/14 Einstein believed that it might be possible to formulate generally-covariant equations, from which equations (8) would follow by introducing a suitable coordinate restriction.[34] While he actually never found such equations corresponding to (8), he did find four non-covariant coordinate restrictions that he believed characteristic for his theory. He obtained these coordinate restrictions from an analysis of the behavior under coordinate transformations of the variational principle, on which his theory was based. Expressed in terms of the Lagrangian H, these four coordinate restrictions are:

$$B_{\mu} = \sum_{\alpha\sigma\nu} \frac{\partial^2}{\partial x_{\sigma} \partial x_{\alpha}} \left(g^{\nu\alpha} \frac{\partial H \sqrt{-g}}{\partial g_{\sigma}^{\mu\nu}} \right) = 0. \tag{10}$$

Einstein regarded these restrictions as making evident the non-general covariance of his theory; indeed he believed them just restrictive enough to avoid the hole-argument. Einstein also required the existence of a gravitational energy-momentum complex (non-tensorial) guaranteeing validity of four energy-momentum conservation equations for the combined matter and gravitational fields. His theory thus involved 10 field equations, 4 coordinate restrictions, and 4 conservation equations — in all 18 equations for the 10 gravitational potentials $g_{\mu\nu}$.

Einstein used the consistency of this overdetermined system as a criterion for the choice of a Lagrangian, imposing the condition that the field equations together with the energy-momentum conservation equations should yield the coordinate restric-

33 See (Stachel 1989; 71–81, sections 3 and 4).

34 See, e.g., (Einstein 1914a, 177–178). It is unclear whether Einstein expected the unknown generally-covariant equations to be of higher order than second.

tions (10). For this purpose, he assumed a general Lagrangian H depending on $g_{\mu\nu}$ and $g_{\mu\nu,\kappa}$, and then examined the four equations implied by the assumption of energy-momentum conservation for the field equations resulting from this Lagrangian. Formulating energy-momentum conservation as the requirement that the covariant divergence of the energy-momentum tensor density $\mathfrak{T}_\sigma{}^\nu$ has to vanish, and using the field equations (8), he first obtained:

$$\begin{aligned}\nabla_\nu \mathfrak{T}_\sigma{}^\nu \equiv \sum_{\nu\tau} \frac{\partial}{\partial x_\nu}(g^{\tau\nu}\mathfrak{T}_{\sigma\tau}) + \frac{1}{2}\sum_{\mu\nu}\frac{\partial g^{\mu\nu}}{\partial x_\sigma}\mathfrak{T}_{\mu\nu} = 0 \Rightarrow \\ \sum_{\nu\tau}\frac{\partial}{\partial x_\nu}(g^{\tau\nu}\mathfrak{E}_{\sigma\tau}) + \frac{1}{2}\sum_{\mu\nu}\frac{\partial g^{\mu\nu}}{\partial x_\sigma}\mathfrak{E}_{\mu\nu} = 0,\end{aligned} \tag{11}$$

and then:

$$\sum_\nu \frac{\partial S_\sigma^\nu}{\partial x_\nu} - B_\sigma = 0, \tag{12}$$

with B_σ given by (10) and:

$$S_\sigma^\nu = \sum_{\mu\tau}\left(g^{\nu\tau}\frac{\partial H\sqrt{-g}}{\partial g^{\sigma\tau}} + g_\mu^{\nu\tau}\frac{\partial H\sqrt{-g}}{\partial g_\mu^{\sigma\tau}} + \frac{1}{2}\partial_\sigma^\nu H\sqrt{-g} - \frac{1}{2}g_\sigma^{\mu\tau}\frac{\partial H\sqrt{-g}}{\partial g_\nu^{\mu\tau}}\right). \tag{13}$$

By requiring that:

$$S_\sigma^\nu \equiv 0, \tag{14}$$

an equation that indeed is satisfied for the Lagrangian (6), it follows that (12) entails no new conditions beyond (10). In other words, for the "right" Lagrangian, the coordinate restrictions required by the hole-argument follow from energy-momentum conservation. In late 1915 Einstein found that his argument for the uniqueness of the Lagrangian, and thus for the uniqueness of the field equations, is fallacious;[35] and this insight helped to motivate him to return to generally-covariant field equations.

If one disregards the wealth of successful predictions of Newtonian gravitation theory that also buttressed Einstein's theory of 1913/14, that theory might appear almost as speculative as Mie's theory of matter. On the one hand, Einstein had been able to make several predictions based on his theory, such as the perihelion shift of Mercury, the deflection of light in a gravitational field, and gravitational redshift, that, at least in principle, could be empirically checked. On the other hand, none of these conclusions had actually received such support by the time Hilbert turned to Einstein's work: indeed, the calculated perihelion shift was in disaccord with observation.

35 For a historical discussion, see (Norton 1984) and "Untying the Knot …" (in vol. 2 of this series).

2.3 Hilbert's Research Program

To a mathematician of Hilbert's competence, Einstein's 1913/1914 theory must have appeared somewhat clumsy. In particular, it left several specifically mathematical questions open, such as the putative existence of the corresponding generally-covariant equations mentioned above; how the field equations (8) result from these generally-covariant equations by means of the coordinate restrictions (10); whether the hole argument for generally-covariant equations is better applied to boundary values on an open space-like hypersurface (the Cauchy problem) or a closed hypersurface (Einstein's formulation); and the closely–related question of the number of independent equations for the gravitational potentials in Einstein's system. Such questions presumably suggested to Hilbert a rather well-circumscribed research program that, taken together with his interest in Mie's theory of matter, amounted to the search for an "axiomatic synthesis" of the two speculative physical theories.

In consequence, Hilbert's initial program presumably comprised:[36]

1. a generally-covariant reformulation of both Mie's and Einstein's theories with the intention of deriving both from a single variational principle for a Lagrangian that depends on both Mie's electrodynamical and Einstein's gravitational variables;
2. an examination of the possibility of replacing Einstein's unspecified energy-momentum tensor for matter by one following from Mie's Lagrangian;
3. a further examination of the non-uniqueness of solutions to generally-covariant equations, involving a study of the question of the number of independent equations, and finally
4. the identification of coordinate restrictions appropriate to delimit a unique solution and an examination of their relation to energy-momentum conservation.

Even prior to looking at Hilbert's attempt to realize such a synthesis of Mie's and Einstein's approaches, it is clear that such a program would fit perfectly into Hilbert's axiomatic approach to physics. Indeed, the realization of this suggested initial program would: constitute a clarification of the logical and mathematical foundations of already existing physical theories in their own terms; represent the synthesis of different theories by combination of logically independent elements within one and the same formalism (in this case incorporation of Mie's variables and Einstein's variables in the same Lagrangian); replace the unspecified character of the material sources entering Einstein's theory with a daring theory of their electromagnetic nature, formulated in mathematical terms, thus shifting the boundary between experience and mathematical deduction in favor of the latter.

Unfortunately, there is no direct evidence that Hilbert developed and pursued some such research program in the course of his work in the second half of 1915 on Mie's and Einstein's theories. We have no "Göttingen notebook" that would be equivalent to Einstein's "Zurich Notebook," documenting in detail the heuristics that Hil-

36 For a similar attempt to reconstruct Hilbert's research program, see (Sauer 1999, 557–559).

bert followed.[37] However, now we have the first proofs of Hilbert's first communication that (as we have argued)[38] provide a glimpse into his thinking prior to his assimilation of Einstein's definitive paper on general relativity. In the next section we shall argue that the proofs version of Hilbert's theory can be interpreted as the result of pursuing just such a research program as that sketched above.

3. HILBERT'S ATTEMPT AT A THEORY OF EVERYTHING: THE PROOFS OF HIS FIRST PAPER

In this section we shall attempt to reconstruct Hilbert's heuristics from the Proofs and published versions of his first paper (Hilbert 1916), hereafter, Proofs and Paper 1. We will begin by reconstructing from the Proofs and other contemporary documents, the first step in the realization of Hilbert's program. This crucial step, an attempt to explore the first two points of the program, was the establishment of a relation between Mie's energy-momentum tensor and the variational derivative with respect to the metric of Mie's Lagrangian.[39] Next, we attempt to reconstruct Hilbert's calculation of Mie's energy-momentum tensor from the Born-Mie Lagrangian. We then examine the consequences of this derivation for the concept of energy, and thus for the further exploration of the second point of his program. We then discuss how these results suggest a new perspective on the relation between Mie's and Einstein's theories, from which gravitation appears more fundamental than electrodynamics. Seen from this perspective, the third point of Hilbert's program, the question of uniqueness of solutions to generally-covariant equations, took on a new significance: Hilbert turned Einstein's argument that only a non-covariant theory can make physical sense into an instrument for the synthesis of electromagnetism and gravitation. Coming to the fourth point of Hilbert's program, we show how he united his energy concept with the requirement of restricting general covariance. Finally, after examining Hilbert's attempt to derive the electromagnetic field equations from the gravitational ones, we discuss Hilbert's rearrangement of his results in the form of an axiomatically constructed theory, which he presented in the Proofs of Paper 1.

3.1 The First Result

At some point in late summer or fall of 1915, Hilbert must have discovered a relation between the energy-momentum tensor following from Mie's theory of matter, the Born-Mie Lagrangian L, and the metric tensor representing the gravitational poten-

---

37 Einstein's search for gravitational field equations in the winter of 1912/13 is documented in the so-called Zurich Notebook, partially published as Doc. 10 of (CPAE 4). Einstein's research project has been reconstructed in volumes 1 and 2 of this series. See, in particular, "Pathways out of Classical Physics …" (in vol. 1 of this series).

38 In (Corry, Renn, and Stachel 1997).

39 Henceforth, mention of the variational derivative of a Lagrangian, without further indication, always means with respect to the metric tensor.

tials in Einstein's theory of gravitation. In the Proofs and the published version of Paper 1, as well as in his contemporary correspondence, Hilbert emphasized the significance of this discovery for his understanding of the relation between Mie's and Einstein's theories. In the Proofs he wrote:

> Mie's electromagnetic energy tensor is nothing but the generally invariant tensor that results from differentiation of the invariant L with respect to the gravitational potentials $g^{\mu\nu}$ in the limit (25) [i.e. the equation $g_{\mu\nu} = \delta_{\mu\nu}$] — a circumstance that gave me the first hint of the necessary close connection between Einstein's general relativity theory and Mie's electrodynamics, and which convinced me of the correctness of the theory here developed.[40]

Hilbert expressed himself similarly in a letter of 13 November 1915 to Einstein:

> I derived most pleasure in the discovery, already discussed with Sommerfeld, that the usual electrical energy results when a certain absolute invariant is differentiated with respect to the gravitation potentials and then g is set = 0,1.[41]

On the basis of our suggested reconstruction of Hilbert's research program, it is possible to suggest what might have led him to this relation. We assume that he attempted to realize the first two steps, that is to reformulate Mie's Lagrangian in a generally-covariant setting and replace the energy-momentum tensor term in Einstein's variational principle by a term corresponding to Mie's theory. Considering (9), this would imply an expression such as $\delta H + \delta L$ under the integral, where H corresponds to Einstein's original Lagrangian and L to a generally-covariant form of Mie's Lagrangian. If the variation of Mie's Lagrangian is regarded as representing the energy-momentum tensor term, one obtains:

$$\delta L = -\kappa \sum_{\mu\nu} \mathfrak{T}_{\mu\nu} \delta g^{\mu\nu}, \tag{15}$$

where $\mathfrak{T}_{\mu\nu}$ should now be the energy-momentum tensor of Mie's theory. It may well have been an equation of this form, following from the attempt to replace the unspecified source-term in Einstein's field equations by a term depending on the generally-covariant form of Mie's Lagrangian, that first suggested to Hilbert that the energy-momentum tensor of Mie's theory could be the variational derivative of Mie's Lagrangian.

---

40 "der Mie'sche elektromagnetische Energietensor ist also nichts anderes als der durch Differentiation der Invariante L nach den Gravitationspotentialen $g^{\mu\nu}$ entstehende allgemein invariante Tensor beim Übergang zum Grenzfall (25) [i.e. the equation $g_{\mu\nu} = \delta_{\mu\nu}$] — ein Umstand, der mich zum ersten Mal auf den notwendigen engen Zusammenhang zwischen der Einsteinschen allgemeinen Relativitätstheorie und der Mie'schen Elektrodynamik hingewiesen und mir die Überzeugung von der Richtigkeit der hier entwickelten Theorie gegeben hat." (Proofs, 10)

41 "Hauptvergnügen war für mich die schon mit Sommerfeld besprochene Entdeckung, dass die gewöhnliche elektrische Energie herauskommt, wenn man eine gewisse absolute Invariante mit den Gravitationspotentialen differenziert und [d]ann $g = 0, 1$ setzt." David Hilbert to Einstein, 13 November 1915, (CPAE 8, 195). Unless otherwise noted, all translations are based on those in the companion volumes to the Einstein edition, but often modified.

If he followed the program outlined above, Hilbert would have assumed that the Lagrangian has the form:

$$H = K + L, \tag{16}$$

where K represents the gravitational part and L the electromagnetic. Indeed, this form of the Lagrangian is used both in the Proofs and the published version of Paper 1.[42]

In Paper 1, Hilbert derived a relation of the form:

$$-2\sum_{\mu}\frac{\partial\sqrt{g}L}{\partial g^{\mu\nu}}g^{\mu m} = T_{\nu}^{m}, \tag{17}$$

where T_{ν}^{m} stands for the energy-momentum tensor density of Mie's theory.[43] This relation, which is exactly what one would expect on the basis of (15), could have suggested to Hilbert that a deep connection must exist between the nature of spacetime as represented by the metric tensor and the structure of matter as represented by Mie's theory.

3.2 Mie's Energy-Momentum Tensor as a Consequence of Generally-Covariant Field Equations

The strategy Hilbert followed to derive (17) can be reconstructed from the two versions of his paper. It consisted in following as closely as possible the standard variational techniques applied, for instance, to derive Lagrange's equations from a variational principle.[44] In Hilbert's paper, a similar variational problem forms the core of his theory. He describes his basic assumptions in two axioms:[45]

> Axiom I (Mie's axiom of the world function): *The law governing physical processes is determined through a world function H that contains the following arguments*:
>
> $$\begin{array}{ll} g_{\mu\nu},\ g_{\mu\nu l} = \dfrac{\partial g_{\mu\nu}}{\partial w_l}, & g_{\mu\nu lk} = \dfrac{\partial^2 g_{\mu\nu}}{\partial w_l \partial w_k}, \\ q_s\ ,\ q_{sl} = \dfrac{\partial q_s}{\partial w_l} & (l, k = 1, 2, 3, 4), \end{array} \tag{18}$$
>
> *where the variation of the integral*

42 In the Proofs it was presumably introduced on the upper part of p. 8, which unfortunately is cut off.

43 See (Proofs, 10; Hilbert 1916, 404). Note that Hilbert uses an imaginary fourth coordinate, so that the minus sign emerges automatically in the determinant of the metric; he does not explicitly introduce the energy-momentum tensor T_{ν}^{m}.

44 See, for example, (Caratheodory 1935).

45 See also (Hilbert 1916, 396).

$$\int H\sqrt{g}\,d\tau \qquad (19)$$
$$(g = |g_{\mu\nu}|, \quad d\tau = dw_1 dw_2 dw_3 dw_4)$$

must vanish for each of the fourteen potentials $g_{\mu\nu}, q_s$.[46]

[The w_s are Hilbert's notation for an arbitrary system of coordinates.]

Axiom II (axiom of general invariance): *The world function* H *is invariant with respect to an arbitrary transformation of the world parameters* w_s.[47]

Starting from an arbitrary invariant J, Hilbert formed a differential expression from it depending on $g^{\mu\nu}, g_l^{\mu\nu}, g_{lk}^{\mu\nu}, q_s, q_{sk}$, which in the published version of his paper he called PJ. He defined the operator P as follows:[48]

$$P = P_g + P_q,$$
$$P_g = \sum_{\mu,\nu,l,k}\left(p^{\mu\nu}\frac{\partial}{\partial g^{\mu\nu}} + p_l^{\mu\nu}\frac{\partial}{\partial g_l^{\mu\nu}} + p_{lk}^{\mu\nu}\frac{\partial}{\partial g_{lk}^{\mu\nu}}\right), \qquad (20)$$
$$P_q = \sum_{l,k}\left(p_l\frac{\partial}{\partial q_l} + p_{lk}\frac{\partial}{\partial q_{lk}}\right),$$

where $p^{\mu\nu}$ and p_l are arbitrary variations of the metric tensor and the electromagnetic four-potentials, respectively. Thus:

$$PJ = \sum_{\mu,\nu,l,k}\left(p^{\mu\nu}\frac{\partial J}{\partial g^{\mu\nu}} + p_l^{\mu\nu}\frac{\partial J}{\partial g_l^{\mu\nu}} + p_{lk}^{\mu\nu}\frac{\partial J}{\partial g_{lk}^{\mu\nu}} + p_l\frac{\partial J}{\partial q_l} + p_{lk}\frac{\partial J}{\partial q_{lk}}\right). \qquad (21)$$

In the mathematical terminology of the time, PJ is a "polarization" of J.[49]

As we shall see, it is possible to derive from PJ identities that realize Hilbert's goal, the derivation of (17). His procedure is described more explicitly in the published version of Paper 1, and since we assume that on this point there was no significant development of Hilbert's thinking after the Proofs, our reconstruction will make use of the published version.

In modern terminology, if $p^{\mu\nu}$ and p_l are those special variations generated by dragging the metric and the electromagnetic potentials over the manifold with some vector field p^s; i.e., if they are the Lie derivatives of the metric and the electromagnetic potentials with respect to p^s,[50] then PJ must be the Lie derivative of J with

---

46 "Axiom I (Mie's Axiom von der Weltfunktion): *Das Gesetz des physikalischen Geschehens bestimmt sich durch eine Weltfunktion H, die folgende Argumente enthält*: [(18); (1) and (2) in the original text] *und zwar muß die Variation des Integrals* [(19)] *für jedes der 14 Potentiale* $g_{\mu\nu}, q_s$ verschwinden." (Proofs, 2) The q_s are the electromagnetic four potentials.

47 "Axiom II (Axiom von der allgemeinen Invarianz): *Die Weltfunktion* H *ist eine Invariante gegenüber einer beliebigen Transformation der Weltparameter* w_s." (Proofs, 2)

48 See (Hilbert 1916, 398–399). Compare (Proofs, 4 and 7).

49 See, e.g., (Kerschensteiner 1887, §2).

respect to p^s. On the other hand, since J is a scalar invariant, the Lie derivative of this scalar with respect to p^s can be written directly, so that:

$$\sum_s \frac{\partial J}{\partial w_s} p^s = PJ. \tag{22}$$

With a little work,[51] equation (22) can be rewritten in the form of equation (23) below. This is the content of Hilbert's Theorem II, both in the Proofs and in Paper 1:

> Theorem II. If J is an invariant depending on $g^{\mu\nu}$, $g_l^{\mu\nu}$, $g_{lk}^{\mu\nu}$, q_s, q_{sk}, then the following is always identically true in all its arguments and for every arbitrary contravariant vector p^s:
>
> $$\begin{aligned}\sum_{\mu,\nu,l,k}&\left(\frac{\partial J}{\partial g^{\mu\nu}}\Delta g^{\mu\nu} + \frac{\partial J}{\partial g_l^{\mu\nu}}\Delta g_l^{\mu\nu} + \frac{\partial J}{\partial g_{lk}^{\mu\nu}}\Delta g_{lk}^{\mu\nu}\right)\\ &+ \sum_{s,k}\left(\frac{\partial J}{\partial q_s}\Delta q_s + \frac{\partial J}{\partial q_{sk}}\Delta q_{sk}\right) = 0;\end{aligned} \tag{23}$$
>
> where
>
> $$\begin{aligned}\Delta g^{\mu\nu} &= \sum_m (g^{\mu m} p_m^\nu + g^{\nu m} p_m^\mu),\\ \Delta g_l^{\mu\nu} &= -\sum_m g_m^{\mu\nu} p_l^m + \frac{\partial \Delta g^{\mu\nu}}{\partial w_l},\\ \Delta g_{lk}^{\mu\nu} &= -\sum_m (g_m^{\mu\nu} p_{lk}^m + g_{lm}^{\mu\nu} p_k^m + g_{km}^{\mu\nu} p_l^m) + \frac{\partial^2 \Delta g^{\mu\nu}}{\partial w_l \partial w_k},\\ \Delta q_s &= -\sum_m q_m p_s^m,\\ \Delta q_{sk} &= -\sum_m q_{sm} p_k^m + \frac{\partial \Delta q_s}{\partial w_k}.\end{aligned} \tag{24}$$ [52]

Hilbert next applies Theorem II to the electromagnetic part L of his Lagrangian $H = K + L$, with the assumption that L only depends on the metric $g^{\mu\nu}$, the elec-

50 Here $p^{\mu\nu}$ corresponds, in modern terms, to the Lie derivative of the contravariant form of the metric tensor with respect to the arbitrary vector p^j. Hilbert writes:

$$p^{\mu\nu} = \sum_s (g_s^{\mu\nu} p^s - g^{\mu s} p_s^\nu - g^{\nu s} p_s^\mu), \quad \left(p_s^j = \frac{\partial p^j}{\partial w_s}\right),$$

and similarly for the Lie derivatives of the electromagnetic potentials. While the term "Lie derivative" was only introduced in 1933 by W. Slebodzinski (see Slebodzinski 1931), it was well known in Hilbert's time that the basic idea came from Lie; see for example (Klein 1917, 471): "For this purpose one naturally determines, as Lie in particular has done in his numerous relevant publications, the formal changes that result from an arbitrary infinitesimal transformation." ("Zu diesem Zwecke bestimmt man natürlich, wie dies insbesondere Lie in seinen zahlreichen einschlägigen Veröffentlichungen getan hat, die formellen Änderungen, welche sich bei einer beliebigen infinitesimalen Transformation ... ergeben ... .") According to Schouten, the name "Lie differential" was proposed by D. Van Dantzig; see (Schouten and Struik 1935, 142).

tromagnetic potentials q_s and their derivatives q_{sk}, but *not* on the derivatives of the metric tensor. This gives the identity:[53]

$$\sum_{\mu,\nu,m}\frac{\partial L}{\partial g^{\mu\nu}}(g^{\mu m}p_m^{\nu}+g^{\nu m}p_m^{\mu})-\sum_{s,m}\frac{\partial L}{\partial q_s}q_m p_s^m$$
$$-\sum_{s,k,m}\frac{\partial L}{\partial q_{sk}}(q_{sm}p_k^m+q_{mk}p_s^m+q_m p_{sk}^m) = 0. \tag{25}$$

Since the vector field p^s is arbitrary, its coefficients as well as the coefficients of its first and second derivatives must vanish identically. Hilbert drew two conclusions, which he interpreted as strong links between a generally-covariant variational principle and Mie's theory of matter. The first concerns the form in which the electromag-

51 See (Proofs, 7–8; Hilbert 1916, 398). The equivalence of (22) and (23) is shown as follows: Since J depends on w_s through $g^{\mu\nu}$, $g_m^{\mu\nu}$, $g_{mk}^{\mu\nu}$, q_m and q_{mk} it follows that:

$$\frac{\partial J}{\partial w_s} = \frac{\partial J}{\partial g^{\mu\nu}}\cdot g_s^{\mu\nu}+\frac{\partial J}{\partial g_m^{\mu\nu}}\cdot g_{sm}^{\mu\nu}+\frac{\partial J}{\partial g_{mk}^{\mu\nu}}\cdot g_{smk}^{\mu\nu}+\frac{\partial J}{\partial q_m}\cdot q_{ms}+\frac{\partial J}{\partial q_{mk}}\cdot q_{mks}.$$

On the other hand, PJ is the Lie derivative of J through its dependence on $g^{\mu\nu}$, $g_m^{\mu\nu}$ q_m and q_{mk}, so:

$$PJ = \frac{\partial J}{\partial g^{\mu\nu}}\cdot p^{\mu\nu}+\frac{\partial J}{\partial g_m^{\mu\nu}}\cdot p_m^{\mu\nu}+\frac{\partial J}{\partial g_{mk}^{\mu\nu}}\cdot p_{mk}^{\mu\nu}+\frac{\partial J}{\partial q_m}\cdot p_m+\frac{\partial J}{\partial q_{mk}}\cdot p_{mk}$$

where $p^{\mu\nu}$ $p_m^{\mu\nu}$, $p_{mk}^{\mu\nu}$, p_m and p_{mk} stand for the Lie derivatives with respect to the vector field p^k of $g^{\mu\nu}$, $g_m^{\mu\nu}$, $g_{mk}^{\mu\nu}$, q_m and q_{mk} respectively (Hilbert's notation). Rewriting (24) in terms of the definition of the Lie derivatives of $g^{\mu\nu}$, $g_m^{\mu\nu}$, $g_{mk}^{\mu\nu}$, q_m and q_{mk}, we easily get:

$$\begin{aligned}\Delta g^{\mu\nu} &= \sum_m g_m^{\mu\nu}p^m-p^{\mu\nu},\\ \Delta g_l^{\mu\nu} &= \sum_m g_{ml}^{\mu\nu}p^m-p_l^{\mu\nu},\\ \Delta g_{lk}^{\mu\nu} &= \sum_m g_{mlk}^{\mu\nu}p^m-p_{lk}^{\mu\nu},\\ \Delta q_s &= \sum_m q_{sm}p^m-p_s,\\ \Delta q_{sk} &= \sum_m q_{smk}p^m-p_{sk}.\end{aligned}$$

Inserting these expressions into (23), and using the equations for $\frac{\partial J}{\partial w_s}$ and PJ at the beginning of this note, one sees that (23) reduces to:

$$\frac{\partial J}{\partial w_s}\cdot p^s-PJ = 0,$$

which is equivalent to (22).

netic potentials enter the Lagrangian, the second concerns the relation between this Lagrangian and Mie's energy-momentum tensor.

From Hilbert's requirements on L—that it be a generally-invariant scalar that does not depend on the derivatives of the metric tensor—he was able to show that the derivatives of the electromagnetic potentials can only enter it in the form characteristic of Mie's theory (see (5)). Setting the coefficients of p^m_{sk} in (25) equal to zero, and remembering that $p^m_{sk} = p^m_{ks}$, one obtains:

$$\left(\frac{\partial L}{\partial q_{sk}} + \frac{\partial L}{\partial q_{ks}}\right) q_m = 0. \tag{26}$$

Since q_m cannot vanish identically, it follows that:

$$\frac{\partial L}{\partial q_{sk}} + \frac{\partial L}{\partial q_{ks}} = 0, \tag{27}$$

which mean that the q_{ik} only enter L in the antisymmetric combination familiar from Mie's theory:

$$M_{ks} = q_{sk} - q_{ks}. \tag{28}$$

Thus, apart from the potentials themselves, L depends only on the components of the tensor M:

$$M = \mathrm{Rot}(q_s), \tag{29}$$

the familiar electromagnetic "six vector." Hilbert emphasized:

> *This result here derives essentially as a consequence of the general invariance, that is, on the basis of axiom II.*[54]

In order to explicitly establish the relation between his theory and Mie's, Hilbert points out that L must be a function of four invariants.[55] Hilbert only gave what he considered to be the "two simplest" of the generally-covariant generalizations of these invariants:

$$Q = \sum_{k,\,l,\,m,\,n} M_{mn} M_{lk} g^{mk} g^{nl} \tag{30}$$

---

52 "Theorem II. Wenn J eine von $g^{\mu\nu}, g_l^{\mu\nu}, g_{lk}^{\mu\nu}, q_s, q_{sk}$ abhängige Invariante ist, so gilt stets identisch in allen Argumenten und für jeden willkürlichen kontravarianten Vektor p^s [(23)] dabei ist: [(24)]."

53 See (Proofs, 9; Hilbert 1916, 403).

54 "*Dieses Resultat ergibt sich hier wesentlich als Folge der allgemeinen Invarianz, also auf Grund von Axiom II.*" (Proofs, 10) In the published version this passage reads: "*This result, which determines the character of Maxwell's equations in the first place, here derives essentially as a consequence of the general invariance, that is, on the basis of axiom II.*" ("*Dieses Resultat, durch welches erst der Charakter der Maxwellschen Gleichungen bedingt ist, ergibt sich hier wesentlich als Folge der allgemeinen Invarianz, also auf Grund von Axiom II.*") See (Hilbert 1916, 403).

55 See (Proofs, 13, and Hilbert 1916, 407). Here Hilbert followed the papers of Mie and Born; see, in particular, (Born 1914).

and:

$$q = \sum_{k,l} q_k q_l g^{kl}. \tag{31}$$

According to Hilbert, the simplest expression that can be formed by analogy to the gravitational part of the Lagrangian K is:[56]

$$L = \alpha Q + f(q), \tag{32}$$

where $f(q)$ is any function of q and α a constant. In order to recover Mie's main example (see (1)) from this more general result, Hilbert considers the following specific functional dependence:

$$L = \alpha Q + \beta q^3, \tag{33}$$

which corresponds to the Lagrangian given by Mie. In contrast to Mie, Hilbert does not even allude to the physical problems associated with this Lagrangian. And in contrast to Einstein, at no point does Hilbert introduce the Newtonian coupling constant into his equations, so that his treatment of gravitation remains as "formalistic" as that of electromagnetism.

The second consequence Hilbert drew from (25), which corresponds to what we have called above "Hilbert's first results" (see (17)), concerns Mie's energy-momentum tensor. Setting the coefficient of p_m^ν equal to zero and using (27), he obtained:[57]

$$2\sum_\mu \frac{\partial L}{\partial g^{\mu\nu}} g^{\mu m} - \frac{\partial L}{\partial q_m} q_\nu - \sum_s \frac{\partial L}{\partial M_{ms}} M_{\nu s} = 0, \quad (\mu = 1, 2, 3, 4). \tag{34}$$

Noting that:

$$2\sum_\mu \frac{\partial L}{\partial g^{\mu\nu}} g^{\mu m} = \frac{2}{\sqrt{g}} \cdot \sum_\mu \frac{\partial \sqrt{g} L}{\partial g^{\mu\nu}} g^{\mu m} + L \cdot \delta_\nu^{\ m}, \tag{35}$$

(34) can be rewritten:

$$-2\sum_\mu \frac{\partial \sqrt{g} L}{\partial g^{\mu\nu}} g^{\mu m} = \sqrt{g}\left\{ L\delta_\nu^m - \frac{\partial L}{\partial q_m} q_\nu - \sum_s \frac{\partial L}{\partial M_{ms}} M_{\nu s} \right\}, \tag{36}$$
$$(\mu = 1, 2, 3, 4) \quad (\delta_\nu^\mu = 0,\ \mu \neq \nu,\ \delta_\mu^\mu = 1).$$

The right-hand side of this equation is the generally-covariant generalization of Mie's energy-momentum tensor. It is this equation that inspired Hilbert's remark about the

56 Note that Q is the term that gives rise to Maxwell's equations and that q cannot be used if the resulting theory is to be gauge invariant. See (Born and Infeld 1934).

57 See (Proofs, 10; Hilbert 1916, 404).

"Umstand, der mich zum ersten Mal auf den notwendigen engen Zusammenhang zwischen der Einsteinschen allgemeinen Relativitätstheorie und der Mie'schen Elektrodynamik hingewiesen ... hat", quoted above (p. 873). Hilbert had shown that characteristic properties of Mie's Lagrangian follow from its generally-covariant generalization, a result he interpreted as indicating that gravitation must be conceived as being more fundamental than electromagnetism, as his later work indicates.

3.3 The Definition of Energy

While (36) shows a strong link between a generally-covariant L and Mie's energy momentum tensor, it does not answer the question of how energy-momentum conservation is to be conceived in Hilbert's theory. Hilbert's theory does not allow the interpretation of an energy-momentum tensor for matter as an external source, as does that of Einstein; so Hilbert could not start from a conservation law for matter in Minkowski spacetime and simply generalize it to the case in which a gravitational field is present. Such a procedure would have conflicted with Hilbert's heuristic, according to which matter itself is conceived in terms of electromagnetic fields that, in turn, arise in conjunction with, or even as an effect of, gravitational fields.

Hilbert's heuristic for finding an appropriate definition of energy seems to be governed by a formal criterion related to his understanding of energy conservation in classical physics, as well as by a criterion with a more specific physical meaning related to the results he expected from Mie's theory. Hilbert's formal criterion is well described in a passage in his summer-semester 1916 lectures on the foundations of physics, a passage which occurs in a discussion of energy-momentum conservation in Mie's theory:

> The energy concept comes from just writing Lagrange's equations in the form of a divergence, and defining as energy what is represented as divergent.[58]

As for Hilbert's physical criterion, any definition of the energy must be compatible with his insight that the variational derivative of Mie's Lagrangian yields the electromagnetic energy-momentum tensor.

Hilbert's treatment of energy conservation in the Proofs and in Paper 1 is not easy to follow. This difficulty was felt by Hilbert's contemporaries; both Einstein and Klein had their problems with it.[59] Nevertheless, as will become clear in what follows, Hilbert's discussion was guided by the heuristic criteria mentioned above. He proceeded in three steps:

- he first identified an energy expression consisting of a sum of divergence terms (Satz 1 in the Proofs):

---

58 "Der Energiebegriff kommt eben daher, dass man die Lagrangeschen Gleichungen in Divergenzform schreibt, und das, was unter der Divergenz steht, als Energie definiert." Die Grundlagen der Physik I, Ms. Vorlesung SS 1916, 98 (D. Hilbert, Bibliothek des Mathematischen Seminars, Universität Göttingen); from here on "SS 1916 Lectures."

- he then formulated a divergence equation for his energy expression in analogy to classical and special-relativistic results (Satz 2 in the Proofs), and imposed this equation as a requirement implying coordinate restrictions (Axiom III):
- finally, he showed that his energy expression can be related to Mie's energy-momentum tensor (the real justification of his choice).

Here we focus on the first and last of these points, deferring the issue of coordinate restrictions to a subsequent section ("Energy-momentum conservation and coordinate restrictions").

As in his derivation of the connection between Mie's energy-momentum tensor and the variational derivative of the Lagrangian, Hilbert's starting point was his generally-covariant variational principle. However, he now proceeded somewhat differently. Instead of focussing on the electromagnetic part L, he considered the entire Lagrangian H, but now neglected the derivatives with respect to the electromagnetic potentials, i.e. the contribution of the term P_q to P (see (20)). Accordingly, Hilbert forms the expression:[60]

$$J^{(p)} = \sum_{\mu,\nu} \frac{\partial H}{\partial g^{\mu\nu}} p^{\mu\nu} + \sum_{\mu,\nu,k} \frac{\partial H}{\partial g_k^{\mu\nu}} p_k^{\mu\nu} + \sum_{\mu,\nu,k,l} \frac{\partial H}{\partial g_{kl}^{\mu\nu}} p_{kl}^{\mu\nu}, \tag{37}$$

where $p^{\mu\nu}$ corresponds, as we have seen, to the Lie derivative of the metric tensor with respect to the arbitrary vector p^j. By partial integration, Hilbert transforms this expression into:

$$\sqrt{g}J^{(p)} = -\sum_{\mu,\nu} H \frac{\partial \sqrt{g}}{\partial g^{\mu\nu}} p^{\mu\nu} + E + D^{(p)}, \tag{38}$$

with:

---

59 In (Klein 1917, 475), Klein quotes from a letter he had written to Hilbert concerning the latter's energy expression in Paper 1: "But I find your equations so complicated that I have not attempted to redo your calculations." ("Ich finde aber Ihre Formeln so kompliziert, daß ich die Nachrechnung nicht unternommen habe.") In a letter, in which Einstein asked Hilbert for a clarification of the latter's energy theorem, he wrote: "Why do you make it so hard for poor mortals by withholding the technique behind your ideas? It surely does not suffice for the thoughtful reader if, although able to verify the correctness of your equations, he cannot get a clear view of the overall plan of the analysis." ("Warum machen Sie es dem armen Sterblichen so schwer, indem Sie ihm die Technik Ihres Denkens vorenthalten? Es genügt doch dem denkenden Leser nicht, wenn er zwar die Richtigkeit Ihrer Gleichungen verifizieren aber den Plan der ganzen Untersuchung nicht überschauen kann.") See Einstein to David Hilbert, 30 May 1916, (CPAE 8, 293). In a letter to Paul Ehrenfest, Einstein expressed himself even more drastically with respect to what he perceived as the obscurity of Hilbert's heuristic: "Hilbert's description doesn't appeal to me. It is unnecessarily specialized as concerns "matter," unnecessarily complicated, and not above-board (=Gauss-like) in structure (feigning the super-human through camouflaging the methods)." ("Hilbert's Darstellung gefällt mir nicht. Sie ist unnötig speziell, was die 'Materie' anbelangt, unnötig kompliziert, nicht ehrlich (=Gaussisch) im Aufbau (Vorspiegelung des Übermenschen durch Verschleierung der Methoden).") See Einstein to Paul Ehrenfest, 24 May 1916, (CPAE 8, 288).

60 See (Proofs, 5ff.).

$$
\begin{aligned}
E = & \sum_{\mu, \nu, s, k, l}\left(H \frac{\partial \sqrt{g}}{\partial g^{\mu\nu}} g_s^{\mu\nu} + \sqrt{g}\frac{\partial H}{\partial g^{\mu\nu}} g_s^{\mu\nu} + \sqrt{g}\frac{\partial H}{\partial g_k^{\mu\nu}} g_{sk}^{\mu\nu} + \sqrt{g}\frac{\partial H}{\partial g_{kl}^{\mu\nu}} g_{skl}^{\mu\nu}\right) p^s \\
& - \sum_{\mu, \nu, s} (g^{\mu s} p_s^{\nu} + g^{\nu s} p_s^{\mu})[\sqrt{g}H]_{\mu\nu} \\
& + \sum_{\mu, \nu, s, k, l}\left(\frac{\partial \sqrt{g}H}{\partial g_k^{\mu\nu}} g_s^{\mu\nu} + \frac{\partial \sqrt{g}H}{\partial g_{kl}^{\mu\nu}} g_{sl}^{\mu\nu} - g_s^{\mu\nu}\frac{\partial}{\partial w_l}\frac{\partial \sqrt{g}H}{\partial g_{kl}^{\mu\nu}}\right) p_k^s,
\end{aligned}
\tag{39}
$$

and:

$$
\begin{aligned}
D^{(p)} = \sum_{\mu, \nu, s, k, l} \Bigg\{ & -\frac{\partial}{\partial w_k}\left(\sqrt{g}\frac{\partial H}{\partial g_k^{\mu\nu}}(g^{\mu s} p_s^{\nu} + g^{\nu s} p_s^{\mu})\right) \\
& + \frac{\partial}{\partial w_k}\left((p_s^{\nu} g^{\nu s} + p_s^{\mu} g^{\nu s})\frac{\partial}{\partial w_l}\left(\sqrt{g}\frac{\partial H}{\partial g_{kl}^{\mu\nu}}\right)\right) \\
& + \frac{\partial}{\partial w_l}\left(\sqrt{g}\frac{\partial H}{\partial g_{kl}^{\mu\nu}}\left(\frac{\partial p^{\mu\nu}}{\partial w_k} - g_{sk}^{\mu\nu} p^s\right)\right)\Bigg\}.
\end{aligned}
\tag{40}
$$

Hilbert had thus succeeded in splitting off a divergence term $D^{(p)}$ from the original expression $J^{(p)}$. By integrating over some region, $D^{(p)}$ could be converted into a surface term, and thus eliminated by demanding that p^s and its derivatives vanish on the boundary of that region.[61] So it would be possible to extract an energy expression from the remainder of $J^{(p)}$ if a way could be found to deal with the first term $\sum_{\mu, \nu} H \frac{\partial}{\partial g^{\mu\nu}} \sqrt{g} p^{\mu\nu}$.

Ultimately, the justification for choosing E as the energy expression depends, of course, on the possibility of a physical interpretation of this expression. As we shall see, for Hilbert this meant an interpretation in terms of Mie's theory. But, first of all, he had to show that E can be represented as a sum of divergences. For this purpose, Hilbert introduced yet another decomposition of $J^{(p)}$, derived from a generalization of (37). As we have indicated earlier, this equation may be identified as a special case of a "polarization" of the Lagrangian H with respect to the contravariant form of the metric $g^{\mu\nu}$: If one takes an arbitrary contravariant tensor $h^{\mu\nu}$, one obtains for the "first polar" of H:

$$
J^{(h)} = \sum_{\mu, \nu} \frac{\partial H}{\partial g^{\mu\nu}} h^{\mu\nu} + \sum_{\mu, \nu, k} \frac{\partial H}{\partial g_k^{\mu\nu}} h_k^{\mu\nu} + \sum_{\mu, \nu, k, l} \frac{\partial H}{\partial g_{kl}^{\mu\nu}} h_{kl}^{\mu\nu}.
\tag{41}
$$

Applying integration by parts to this expression, Hilbert obtained:

61 Die Grundlagen der Physik II, Ms. Vorlesung WS 1916/17, 186 ff. (D. Hilbert, Bibliothek des Mathematischen Seminars, Universität Göttingen); from here on "WS 1916/17 Lectures."

$$\sqrt{g}J^{(h)} = -\sum_{\mu,\nu} H\frac{\partial\sqrt{g}}{\partial g^{\mu\nu}}h^{\mu\nu} + \sum_{\mu,\nu}[\sqrt{g}H]_{\mu\nu}h^{\mu\nu} + D^{(h)}; \tag{42}$$

here

$$[\sqrt{g}H]_{\mu\nu} = \frac{\partial\sqrt{g}H}{\partial g^{\mu\nu}} - \sum_k \frac{\partial}{\partial w_k}\frac{\partial\sqrt{g}H}{\partial g_k^{\mu\nu}} + \sum_{k,l}\frac{\partial^2}{\partial w_k \partial w_l}\frac{\partial\sqrt{g}H}{\partial g_{kl}^{\mu\nu}} \tag{43}$$

is the Lagrangian variational derivative of H, the vanishing of which is the set of gravitational field equations; and:

$$\begin{aligned} D^{(h)} = &\sum_{\mu,\nu,k}\frac{\partial}{\partial w_k}\left(\frac{\partial\sqrt{g}H}{\partial g_k^{\mu\nu}}h^{\mu\nu}\right) + \sum_{\mu,\nu,k,l}\frac{\partial}{\partial w_k}\left(\frac{\partial\sqrt{g}H}{\partial g_{kl}^{\mu\nu}}h_l^{\mu\nu}\right) \\ &- \sum_{\mu,\nu,k,l}\frac{\partial}{\partial w_l}\left(h^{\mu\nu}\frac{\partial}{\partial w_k}\frac{\partial\sqrt{g}H}{\partial g_{kl}^{\mu\nu}}\right) \end{aligned} \tag{44}$$

i.e. another divergence expression. Obviously, $J^{(h)}$ turns into $J^{(p)}$ if one sets $h^{\mu\nu}$ equal to $p^{\mu\nu}$, thus yielding the desired alternative decomposition:

$$\sqrt{g}J^{(p)} = -\sum_{\mu,\nu} H\frac{\partial\sqrt{g}}{\partial g^{\mu\nu}}p^{\mu\nu} + D^{(h)}\big|_{h=p}. \tag{45}$$

Comparing (45) with (38), it becomes clear that E indeed can be written as a divergence, and thus represents a candidate for the energy expression. In the Proofs this conclusion is presented as one of two properties justifying this designation:

> Call the expression E the energy form. To justify this designation, I prove two properties that the energy form enjoys.
>
> If we substitute the tensor $p^{\mu\nu}$ for $h^{\mu\nu}$ in identity (6) [i.e. (42)] then, taken together with (9) [i.e. (39)] it follows, provided the gravitational equations (8) [i.e. (51) below] are satisfied:
>
> $$E = (D^{(h)})_{h=p} - D^{(p)} \tag{46}$$
>
> or
>
> $$\begin{aligned} E = \sum\Bigg\{&\frac{\partial}{\partial w_k}\left(\sqrt{g}\frac{\partial H}{\partial g_k^{\mu\nu}}g_s^{\mu\nu}p^s\right) - \frac{\partial}{\partial w_k}\left(\frac{\partial}{\partial w_l}\left(\sqrt{g}\frac{\partial H}{\partial g_{kl}^{\mu\nu}}\right)g_s^{\mu\nu}p^s\right) \\ &+ \frac{\partial}{\partial w_l}\left(\sqrt{g}\frac{\partial H}{\partial g_{kl}^{\mu\nu}}g_{sk}^{\mu\nu}p^s\right)\Bigg\}, \end{aligned} \tag{47}$$
>
> that is, we have the proposition:
>
> Proposition 1: In virtue of the gravitational equations the energy form E becomes a sum of differential quotients with respect to w_s, that is, it acquires the character of a divergence.[62]

Whereas (47) for an arbitrary H involves an arbitrary combination of electromagnetic and gravitational contributions, Hilbert makes an ansatz $H = K + L$ that allows him to separate these two contributions; in particular, to relate E to his result concerning the energy-momentum tensor of Mie's theory. Accordingly, at this point, he presumably introduces in a missing part of the Proofs (as he does in the corresponding part of Paper 1) the splitting of the Lagrangian (16), and introduces the condition that L not depend on $g_s^{\mu\nu}$.[63] Finally, he writes down explicitly the electromagnetic part of the energy:

> Because K depends only on $g^{\mu\nu}, g_s^{\mu\nu}, g_{lk}^{\mu\nu}$, therefore in ansatz (17) [i.e. (16)], due to (13) [i.e. (47)], the energy E can be expressed solely as a function of the gravitational potentials $g^{\mu\nu}$ and their derivatives, provided L is assumed to depend not on $g_s^{\mu\nu}$, but only on $g^{\mu\nu}, q_s, q_{sk}$. On this assumption, which we shall always make in the following, the definition of the energy (10) [i.e. (39)] yields the expression
>
> $$E = E^{(g)} + E^{(e)}, \tag{48}$$
>
> where the "gravitational energy" $E^{(g)}$ depends only on $g^{\mu\nu}$ and their derivatives, and the "electrodynamic energy" $E^{(e)}$ takes the form
>
> $$E^{(e)} = \sum_{\mu,\nu,s} \frac{\partial\sqrt{g}L}{\partial g^{\mu\nu}}(g_s^{\mu\nu}p^s - g^{\mu s}p_s^{\nu} - g^{\nu s}p_s^{\mu}), \tag{49}$$
>
> which proves to be *a general invariant multiplied by* $\sqrt{g}$.[64]

(The term in parentheses in equation (49) is $p^{\mu\nu}$, the Lie derivative of the contravariant metric with respect to the vector p^s.)

Hilbert's final expression (49) satisfies what we called his "physical criterion" for finding a definition of the energy since the term $\frac{\partial}{\partial g^{\mu\nu}}\sqrt{g}L$ corresponds—apart from the factor -2 —to the left-hand side of (36), and thus to Mie's energy momentum ten-

---

62 "Der Ausdruck E heiße die Energieform. Um diese Bezeichnung zu rechtfertigen, beweise ich zwei Eigenschaften, die der Energieform zukommen.
Setzen wir in der Identität (6) [i.e. (42)] für $h^{\mu\nu}$ den Tensor $p^{\mu\nu}$ ein, so folgt daraus zusammen mit (9) [(39)], sobald die Gravitationsgleichungen (8) erfüllt sind: [(46); (12) in the original text] or [(47); (13) in the original text] d. h. es gilt der Satz:
Satz 1. Die Energieform E wird vermöge der Gravitationsgleichungen einer Summe von Differentialquotienten nach w_s gleich, d. h. sie erhält Divergenzcharakter." See (Proofs, 6).

63 Compare (Hilbert 1916, 402) with (Proofs, 8), and see the discussion in "Einstein Equations and Hilbert Action …" (in this volume).

64 "Da K nur von $g^{\mu\nu}, g_s^{\mu\nu}, g_{lk}^{\mu\nu}$ abhängt, so läßt sich beim Ansatz (17) die Energie E wegen (13) lediglich als Funktion der Gravitationspotentiale $g^{\mu\nu}$ und deren Ableitungen ausdrücken, sobald wir L nicht von $g_s^{\mu\nu}$, sondern nur von $g^{\mu\nu}, q_s, q_{sk}$ abhängig annehmen. Unter dieser Annahme, die wir im Folgenden stets machen, liefert die Definition der Energie (10) den Ausdruck [(48); (18) in the original text] wo die "Gravitationsenergie" $E^{(g)}$ nur von $g^{\mu\nu}$ und deren Ableitungen abhängt und die "elektrodynamische Energie" $E^{(e)}$ die Gestalt erhält [(49); (19) in the original text] in der sie sich als *eine mit* $\sqrt{g}$ *multiplizierte allgemeine Invariante* erweist." (Proofs, 8)

sor. Hilbert's definition of energy had thus been given a "physical justification" in terms of Mie's theory. But—apart from merely formal similarities—its relation to energy-momentum conservation in classical and special-relativistic theories remains entirely unclear. In the Proofs, as we shall see below, Hilbert's energy expression served still another and even more important function, that of determining admissible coordinate systems.

3.4 Hilbert's Revision of Mie's Program and the Roots of his Leitmotiv in Einstein's Work

Apparently Hilbert was convinced that the relation he established between the variational derivative of the Lagrangian and the energy-momentum tensor (see (36)) singled out Mie's theory as having a special relation to the theory of gravitation.[65] In fact, as we have seen, this conclusion is only justified insofar as one imposes on the electrodynamic term in the Lagrangian the condition that it does not depend on $g_s^{\mu\nu}$. Nevertheless, this result apparently suggested to Hilbert that gravitation may be the more fundamental physical process and that it might be possible to conceive of electromagnetic phenomena as "effects of gravitation."[66] Such an interpretation, which was in line with the reductionist perspective implied by his understanding of the axiomatization of physics, led to a revision of Mie's original aim of basing all of physics on electromagnetism.

In the light of this possibility, the third point of Hilbert's initial research program, the question of the number of independent equations in a generally-covariant theory, must have taken on a new and increased significance. Einstein's hole argument, when applied to Hilbert's formalism, suggests that the fourteen generally-covariant field equations for the 14 gravitational and electromagnetic potentials do not have a unique solution for given boundary values. Consequently, 4 identities must exist between the 14 field equations; and 4 additional, non-covariant equations would be required in order to assure a unique solution; and if these 4 identities were somehow equivalent to the 4 equations for the electromagnetic potentials, then the latter could be considered as a consequence of the 10 gravitational equations by virtue of the unique properties of a generally-covariant variational principle, and Hilbert would indeed be entitled to claim that electromagnetism is an effect of gravitation.

As we have seen, the non-uniqueness of solutions to generally-covariant field equations and the conclusion that such field equations must obey 4 identities, are both issues raised by Einstein in his publications of 1913/14. These writings and his 1915 Göttingen lectures, which Hilbert attended, offered rich sources of information about Einstein's theory. In addition the physicist Paul Hertz, then a participant in the group

---

65 In fact, this relation between the special-relativistic stress-energy tensor and the variational derivative of the general-relativistic generalization of a Lagrangian giving rise to this stress-energy tensor is quite general, as was pointed out many years later in (Rosenfeld 1940, 1–30; and Belinfante 1939, 887). See also (Vizgin 1989, 304; 1994).

66 See (Proofs, 3) and (Hilbert 1916, 397).

centered around Hilbert in Göttingen, may also have kept Hilbert informed about Einstein's thinking on these issues. For example, in a letter to Hertz of August 1915, Einstein raised the problem of solving hyperbolic partial differential equations for arbitrary boundary values and discussed the necessity of introducing four additional equations to restore causality for a set of generally-covariant field equations.[67]

Einstein's treatment of these issues thus forms the background to the crucial theorem, on which Hilbert's entire approach is based, his *Leitmotiv*, labelled "Theorem I" in the Proofs:

> The guiding motive for setting up the theory is given by the following theorem, the proof of which I shall present elsewhere.
>
> Theorem I. If J is an invariant under arbitrary transformations of the four world parameters, containing n quantities and their derivatives, and if one forms from
>
> $$\delta \int J \sqrt{g} d\tau = 0 \tag{50}$$
>
> the n variational equations of Lagrange with respect to each of the n quantities, then in this invariant system of n differential equations for the n quantities there are always four that are a consequence of the remaining $n-4$ —in the sense that, among the n differential equations and their total derivatives, there are always four linear and mutually independent combinations that are satisfied identically.[68]

For a Lagrangian H depending on the gravitational and the electrodynamic potentials and their derivatives, Hilbert derived 10 field equations for the gravitational potentials $g^{\mu\nu}$ and 4 for the electrodynamic potentials q_s from such a variational principle (50):

$$\frac{\partial \sqrt{g} H}{\partial g^{\mu\nu}} = \sum_k \frac{\partial}{\partial w_k} \frac{\partial \sqrt{g} H}{\partial g_k^{\mu\nu}} - \sum_{k,l} \frac{\partial^2}{\partial w_k \partial w_l} \frac{\partial \sqrt{g} H}{\partial g_{kl}^{\mu\nu}}, \quad (\mu, \nu = 1, 2, 3, 4), \tag{51}$$

$$\frac{\partial \sqrt{g} H}{\partial q_h} = \sum_k \frac{\partial}{\partial w_k} \frac{\partial \sqrt{g} H}{\partial q_{hk}}, \quad (h = 1, 2, 3, 4). \tag{52}$$

---

67 Einstein to Paul Hertz, 22 August 1915, (CPAE 8, 163–164). See (Howard and Norton 1993) for an extensive historical discussion.

68 "Das Leitmotiv für den Aufbau der Theorie liefert der folgende mathematische Satz, dessen Beweis ich an einer anderen Stelle darlegen werde.
Theorem I. Ist J eine Invariante bei beliebiger Transformation der vier Weltparameter, welche n Größen und ihre Ableitungen enthält, und man bildet dann aus [(50)] in Bezug auf jene n Größen die n Lagrangeschen Variationsgleichungen, so sind in diesem invarianten System von n Differentialgleichungen für die n Größen stets vier eine Folge der $n-4$ übrigen — in dem Sinne, daß zwischen den n Differentialgleichungen und ihren totalen Ableitungen stets vier lineare, von einander unabhängige Kombinationen identisch erfüllt sind." (Proofs, 2–3) See (Hilbert 1916, 396–397). See (Rowe 1999) for a discussion of the debate on Hilbert's Theorem I among Göttingen mathematicians.

In both the Proofs and Paper 1, Hilbert erroneously claimed that one can consider the last four equations to be a consequence of the 4 identities that must hold, according to his Theorem I, between the 14 differential equations:

> Let us call equations (4) [i.e. (51)] the fundamental equations of gravitation, and equations (5) [i.e. (52)] the fundamental electrodynamic equations, or generalized Maxwell equations. Due to the theorem stated above, the four equations (5) [i.e. (52)] can be viewed as a consequence of equations (4) [i.e. (51)]; that is, because of that mathematical theorem we can immediately assert the claim *that in the sense explained above electrodynamic phenomena are effects of gravitation*. I regard this insight as the simple and very surprising solution of the problem of Riemann, who was the first to search for a theoretical connection between gravitation and light.[69]

We shall come back to this claim later, in connection with Hilbert's proof of a special case of Theorem I.

The fact that Hilbert did not give a proof of this theorem makes it difficult to assess its heuristic roots. No doubt, of course, some of these roots lay in Hilbert's extensive mathematical knowledge, in particular, of the theory of invariants. But the lack of a proof in Paper 1, as well as the peculiar interpretation of it in the Proofs, make it plausible that the theorem also had roots in Einstein's hole argument on the ambiguity of solutions to generally-covariant field equations.

In fact, in the Proofs, Hilbert placed the implications of Theorem I for his field theory in the context of the problem of causality, as Einstein had done for the hole argument. But while the hole argument was formulated in terms of a boundary value problem for a closed hypersurface, Hilbert posed the question of causality in terms of an initial value problem for an open one, thus adapting it to Cauchy's theory of systems of partial differential equations:

> Since our mathematical theorem shows that the axioms I and II [essentially amounting to the variational principle (50), see the discussion below] considered so far can produce only ten essentially independent equations; and since, on the other hand, if general invariance is maintained, more than ten essentially independent equations for the 14 potentials $g_{\mu\nu}$, q_s are not at all possible; therefore—provided that we want to retain the determinate character of the basic equation of physics corresponding to Cauchy's theory of differential equations— the demand for four further non-invariant equations in addition to (4) [i.e. (51)] and (5) [i.e. (52)] is imperative.[70]

Hilbert's counting of needed equations closely parallels Einstein's: the number of field equations (10 in Einstein's case and 14 in Hilbert's) plus 4 coordinate restrictions to make sure that causality is preserved. Since Hilbert, in contrast to Einstein,

69 "Die Gleichungen (4) mögen die Grundgleichungen der Gravitation, die Gleichungen (5) die elektrodynamischen Grundgleichungen oder die verallgemeinerten Maxwellschen Gleichungen heißen. Infolge des oben aufgestellten Theorems können die vier Gleichungen (5) als eine Folge der Gleichungen (4) angesehen werden, d. h. wir können unmittelbar wegen jenes mathematischen Satzes die Behauptung aussprechen, *daß in dem bezeichneten Sinne die elektrodynamischen Erscheinungen Wirkungen der Gravitation sind*. In dieser Erkenntnis erblicke ich die einfache und sehr überraschende Lösung des Problems von Riemann, der als der Erste theoretisch nach dem Zusammenhang zwischen Gravitation und Licht gesucht hat." (Proofs, 3; Hilbert 1916, 397–398)

had started from a generally-covariant variational principle, he obtained, in addition, 4 identities that, he claimed, imply the electrodynamic equations (52).

Additional evidence for our conjecture that Einstein's hole argument was one of the roots of Hilbert's theorem (and thus of its later elaboration by Emmy Noether) is provided by other contemporary writings of Hilbert, which will be discussed below in connection with Hilbert's second paper, in which the problem of causality is addressed explicitly.[71]

3.5 Energy-Momentum Conservation and Coordinate Restrictions

As we shall see in this section, the Proofs show that Hilbert was convinced that causality requires four supplementary non-covariant equations to fix the admissible coordinate systems. In identifying these coordinate restrictions, he again followed closely in Einstein's tracks. As did the latter, Hilbert invoked energy-momentum conservation in order to justify physically the choice of a preferred reference frame. After formulating his version of energy-momentum conservation, he introduced the following axiom:

> Axiom III (axiom of space and time). *The spacetime coordinates are those special world parameters for which the energy theorem* (15) [i.e. (57) below] *is valid*.
>
> According to this axiom, space and time in reality provide a special labeling of the world's points such that the energy theorem holds.
>
> Axiom III implies the existence of equations (16) [$d^{(g)}\sqrt{g}H\ /\ dw_s\ =\ 0$]: these four differential equations (16) complete the gravitational equations (4) [i.e. (51)] to give a system of 14 equations for the 14 potentials $g_{\mu\nu}$, q_s, *the system of fundamental equations of physics*. Because of the agreement in number between equations and potentials to be determined, the principle of causality for physical processes is also guaranteed, revealing to us the closest connection between the energy theorem and the principle of causality, since each presupposes the other.[72]

The strategy Hilbert followed to extract these coordinate restrictions from the requirement of energy conservation closely followed that of Einstein's *Entwurf* theory of 1913/14. Even before he developed the hole argument, energy-momentum conservation played a crucial role in justifying the lack of general covariance of his

---

70 "Indem unser mathematisches Theorem lehrt, daß die bisherigen Axiome I und II für die 14 Potentiale nur zehn wesentlich von einander unabhängige Gleichungen liefern können, andererseits bei Aufrechterhaltung der allgemeinen Invarianz mehr als zehn wesentlich unabhängige Gleichungen für die 14 Potentiale $g_{\mu\nu}$, q_s garnicht möglich sind, so ist, wofern wir der Cauchyschen Theorie der Differentialgleichungen entsprechend den Grundgleichungen der Physik den Charakter der Bestimmtheit bewahren wollen, die Forderung von vier weiteren zu (4) und (5) hinzutretenden nicht invarianten Gleichungen unerläßlich." (Proofs, 3–4)

71 See, e.g., his SS 1916 Lectures, in particular p. 108, as well as an undated typescript preserved at Göttingen, in SUB Cod. Ms. 642, entitled *Das Kausalitätsprinzip in der Physik*, henceforth cited as the "Causality Lecture." Page 4 of this typescript, describing a construction equivalent to Einstein's hole argument, is discussed below.

gravitational field equations. He was convinced that energy-momentum conservation actually required a restriction of the covariance group.[73] An the beginning of 1914, after having formulated the hole argument, he described the connection between coordinate restrictions and energy-momentum conservation in the *Entwurf* theory as follows:

> Once we have realized that an acceptable theory of gravitation necessarily implies a specialization of the coordinate system, it is also easily seen that the gravitational equations given by us are based upon a special coordinate system. Differentiation of equations (II) with respect to x_ν [the field equations in the form
>
> $$\sum_{\alpha\beta\mu}\frac{\partial}{\partial x_\alpha}\left(\sqrt{-g}\gamma_{\alpha\beta}g_{\sigma\mu}\frac{\partial\gamma_{\mu\nu}}{\partial x_\beta}\right) = \kappa(\mathfrak{T}_{\sigma\nu}+\mathfrak{t}_{\sigma\nu})]$$
>
> and summation over ν, and taking into account equations (III), [the conservations equations in the form
>
> $$\sum_\nu\frac{\partial}{\partial x_\nu}(\mathfrak{T}_{\sigma\nu}+\mathfrak{t}_{\sigma\nu}) = 0\,] \tag{53}$$
>
> yields the relations (IV)
>
> $$\left[\sum_{\alpha\beta\mu\nu}\frac{\partial^2}{\partial x_\nu\partial x_\alpha}\left(\sqrt{-g}\gamma_{\alpha\beta}g_{\sigma\mu}\frac{\partial\gamma_{\mu\nu}}{\partial x_\beta}\right)= 0\right], \tag{54}$$
>
> that is, four differential conditions for the quantities $g_{\mu\nu}$, which we write in the abbreviated form
>
> $$B_\sigma = 0. \tag{55}$$
>
> These quantities B_σ do not form a generally-covariant vector, as will be shown in §5. From this one can conclude that the equations $B_\sigma = 0$ represent a real restriction on the choice of coordinate system.[74]

In a later 1914 paper, Einstein discussed the physical significance and the transformation properties of the gravitational energy-momentum term $\mathfrak{t}_\sigma^\nu$:

> According to the considerations of §10, the equations (42 c) [i.e. (53)] represent the conservation laws of momentum and energy for matter and gravitational field combined. The $\mathfrak{t}_\sigma^\nu$ are those quantities, related to the gravitational field, which are analogies in physical

72 "Axiom III (Axiom von Raum und Zeit). *Die Raum-Zeitkoordinaten sind solche besonderen Weltparameter, für die der Energiesatz (15) gültig ist.*
Nach diesem Axiom liefern in Wirklichkeit Raum und Zeit eine solche besondere Benennung der Weltpunkte, daß der Energiesatz gültig ist.
Das Axiom III hat das Bestehen der Gleichungen (16) zur Folge: diese vier Differentialgleichungen (16) vervollständigen die Gravitationsgleichungen (4) zu einem System von 14 Gleichungen für die 14 Potentiale $g^{\mu\nu}$, q_s: *dem System der Grundgleichungen der Physik*. Wegen der Gleichzahl der Gleichungen und der zu bestimmenden Potentiale ist für das physikalische Geschehen auch das Kausalitätsprinzip gewährleistet, und es enthüllt sich uns damit der engste Zusammenhang zwischen dem Energiesatz und dem Kausalitätsprinzip, indem beide sich einander bedingen." (Proofs, 7)

73 See, e.g., (Einstein 1913, 1258).

> interpretation to the components $\mathfrak{T}_\sigma^\nu$ of the energy tensor (V-Tensor) [i.e. tensor density]. It is to be emphasized that the $\mathfrak{t}_\sigma^\nu$ do not have tensorial covariance under arbitrary admissible [coordinate] transformations but only under linear transformations. Nevertheless, we call ($\mathfrak{t}_\sigma^\nu$) the energy tensor of the gravitational field.[75]

Similarly, Hilbert notes that his energy-form is invariant with respect to linear transformations; he shows that E can be decomposed with respect to the vector p^j as follows (Proofs, 6):

$$E = \sum_s e_s p^s + \sum_{s,l} e_s^l p_l^s \tag{56}$$

where e_s and e_s^l are independent of p^j. If one compares this expression with Einstein's (53), then the analogy between the two suggests that the two-index object e_s^l should play the same role in Hilbert's theory as does the total energy-momentum tensor in Einstein's theory, satisfying a divergence equation of the form:

$$\sum_l \frac{\partial e_s^l}{\partial w_l} = 0. \tag{57}$$

Hilbert shows that this equation holds only if e_s vanishes, in which case:

$$E = \sum_{s,l} e_s^l p_l^s. \tag{58}$$

This equation can be related to energy conservation; Hilbert calls this the "normal form" of the energy. The fact that the last two equations imply each other was, for Hilbert, apparently a decisive reason for calling E the energy form. Indeed, this equivalence is the subject of his second theorem about the energy-form. Although the relevant part of the Proofs is missing,[76] Hilbert's theorem and its proof can be reconstructed:

---

74 "Nachdem wir so eingesehen haben, daß eine brauchbare Gravitationstheorie notwendig einer Spezialisierung des Koordinatensystems bedarf, erkennen wir auch leicht, daß bei den von uns angegebenen Gravitationsgleichungen ein spezielles Koordinatensystem zugrunde liegt. Aus den Gleichungen (II) folgen nämlich durch Differentiation nach x_ν und Summation über ν unter Berücksichtigung der Gleichungen (III) die Beziehungen (IV) also vier Differentialbedingungen für die Größen $g_{\mu\nu}$, welche wir abgekürzt $B_\sigma = 0$ schreiben wollen.
Diese Größen B_σ bilden, wie in §5 gezeigt ist, keinen allgemein-kovarianten Vektor. Hieraus kann geschlossen werden, daß die Gleichungen $B_\sigma = 0$ eine wirkliche Bedingung für die Wahl des Koordinatensystems darstellen." (Einstein and Grossmann 1914, 218–219)

75 "Die Gleichungen (42 c) drücken nach den in §10 gegebenen Überlegungen die Erhaltungssätze des Impulses und der Energie für Materie und Gravitationsfeld zusammen aus. $\mathfrak{t}_\sigma^\nu$ sind diejenigen auf das Gravitationsfeld bezüglichen Größen, welche den Komponenten $\mathfrak{T}_\sigma^\nu$ des Energietensors (V-Tensors) [i.e. tensor density] der physikalischen Bedeutung nach analog sind. Es sei hervorgehoben, daß die $\mathfrak{t}_\sigma^\nu$ nicht beliebigen berechtigten, sondern nur linearen Transformationen gegenüber Tensorkovarianz besitzen; trotzdem nennen wir ($\mathfrak{t}_\sigma^\nu$) den Energietensor des Gravitationsfeldes." (Einstein 1914b, 1077)

76 The top portion of the Proofs, p. 7, is missing.

Theorem 2 must have asserted that:

$$e_s = \frac{\partial e_s^l}{\partial w_l} \tag{59}$$

This assertion is easily proven by following the lines indicated in the surviving portion of Hilbert's argument. From (38) and (56) it follows that:

$$\sqrt{g}J^{(p)} + H\frac{\partial\sqrt{g}}{\partial g^{\mu\nu}}p^{\mu\nu} = e_s p^s + e_s^l p_l^s + D^{(p)}, \tag{60}$$

which can be rewritten as:

$$\sqrt{g}J^{(p)} + H\frac{\partial\sqrt{g}}{\partial g^{\mu\nu}}p^{\mu\nu} = \left(e_s - \frac{\partial e^l{}_s}{\partial w_l}\right)p^s + \overline{D^{(p)}}, \tag{61}$$

where $\overline{D^{(p)}}$ is still a divergence. If now the integral over a region Ω, on the boundary of which p^s and its first derivative vanish, is taken on both sides, then the surface terms vanish. Thus one obtains in view of (42):

$$\int_\Omega [\sqrt{g}H]_{\mu\nu}p^{\mu\nu}dx^4 = \int_\Omega \left(e_s - \frac{\partial e^l{}_s}{\partial w_l}\right)p^s(dx^4). \tag{62}$$

But the left-hand side vanishes when the gravitational field equations hold, and p^s is an arbitrary vector field, from which (59) follows.

Theorem 2 provides Hilbert with the desired coordinate restrictions:

> This theorem shows that the divergence equation corresponding to the energy theorem of the old theory
>
> $$\sum_l \frac{\partial e_s^l}{\partial w_l} = 0 \tag{63}$$
>
> holds if and only if the four quantities e_s vanish ... .[77]

After these preparations, Hilbert introduces Axiom III, quoted at the beginning of this section, which establishes a distinction between the arbitrary world parameters w_l and the restricted class of coordinates that constitute "a spacetime reference system." In fact, the latter are those world parameters satisfying the coordinate restrictions $e_s = 0$ following from Hilbert's energy condition. In analogy to the "justified coordinate transformations" of Einstein's 1913/14 theory leading from one "adapted

77 "Dieser Satz zeigt, daß die dem Energiesatz der alten Theorie entsprechende Divergenzgleichung [(63); (15) in the original text] dann und nur dann gelten kann, wenn die vier Größen e_s verschwinden ..." (Proofs, 7).

coordinate system" to another, Hilbert introduced spacetime transformations that lead from one "normal form" of the energy to another:

> To the transition from one spacetime reference system to another one corresponds the transformation of the energy form from one so-called "normal form"
>
> $$E = \sum_{s,l} e_s^l p_l^s \tag{64}$$
>
> to another normal form.[78]

The claim that Hilbert's introduction of coordinate restrictions was guided by the goal of recovering the ordinary divergence form of energy-momentum conservation is supported by his later use of this argument in a discussion with Felix Klein. In a letter to Hilbert, Klein recounted how, at a meeting of the Göttingen Academy, he had argued that, for the energy balance of a field, one should take into account only the energy tensor of matter (including that of the electromagnetic field) without ascribing a separate energy-momentum tensor to the gravitational field.[79] This suggestion was taken up by Carl Runge, who had given an expression for energy-momentum conservation that, in his letter to Hilbert, Klein called "regular" and found similar to what happens in the "elementary theory."[80] Starting from an expression for the covariant divergence of the stress-energy tensor:

$$\sum_{\mu\nu}\left(\sqrt{g}T_{\mu\nu}g_{\sigma}^{\mu\nu} + 2\frac{\partial}{\partial w_\nu}(\sqrt{g}T_{\mu\sigma}g^{\mu\nu})\right) = 0 \qquad \sigma = 1, 2, 3, 4 \tag{65}$$

Runge obtained his "regular" expression by imposing the four equations:

$$\sum_{\mu\nu}\sqrt{g}T_{\mu\nu}g_{\sigma}^{\mu\nu} = 0, \tag{66}$$

thus specifying a preferred class of coordinate systems. In his response, Hilbert sent Klein three pages of the Proofs to show that he had anticipated Runge's line of reasoning:

> I send you herewith my first proofs [footnote: Please kindly return these to me as I have no other record of them.] (3 pages) of my first communication, in which I also implemented Runge's ideas; in particular with theorem 1, p. 6, in which the divergence character of the energy is proven. I later omitted the whole thing as the thing did not seem to me to be fully mature. I would be very pleased if progress could now be made. For this it is necessary to retrieve the old energy conservation laws in the limiting case of Newtonian theory.[81]

---

78 "Dem Übergang von einem Raum-Zeit-Bezugssystem zu einem anderen entspricht die Transformation der Energieform von einer sogenannten "Normalform" [(64)] auf eine andere Normalform." (Proofs, 7)

79 Felix Klein to David Hilbert, 5 March 1918, (Frei 1985, 142–143).

80 For a discussion of Runge's work, see (Rowe 1999).

Hilbert's final sentence confirms that the recovery of the familiar form of energy conservation was his goal. However, at the time of the Proofs, it was clearly not his aim to eliminate the energy-momentum expression of the gravitational field from the energy balance, as the above reference to Runge might suggest. On the contrary, as we have seen above (see (48)), Hilbert followed Einstein in attempting to treat the contributions to the total energy from the electromagnetic and the gravitational parts on an equal footing.

In summary, Hilbert's first steps in the realization of his research program were the derivation of what he regarded as the unique relation between the variational derivative of Mie's Lagrangian and Mie's energy momentum tensor, and the formulation of a theorem, by means of which he hoped to show that the electromagnetic field equations follow from the gravitational ones. Albeit problematic from a modern perspective, these steps become understandable in the context of Hilbert's application of his axiomatic approach to Einstein's non-covariant theory of gravitation and Mie's theory of matter. These first steps in turn shaped Hilbert's further research. They effected a change of perspective from viewing electrodynamics and gravitation on an equal footing to his vision of deriving electromagnetism from gravitation. As a consequence, the structure of Hilbert's original, non-covariant theory, in spite of the covariance of Hilbert's gravitational equations and the different physical interpretation that he gave to his equations, is strikingly similar to that of Einstein's 1913/14 *Entwurf* theory of gravitation.

3.6 Electromagnetism as an Effect of Gravitation: The Core of Hilbert's Theory

Now we come to the part of Hilbert's program that today is often considered to contain his most important contributions to general relativity: the contracted Bianchi identities and a special case of Noether's theorem. We shall show that, in the original version of Hilbert's theory, these mathematical results actually constituted part of a different physical framework that also affected their interpretation. In a later section, we shall see how these results were transformed, primarily due to the work of Hendrik Antoon Lorentz and Felix Klein, into constituents of general relativity. In the hindsight of general relativity, it appears as if Hilbert first derived the contracted Bianchi identities, applied them to the gravitational field equations with an electromagnetic source-term, and then showed that the electrodynamic variables necessarily satisfy the Maxwell equations. This last result, however, is valid only under addi-

81 "Anbei schicke ich Ihnen meine erste Korrektur [footnote: Bitte dieselbe mir wieder freundlichst zustellen zu wollen, da ich sonst keine Aufzeichnungen habe.] (3 Blätter) meiner ersten Mitteilung, in der ich gerade die Ideen von Runge auch ausgeführt hatte; insbesondere auch mit Satz l, S. 6, in dem der Divergenzcharakter der Energie bewiesen wird. Ich habe aber die ganze Sache später unterdrückt, weil die Sache mir nicht reif erschien. Ich würde mich sehr freuen, wenn jetzt der Fortschritt gelänge. Dazu ist aber nötig im Grenzfalle zur Newtonschen Theorie die alten Energiesätze wiederzufinden." Tilman Sauer suggested that the pages sent to Klein were the three sheets of the Proofs bearing Roman numbers I, II, and III, see (Sauer 1999, 544).

tional assumptions that run counter to Mie's program. From the point of view of general relativity, Hilbert obtained Maxwell's equations as a consequence of the integrability conditions for the gravitational field equations with electromagnetic source term, as if he had treated a special case of Einstein's equations and expressed certain of their general properties in terms of this special case. From Hilbert's point of view, however, he had derived the electrodynamic equations as a consequence of the gravitational ones; his derivation was closely interwoven with other results of his theory that pointed to electromagnetism as an effect of gravitation. For him, the equation, on the basis of which he argued that electrodynamics is a consequence of gravitation, was a result of four ingredients, two of which are other links between gravitation and electrodynamics, and all of which are based on his generally-covariant variational principle:

- a general theorem corresponding to the contracted Bianchi identities,
- the field equations following from the variational principle,
- the relation between Mie's energy-momentum tensor and the variational derivative of the Lagrangian, and
- the way in which the derivatives of the electrodynamic potentials enter Mie's Lagrangian.

In the Proofs, the general theorem is:

> Theorem III. If J is an invariant depending only on the $g^{\mu\nu}$ and their derivatives and if, as above, the variational derivatives of $\sqrt{g}J$ with respect to $g^{\mu\nu}$ are denoted by $[\sqrt{g}J]_{\mu\nu}$, then the expression — in which $h^{\mu\nu}$ is understood to be any contravariant tensor —
>
> $$\frac{1}{\sqrt{g}}\sum_{\mu,\nu}[\sqrt{g}J]_{\mu\nu}h^{\mu\nu} \tag{67}$$
>
> represents an invariant; if in this sum we substitute in place of $h^{\mu\nu}$ the particular tensor $p^{\mu\nu}$ and write
>
> $$\sum_{\mu,\nu}[\sqrt{g}J]_{\mu\nu}p^{\mu\nu} = \sum_{s,l}(i_s p^s + i_s^l p_l^s), \tag{68}$$
>
> where then the expressions
>
> $$\begin{aligned} i_s &= \sum_{\mu,\nu}[\sqrt{g}J]_{\mu\nu}g_s^{\mu\nu}, \\ i_s^l &= -2\sum_{\mu}[\sqrt{g}J]_{\mu s}g^{\mu l} \end{aligned} \tag{69}$$
>
> depend only on the $g^{\mu\nu}$ and their derivatives, then we have
>
> $$i_s = \sum_l \frac{\partial i_s^l}{\partial w_l} \tag{70}$$
>
> in the sense, that this equation is identically fulfilled for all arguments, that is for the $g^{\mu\nu}$ and their derivatives.[82]

Here, (68) follows from an explicit calculation taking into account the definition of $p^{\mu\nu}$; the identity (70) follows if in analogy to (61) one rewrites (68) as:

$$[\sqrt{g}J]_{\mu\nu}p^{\mu\nu} = \left(i_s - \frac{\partial i_s^l}{\partial w_l}\right)p^s + \frac{\partial}{\partial w_l}(i_s^l p^s), \tag{71}$$

and, as in the earlier derivation, carries out a surface integration. Theorem III, in the form of (70), thus corresponds to the contracted Bianchi identities.

Hilbert next applies Theorem III to the Lagrangian $H = K + L$ using his knowledge about its electrodynamic part (see the last two "ingredients" listed above) in order to extract the electrodynamic equations from the identity for L that corresponds to (70). From a modern point of view, it is remarkable that Hilbert did not consider the physical significance of this identity for the gravitational part K of the Lagrangian, but only for the electrodynamic part. For Hilbert, however, this was natural; presumably he was convinced, on the basis of Theorem I, that generally-covariant equations for gravitation are impossible as a "stand-alone" theory. Consequently, it simply made no sense to interpret the gravitational part of these equations by itself.

Assuming the split of the Lagrangian into $K + L$, the gravitational and electrodynamic parts as in (16), he rewrites (51) as:[83]

$$[\sqrt{g}K]_{\mu\nu} + \frac{\partial\sqrt{g}L}{\partial g^{\mu\nu}} = 0. \tag{72}$$

He next applies (69) to the invariant K:

$$i_s = \sum_{\mu,\nu}[\sqrt{g}K]_{\mu\nu}g_s^{\mu\nu}, \tag{73}$$

and

$$i_s^l = -2\sum_{\mu}[\sqrt{g}K]_{\mu s}g^{\mu l}, \quad (\mu = 1, 2, 3, 4). \tag{74}$$

From the modern point of view, it would be natural to invoke the identity (70) in order to derive its implications for the source term of the gravitational field equations, i.e., the second term of (72) in Hilbert's notation. In this way, one would obtain an integrability condition for the gravitational field equations that can be interpreted as representing energy-momentum conservation.

---

82 "Theorem III. Wenn J eine nur von den $g^{\mu\nu}$ und deren Ableitungen abhängige Invariante ist, und, wie oben, die Variationsableitungen von $\sqrt{g}J$ bezüglich $g^{\mu\nu}$ mit $[\sqrt{g}J]_{\mu\nu}$ bezeichnet werden, so stellt der Ausdruck — unter $h^{\mu\nu}$ irgend einen kontravarianten Tensor verstanden — [(67)] eine Invariante dar; setzen wir in dieser Summe an Stelle von $h^{\mu\nu}$ den besonderen Tensor $p^{\mu\nu}$ ein und schreiben [(68)] wo alsdann die Ausdrücke [(69)] lediglich von den $g^{\mu\nu}$ und deren Ableitungen abhängen, so ist [(70)] in der Weise, daß diese Gleichung identisch für alle Argumente, nämlich die $g^{\mu\nu}$ und deren Ableitungen, erfüllt ist." (Proofs, 9; Hilbert 1916, 399)

83 See (Proofs, 11; Hilbert 1916, 405).

Hilbert proceeded differently, using Theorem III to further elaborate what he considered his crucial insight into the relation between Mie's energy-momentum tensor and the variational derivative of L. Consequently he focussed on (36), from which he attempted to extract the equations for the electromagnetic field. In fact, the left-hand side of this equation can (in view of (72) and (74)) be rewritten as $-i_\nu^m$. Consequently, differentiating the right-hand side of (36) with respect to w_m and summing over m, Theorem III yields:

$$\begin{aligned} i_\nu = {} & \sum_m \frac{\partial}{\partial w_m}\Big(-\sqrt{g}L\delta_\nu^m + \frac{\partial\sqrt{g}L}{\partial q_m}q_\nu + \sum_s \frac{\partial\sqrt{g}L}{\partial M_{sm}}M_{s\nu}\Big) \\ = {} & -\frac{\partial\sqrt{g}L}{\partial w_\nu} + \sum_m \Big\{ q_\nu \frac{\partial}{\partial w_m}\Big([\sqrt{g}L]_m + \sum_s \frac{\partial}{\partial w_s}\frac{\partial\sqrt{g}L}{\partial q_{ms}}\Big) \\ & + q_{\nu m}\Big([\sqrt{g}L]_m + \sum_s \frac{\partial}{\partial w_s}\frac{\partial\sqrt{g}L}{\partial q_{ms}}\Big)\Big\} \\ & + \sum_s \Big([\sqrt{g}L]_s - \frac{\partial\sqrt{g}L}{\partial q_s}\Big)M_{s\nu} + \sum_{s,m}\frac{\partial\sqrt{g}L}{\partial M_{sm}}\frac{\partial M_{s\nu}}{\partial w_m}, \end{aligned} \tag{75}$$

where use has been made of:

$$\frac{\partial\sqrt{g}L}{\partial q_m} = [\sqrt{g}L]_m + \sum_s \frac{\partial}{\partial w_s}\frac{\partial\sqrt{g}L}{\partial q_{ms}} \tag{76}$$

and

$$-\sum_m \frac{\partial}{\partial w_m}\frac{\partial\sqrt{g}L}{\partial q_{sm}} = [\sqrt{g}L]_s - \frac{\partial\sqrt{g}L}{\partial q_s}. \tag{77}$$

Here $[\sqrt{g}L]_h$ denotes the Lagrangian derivative of $\sqrt{g}L$ with respect to the electrodynamic potentials q_h:

$$[\sqrt{g}L]_h = \frac{\partial\sqrt{g}L}{\partial q_h} - \sum_k \frac{\partial}{\partial w_k}\frac{\partial\sqrt{g}L}{\partial q_{hk}}; \tag{78}$$

the vanishing of which constitutes the electromagnetic field equations. At this point Hilbert makes use of the last ingredient, the special way in which the derivatives of the potentials enter Mie's Lagrangian. Taking into account (27), one obtains:

$$\sum_{m,s}\frac{\partial^2}{\partial w_m \partial w_s}\frac{\partial\sqrt{g}L}{\partial q_{ms}} = 0, \tag{79}$$

so that (75) can be rewritten as:

$$
\begin{aligned}
i_\nu = -\frac{\partial\sqrt{g}L}{\partial w_\nu} &+ \sum_m \Big(q_\nu \frac{\partial}{\partial w_m}[\sqrt{g}L]_m + M_{m\nu}[\sqrt{g}L]_m\Big) \\
&+ \sum_m \frac{\partial\sqrt{g}L}{\partial q_m} q_{m\nu} + \sum_{s,m} \frac{\partial\sqrt{g}L}{\partial M_{sm}} \frac{\partial M_{s\nu}}{\partial w_m}.
\end{aligned}
\tag{80}
$$

While the right-hand side of this equation only involves the electrodynamic part of the Lagrangian, in view of (73) this is not the case for the left-hand side. Therefore, Hilbert once more uses the field equations, in the form of (72), for i_ν to obtain an expression entirely in terms of the electrodynamic part of the Lagrangian. For this purpose, he first writes:

$$
-\frac{\partial\sqrt{g}L}{\partial w_\nu} = -\sum_{s,m} \frac{\partial\sqrt{g}L}{\partial g^{sm}} g_\nu^{sm} - \sum_m \frac{\partial\sqrt{g}L}{\partial q_m} q_{m\nu} - \sum_{m,s} \frac{\partial\sqrt{g}L}{\partial q_{ms}} \frac{\partial q_{ms}}{\partial w_\nu},
\tag{81}
$$

and then uses (72) and (73) to identify the first term on the right-hand side as i_ν. Hilbert thus reaches his goal of transforming the identity following from Theorem III into an equation involving only the electromagnetic potentials. A further simplification results from noting that the last term on the right-hand side of (81) is, apart from its sign, identical to the last term of (80). (This is because:

$$
\sum_{s,m} \frac{\partial\sqrt{g}L}{\partial M_{sm}} \Big(\frac{\partial M_{s\nu}}{\partial w_m} - \frac{\partial q_{ms}}{\partial w_\nu}\Big) = 0,
\tag{82}
$$

which follows from the definition (28) of M_{sm}.)

Finally, using (80), Hilbert obtains:

$$
\sum_m \Big(M_{m\nu}[\sqrt{g}L]_m + q_\nu \frac{\partial}{\partial w_m}[\sqrt{g}L]_m\Big) = 0.
\tag{83}
$$

Summarizing what he had achieved, Hilbert claimed:

> ... from the gravitational equations (4) [i.e. (51)] there follow indeed the four linearly independent combinations (32) [i.e. (83)] of the basic electrodynamic equations (5) [i.e. (52)] and their first derivatives. *This is the entire mathematical expression of the general claim made above about the character of electrodynamics as an epiphenomenon of gravitation*.[84]

On closer inspection, Hilbert's claim turns out to be problematic. One might try to interpret it in either of two ways: the electromagnetic field equations follow either differentially or algebraically from (83).

---

84 "... aus den Gravitationsgleichungen (4) folgen in der Tat die vier von einander unabhängigen linearen Kombinationen (32) der elektrodynamischen Grundgleichungen (5) und ihrer ersten Ableitungen. *Dies ist der ganze mathematische Ausdruck der oben allgemein ausgesprochenen Behauptung über den Charakter der Elektrodynamik als einer Folgeerscheinung der Gravitation*." (Proofs, 12) In (Hilbert 1916, 406), "ganze" [entire] is corrected to "genaue" [exact] in the last sentence.

In the first case one would have to show that, if these equations hold on an initial hypersurface $w_4 = const$, then they hold everywhere off that hypersurface by virtue of the identities (83). Indeed it follows from these identities that, if these equations hold on $w_4 = 0$:

$$\frac{\partial[\sqrt{g}L]_4}{\partial w_4} = 0, \tag{84}$$

so that, by iteration, $[\sqrt{g}L]_4 = 0$ holds everywhere provided that it holds initially and that the other three field equations hold everywhere. But the time derivatives of the other three field equations,

$$\frac{\partial[\sqrt{g}L]_m}{\partial w_4} \qquad m = 1, 2, 3 \tag{85}$$

remain unrestricted by the identity so that one cannot simply give the electromagnetic field equations on an initial hypersurface and have them continue to hold automatically off it as a consequence of (83).

In the second case, it is clear that the field equations can only hold algebraically by virtue of (83) if the second term vanishes; this implies that the theory is gauge invariant, i.e. that the potentials themselves do not enter the field equations. In that case one indeed obtains an additional identity from gauge invariance:

$$\frac{\partial[\sqrt{g}L]_m}{\partial w_m} = 0. \tag{86}$$

(In the usual Maxwell theory this is the identity that guarantees conservation of the charge-current vector.) However, this cannot have been the argument Hilbert had in mind when stating his claim. First of all, he did not introduce the additional assumptions required—and could not have introduced them because they violated his physical assumptions;[85] and second he did not derive the identity for gauge-invariant electromagnetic Lagrangians that makes this argument work. As illustrated by Klein's later work, the derivation of these identities is closely related to a different perspective on Hilbert's results, a perspective in which electromagnetism is no longer, as in Hilbert's Proofs, treated as an epiphenomenon of gravitation, but in which both are treated in parallel.[86]

In summary, Hilbert's claim that the electromagnetic equations are a consequence of the gravitational ones turns out to be an interpretation forced upon his mathematical results by his overall program rather than being implied by them. In any case, this

---

85 Mie's original theory is in fact not gauge invariant, and in the version adopted by Hilbert one of the invariants involves a function of the electromagnetic potential vector, see (33).

86 Compare Klein's attempt to derive analogous equations for the gravitational and the electromagnetic potentials, from which the Maxwell equations then are derived, (Klein 1917, 472–473).

interpretation is different from that given to the corresponding results in general relativity and usually associated with Hilbert's work.

3.7 The Deductive Structure of the Proofs Version

Having attempted to reconstruct the line of reasoning Hilbert followed while developing the original version of his theory, we now summarize the way in which he presented these results in the Proofs. This serves as a review of the deductive structure of his theory, indicating which results were emphasized by Hilbert, and facilitating a comparison between the Proofs and the published versions.

We begin by recalling the elements of this deductive structure that Hilbert introduced explicitly:

- Axiom I "Mie's Axiom von der Weltfunktion," (see (19))
- Axiom II "Axiom von der allgemeinen Invarianz," (see the passage below (19))
- Axiom III "Axiom von Raum und Zeit," (see the passage above (55))
- Theorem I, Hilbert's *Leitmotiv*, (see (50))
- Theorem II, Lie derivative of the Lagrangian, (see (23))
- Theorem III, contracted Bianchi identities, (see (70))
- Proposition 1, divergence character of the energy expression, (see (47))
- Proposition 2, identity obeyed by the components of the energy expression, (see (59)).

He also used the following assumptions, introduced as part of his deductive structure without being explicitly stated:

- vanishing of the divergence of the energy expression (see (63))
- splitting of the Lagrangian into gravitational and electrodynamical terms (see (16))
- the assumption that the electrodynamical term does not depend on the derivatives of the metric tensor (see (25)).

There are, furthermore, the following physical results, not labelled as theorems:

- the field equations (see (51) and (52))
- the energy expression (see (39)) and the related coordinate restrictions (see (63))
- the form of Mie's Lagrangian (see (27))
- the relation between Mie's energy tensor and Lagrangian (see (36))
- the relation between the electromagnetic and gravitational field equations (see (83)).

The exposition of Hilbert's theory in the Proofs can be subdivided into four sections, to which we give short titles and list under each the relevant elements of his theory:

1. Basic Framework (Proofs, 1–3)

   Axioms I and II, Theorem I, and the field equations for gravitation and electromagnetism

2. Causality and the Energy Expression (Proofs, 3–8)

   the energy expression, Propositions 1 and 2, the divergence character of the energy expression, Axiom III, the coordinate restrictions, the split of the Lagrangian into gravitational and electrodynamical terms, and the structure of the electrodynamical term
3. Basic Theorems (Proofs, 8–9)

   Theorems II and III
4. Implications for Electromagnetism (Proofs, 9–13)

   the form of Mie's Lagrangian, its relation to his energy tensor, and the relation between electromagnetic and gravitational field equations.

The sequence in which Hilbert presented these elements suggests that he considered its implications for electromagnetism as the central results of his theory. Indeed, the gravitational field equations are never explicitly given and only briefly considered at the beginning as part of the general framework, whereas the presentation concludes with three results concerning Mie's theory. The centrality of these electromagnetic implications for him is also clear from his introductory and concluding remarks. Hilbert's initial discussion mentions Mie's electrodynamics first, and closes with the promise of further elaboration of the consequences of his theory for electrodynamics:

> The far reaching ideas and the formation of novel concepts by means of which Mie constructs his electrodynamics, and the prodigious problems raised by Einstein, as well as his ingeniously conceived methods of solution, have opened new paths for the investigation into the foundations of physics.
>
> In the following—in the sense of the axiomatic method—I would like to develop from three simple axioms a new system of basic equations of physics, of ideal beauty, containing, I believe, the solution of the problems presented. I reserve for later communications the detailed development and particularly the special application of my basic equations to the fundamental questions of the theory of electricity.[87]

In his conclusion, Hilbert makes clear what he had in mind here: a solution of the riddles of atomic physics:

> As one can see, the few simple assumptions expressed in axioms I, II, III suffice with appropriate interpretation to establish the theory: through it not only are our views of space, time, and motion fundamentally reshaped in the sense called for by Einstein, but I am also convinced that through the basic equations established here the most

87 "Die tiefgreifenden Gedanken und originellen Begriffsbildungen vermöge derer Mie seine Elektrodynamik aufbaut, und die gewaltigen Problemstellungen von Einstein sowie dessen scharfsinnige zu ihrer Lösung ersonnenen Methoden haben der Untersuchung über die Grundlagen der Physik neue Wege eröffnet.
Ich möchte im Folgenden—im Sinne der axiomatischen Methode—aus drei einfachen Axiomen ein neues System von Grundgleichungen der Physik aufstellen, die von idealer Schönheit sind, und in denen, wie ich glaube, die Lösung der gestellten Probleme enthalten ist. Die genauere Ausführung sowie vor allem die spezielle Anwendung meiner Grundgleichungen auf die fundamentalen Fragen der Elektrizitätslehre behalte ich späteren Mitteilungen vor." (Proofs, 1)

> intimate, hitherto hidden processes in the interior of atoms will receive an explanation; and in particular that generally a reduction of all physical constants to mathematical constants must be possible—whereby the possibility approaches that physics in principle becomes a science of the type of geometry: surely the highest glory of the axiomatic method, which, as we have seen, here takes into its service the powerful instruments of analysis, namely the calculus of variations and the theory of invariants.[88]

Hilbert's final remarks about the status of his theory *vis à vis* Einstein's work on gravitation strikingly parallel Minkowski's assessment of the relation of his four-dimensional formulation to Einstein's special theory; not just providing a mathematical framework for existing results, but developing a genuinely novel physical theory, which, properly understood, turns out to be a part of mathematics.[89]

Fig. 1 provides a graphical survey of the deductive structure of Hilbert's theory. The main elements listed above are connected by arrows; mathematical implications are represented by straight arrows and inferences based on heuristic reasoning by curved arrows. As the figure shows, apart from the field equations, Hilbert's results can be divided into two fairly distinct clusters: one comprises the implications for electromagnetism (right-hand side of the diagram); the other, the implications for the understanding of energy conservation (left-hand side of the diagram). While the assertions concerning energy conservation are not essential for deriving the other results, they depend on practically all the other parts of this theory. The main link between the two clusters is clearly Theorem I. Although no assertion of Hilbert's theory is derived directly from Theorem I, it motivates both the relation between energy conservation and coordinate restrictions and the link between electromagnetism and gravitation.

The analysis of the deductive structure of Hilbert's theory thus confirms that Theorem I is indeed the *Leitmotiv* of the theory. The two clusters of results obviously are also related to what he considered the two main physical touchstones of his theory: Mie's theory of electromagnetism and energy conservation. On the other hand, neither Newton's theory of gravitation nor any other parts of mechanics are mentioned by Hilbert. Einstein's imprint on Hilbert's theory was more of a mathematical or structural nature than a physical one.

---

88 "Wie man sieht, genügen bei sinngemäßer Deutung die wenigen einfachen in den Axiomen I, II, III ausgesprochenen Annahmen zum Aufbau der Theorie: durch dieselbe werden nicht nur unsere Vorstellungen über Raum, Zeit und Bewegung von Grund aus in dem von Einstein geforderten Sinne umgestaltet, sondern ich bin auch der Überzeugung, daß durch die hier aufgestellten Grundgleichungen die intimsten, bisher verborgenen Vorgänge innerhalb des Atoms Aufklärung erhalten werden und insbesondere allgemein eine Zurückführung aller physikalischen Konstanten auf mathematische Konstanten möglich sein muß—wie denn überhaupt damit die Möglichkeit naherückt, daß aus der Physik im Prinzip eine Wissenschaft von der Art der Geometrie werde: gewiß der herrlichste Ruhm der axiomatischen Methode, die hier wie wir sehen die mächtigen Instrumente der Analysis nämlich, Variationsrechnung und Invariantentheorie, in ihre Dienste nimmt." (Proofs, 13)

89 For Minkowski, see (Walter 1999).

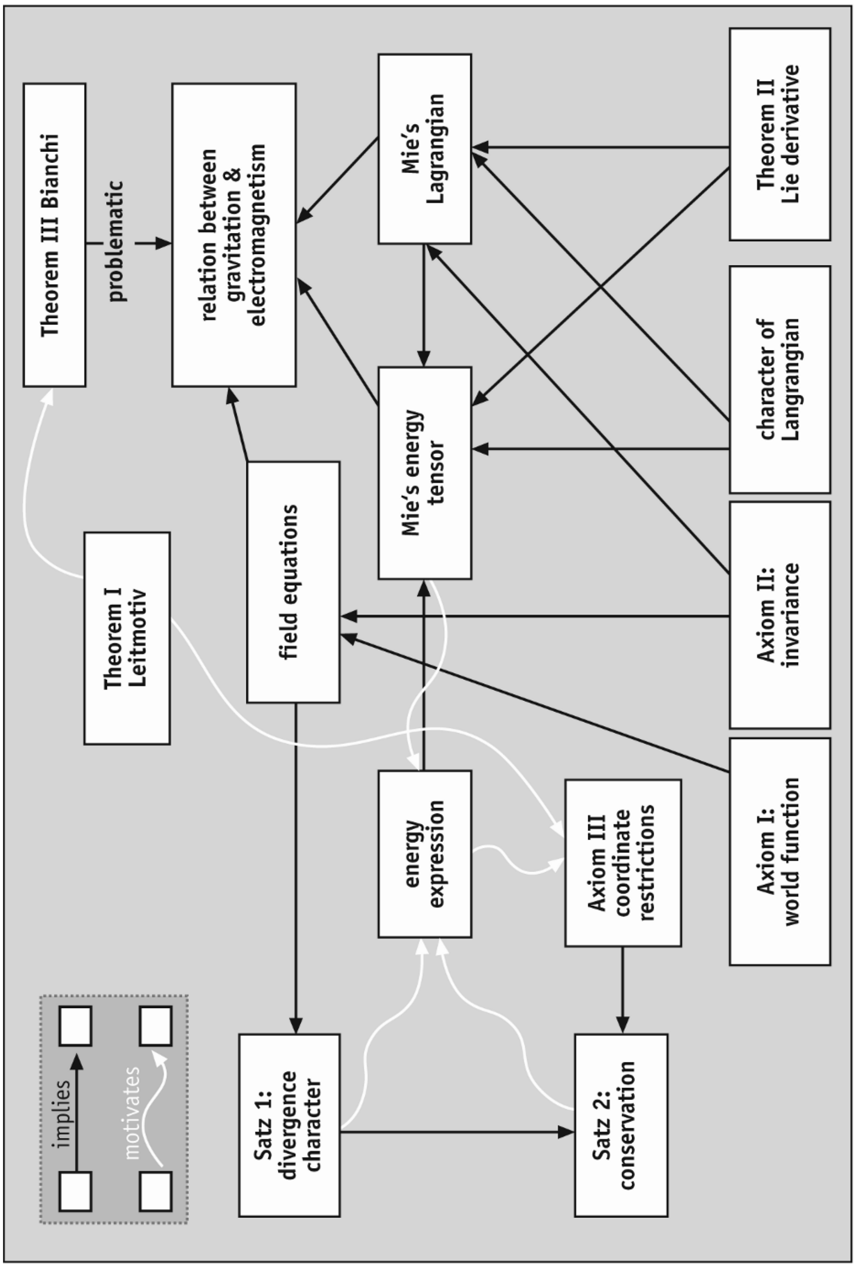


Figure 1: Deductive Structure of the Proofs (1915)

4. HILBERT'S PHYSICS AND EINSTEIN'S MATHEMATICS: THE EXCHANGE OF LATE 1915

4.1 What Einstein Could Learn From Hilbert

The Hilbert-Einstein correspondence begins with Einstein's letter of 7 November 1915.[90] That November was the month during which Einstein's theory of gravitation underwent several dramatic changes documented by four papers he presented to the Prussian Academy, culminating in the definitive version of the field equations in the paper submitted 25 November.[91] On 4 November Einstein submitted his first note, in which he abandoned the *Entwurf* field equations and replaced them with equations derived from the Riemann tensor (Einstein 1915a); he included the proofs of this paper in his letter to Hilbert. In spite of this radical modification of the field equations, the structure of Einstein's theory remained essentially unchanged from that of the non-covariant 1913 *Entwurf* theory. In both, the requirement of energy-momentum conservation is linked to a restriction to adapted coordinate systems. In Einstein's 4 November paper, this restriction implies the following equation (Einstein 1915a, 785):

$$\sum_{\alpha\beta}\frac{\partial}{\partial x_\alpha}\left(g^{\alpha\beta}\frac{\partial lg\sqrt{-g}}{\partial x_\beta}\right) = -\kappa\sum_\sigma T_\sigma^\sigma. \qquad (87)$$

Einstein pointed out one immediate consequence for the choice of an adapted coordinate system:

> Equation (21a) [i.e. (87)] shows the impossibility of so choosing the coordinate system that $\sqrt{-g}$ equals 1, because the scalar of the energy tensor cannot be set to zero.[92]

That the scalar [i.e. the trace] of the energy-momentum tensor cannot vanish is obvious if one takes Einstein's standard example (a swarm of non-interacting particles or incoherent "dust") as the source of the gravitational field: the trace of its energy-momentum tensor equals the mass density of the dust. However, the physical meaning of condition (87) was entirely obscure. It was therefore incumbent upon Einstein to find a physical interpretation of it or to modify his theory once more in order to get rid of it. He soon succeeded in doing both, and formulated his new view in an addendum to the first note, published on 11 November (Einstein 1915b).

On 12 November 1915 he reported his success to Hilbert:

> For the time being, I just thank you cordially for your kind letter. Meanwhile, the problem has made new progress. Namely, it is possible to compel *general* covariance by means of the postulate $\sqrt{-g} = 1$; Riemann's tensor then furnishes the gravitational

---

90 Einstein to David Hilbert, 7 November 1915, (CPAE 8, 191).

91 See (Einstein 1915e).

92 "Aus Gleichung (21a) [i.e. (87)] geht hervor, daß es unmöglich ist, das Koordinatensystem so zu wählen, daß $\sqrt{-g}$ gleich 1 wird; denn der Skalar des Energietensors kann nicht zu null gemacht werden." (Einstein 1915a, 785)

> equations directly. If my present modification (which does not change the equations) is legitimate, then gravitation must play a fundamental role in the structure of matter. My own curiosity is impeding my work![93]

What had happened? Einstein had noticed that the condition $\sum_{\sigma} T_{\sigma}^{\sigma} = 0$, which follows from setting $\sqrt{-g} = 1$ in (87), can be related to an electromagnetic theory of matter: in Maxwell's theory, the vanishing of its trace is a characteristic property of the electromagnetic energy-momentum tensor. Thus, if one assumes all matter to be of electromagnetic origin, the vanishing of its trace becomes a fundamental property of the energy-momentum tensor. This has two important consequences: Condition (87) is no longer an inexplicable restriction on the admissible coordinate systems, and the 4 November field equations can be seen as a particular form of generally-covariant field equations based on the Ricci tensor. From the perspective of the 11 November revision, the condition $\sqrt{-g} = 1$ turns out to be nothing more than an arbitrary but convenient choice of coordinate systems.

The core of Einstein's new theory is strikingly simple. The left-hand side of the gravitational field equations is now simply the Ricci tensor and the right-hand side an energy-momentum tensor, the trace of which has to vanish:[94]

$$R_{\mu\nu} = -\kappa T_{\mu\nu} \qquad \sum_{\sigma} T_{\sigma}^{\sigma} = 0. \tag{88}$$

What distinguishes these field equations from the final equations presented on 25 November is an additional term on the right-hand side of the equations involving the trace of the energy-momentum tensor, which now need not vanish:[95]

$$R_{\mu\nu} = -\kappa\left(T_{\mu\nu} - \frac{1}{2} g_{\mu\nu} T\right). \tag{89}$$

Remarkably enough, in the winter of 1912/13 Einstein had considered the linearized form of these field equations, but discarded them because they were not compatible with his expectation of how the Newtonian limit should result.[96] He had also then considered and rejected field equations of the form (88), just because they imply

---

93 "Ich danke einstweilen herzlich für Ihren freundlichen Brief. [Das] Problem hat unterdessen einen neuen Fortschritt gemacht. Es lässt sich nämlich durch das Postulat $\sqrt{-g} = 1$ die *allgemeine* Kovarianz erzwingen; der Riemann'sche Tensor liefert dann direkt die Gravitationsgleichungen. Wenn meine jetzige Modifikation (die die Gleichungen nicht ändert) berechtigt ist, dann muss die Gravitation im Aufbau der Materie eine fundamentale Rolle spielen. Die Neugier erschwert mir die Arbeit!" Einstein to David Hilbert, 12 November 1915, (CPAE 8, 194).

94 See (Einstein 1915b, 801 and 800).

95 See (Einstein 1915e, 845).

96 See Doc. 10 of (CPAE 4), "Pathways out of Classical Physics ...", "Einstein's Zurich Notebook", (both in vol. 1 of this series), and the "Commentary" (in vol. 2).

the condition $\sum_{\sigma} T_{\sigma}^{\sigma} = 0$. At that time, this condition seemed unacceptable because the trace of the energy-momentum tensor of ordinary matter does *not* vanish.

The prehistory of Einstein's 11 November paper thus confronts us with a puzzle: Why did he consider it to be such a decisive advance beyond his 4 November paper and not just a possible alternative interpretation of his previous results; and why did he now so readily accept the trace-condition $\sum_{\sigma} T_{\sigma}^{\sigma} = 0$ that earlier had led him to reject this very theory? What impelled Einstein's change of perspective in November 1915?

The answer seems to lie in the changed context, within which Einstein formulated his new approach: in particular, his interaction with Hilbert. As will become evident, it would have been quite uncharacteristic of him to adopt the new approach so readily had it not been for current discussions of the electrodynamic worldview and his feeling that he was now in competition with Hilbert.[97]

In his addendum, Einstein directly referred to the supporters of the electrodynamic worldview:

> One now has to remember that, in accord with our knowledge, "matter" is not to be conceived as something primitively given, or physically simple. There even are those, and not just a few, who hope to be able to reduce matter to purely electrodynamic processes, which of course would have to be done in a theory more complete than Maxwell's electrodynamics.[98]

Only this context explains Einstein's highly speculative and fragmentary comments on an electromagnetic model of matter. That, in November 1915, Einstein conceived of a field theory of matter as a goal in its own right is also supported by his correspondence, which makes it clear that this perspective was shaped by his rivalry with Hilbert. We have already cited Einstein's letter to Hilbert, in which he wrote:

> If my present modification (which does not change the equations) is legitimate, then gravitation must play a fundamental role in the structure of matter. My own curiosity is impeding my work![99]

And when, in a letter of 14 November, Hilbert claimed to have achieved the unification of gravitation and electromagnetism, Einstein responded:

---

97 For a discussion of Hilbert's reaction to what he must have seen as an intrusion by Einstein into his domain, see (Sauer 1999, 542–543).

98 "Es ist nun daran zu erinnern, daß nach unseren Kenntnissen die "Materie" nicht als ein primitiv Gegebenes, physikalisch Einfaches aufzufassen ist. Es gibt sogar nicht wenige, die hoffen, die Materie auf rein elektromagnetische Vorgänge reduzieren zu können, die allerdings einer gegenüber Maxwells Elektrodynamik vervollständigten Theorie gemäß vor sich gehen würden." (Einstein 1915b, 799)

99 "Wenn meine jetzige Modifikation (die die Gleichungen nicht ändert) berechtigt ist, dann muss die Gravitation im Aufbau der Materie eine fundamentale Rolle spielen. Die Neugier erschwert mir die Arbeit!" Einstein to David Hilbert, 12 November 1915, (CPAE 8, 194).

> Your investigation interests me tremendously, especially since I often racked my brain to construct a bridge between gravitation and electromagnetics.[100]

A few days later (after calculating the perihelion shift on the basis of the new theory), he expressed himself similarly:

> In these last months I had great success in my work. *Generally covariant* gravitation equations. *Perihelion motions explained quantitatively.* The role of gravitation in the structure of matter. You will be astonished. I worked dreadfully hard; it is remarkable that one can sustain it.[101]

When one examines Einstein's previous writings on gravitation, published and unpublished, one finds no trace of an attempt to unify gravitation and electromagnetism. He had never advocated the electromagnetic worldview. On the contrary, he was apparently disinterested in Mie's attempt at a unification of gravitation and electrodynamics, not finding it worth mentioning in his 1913 review of contemporary gravitation theories.[102]

And soon after completion of the final version of general relativity, Einstein reverted to his earlier view that general relativity could make no assertions about the structure of matter:

> From what I know of Hilbert's theory, it makes use of an assumption about electrodynamic processes that—apart from the treatment of the gravitational field—is closely connected to Mie's. Such a specialized approach is not in accordance with the point of view of general relativity. The latter actually only provides the gravitational field law, and quite unambiguously so when general covariance is required.[103]

Einstein's mid-November 1915 pursuit of a relation between gravitation and electromagnetism was, then, merely a short-lived episode in his search for a relativistic theory of gravitation. Its novelty is confirmed by a footnote in the addendum:

> In writing the earlier paper, I had not yet realized that the hypothesis $\sum T_\mu^\mu = 0$ is, in principle, admissible.[104]

---

100 "Ihre Untersuchung interessiert mich gewaltig, zumal ich mir schon oft das Gehirn zermartert habe, um eine Brücke zwischen Gravitation und Elektromagnetik zu schlagen." Einstein to David Hilbert, 15 November 1915, (CPAE 8, 199).

101 "Ich habe mit grossem Erfolg gearbeitet in diesen Monaten. *Allgemein kovariante* Gravitationsgleichungen. *Perihelbewegungen quantitativ erklärt*. Rolle der Gravitation im Bau der Materie. Du wirst staunen. Gearbeitet habe ich schauderhaft angestrengt; sonderbar, dass man es aushält." Einstein to Michele Besso, 17 November 1915, (CPAE 8, 201).

102 See (Einstein 1913).

103 "Soviel ich von Hilbert's Theorie weiss, bedient sie sich eines Ansatzes für das elektrodynamische Geschehen, der sich [— a]bgesehen von der Behandlung des Gravitationsfeldes — eng an Mie anschliesst. Ein derartiger spezieller Ansatz lässt sich aus dem Gesichtspunkte der allgemeinen Relativität nicht begründen. Letzterer liefert eigentlich nur das Gesetz des Gravitationsfeldes, und zwar ganz eindeutig, wenn man allgemeine Kovarianz fordert." Einstein to Arnold Sommerfeld, 9 December 1915, (CPAE 8, 216).

104 "Bei Niederschrift der früheren Mitteilung war mir die prinzipielle Zulässigkeit der Hypothese $\sum T_\mu^\mu = 0$ noch nicht zu Bewußtsein gekommen." (Einstein 1915b, 800)

It thus seems quite clear that Einstein's temporary adherence to an electromagnetic theory of matter was triggered by Hilbert's work, which he attempted to use in order to solve a problem that had arisen in his own theory, and that he dropped it when he solved this problem in a different way.

So this whole episode might appear to be a bizarre and unnecessary detour. A closer analysis of the last steps of Einstein's path to general relativity shows, however, that the solution depended crucially on this detour, and hence indirectly on Hilbert's work. In fact, Einstein successfully calculated the perihelion shift of Mercury on the basis of his 11 November theory.[105] The condition $\sqrt{-g} = 1$, implied by the assumption of an electromagnetic origin of matter (see (87)), was essential for this calculation, which Einstein considered a striking confirmation of his audacious hypothesis on the constitution of matter, definitely favoring this theory over that of 4 November.[106] The 11 November theory also turned out to be the basis for a new understanding of the Newtonian limit, which allowed Einstein to accept the field equations of general relativity as the definitive solution to the problem of gravitation. Ironically, Hilbert's most important contribution to general relativity may have been enhancing the credibility of a speculative and ultimately untenable physical hypothesis that guided Einstein's final mathematical steps towards the completion of his theory.

Einstein submitted his perihelion paper on 18 November 1915. In a footnote, appended after its completion, Einstein observed that, in fact, the hypothesis of an electromagnetic origin of matter is unnecessary for the perihelion shift calculation. He announced a further modification of his field equations, finally reaching the definitive version of his theory.[107] On the same day, Einstein wrote to Hilbert, acknowledging receipt of Hilbert's work, including a system of field equations:

> The system [of field equations] you give agrees—as far as I can see—exactly with that which I found in the last few weeks and have presented to the Academy.[108]

105 See (Einstein 1915c).

106 See (Einstein 1915d): the abstract of this paper, probably by Einstein, summarizes the issue: "Es wird gezeigt, daß die allgemeine Relativitätstheorie die von Leverrier entdeckte Perihelbewegung des Merkurs qualitativ und quantitativ erklärt. Dadurch wird die Hypothese vom Verschwinden des Skalars des Energietensors der "Materie" bestätigt. Ferner wird gezeigt, daß die Untersuchung der Lichtstrahlenkrümmung durch das Gravitationsfeld ebenfalls eine Möglichkeit der Prüfung dieser wichtigen Hypothese bietet." ("It will be shown that the theory of general relativity explains qualitatively and quantitatively the perihelion motion of Mercury, which was discovered by Leverrier. Thus the hypothesis of the vanishing of the scalar of the energy tensor of "matter" is confirmed. Furthermore, it is shown that the analysis of the bending of light by the gravitational field also offers a way of testing this important hypothesis.")

107 See (Einstein 1915c, 831).

108 "Das von Ihnen gegebene System [of field equations] stimmt - soweit ich sehe - genau mit dem überein, was ich in den letzten Wochen gefunden und der Akademie überreicht habe." Einstein to David Hilbert, 18 November 1915, (CPAE 8, 201–202). For discussion of what Einstein may have received from Hilbert, see below.

Einstein emphasized that the real difficulty had not been the formulation of generally-covariant field equations, but in showing their agreement with a physical requirement: the existence of the Newtonian limit. Stressing his priority, he mentioned that he had considered such equations three years earlier:

> … it was hard to recognize that these equations form a generalisation, and indeed a simple and natural generalisation, of Newton's law. It has just been in the last few weeks that I succeeded in this (I sent you my first communication), whereas 3 years ago with my friend Grossmann I had already taken into consideration the only possible generally covariant equations, which have now been shown to be the correct ones. We had only heavy-heartedly distanced ourselves from it, because it seemed to me that the physical discussion had shown their incompatibility with Newton's law.[109]

Einstein's statement not only characterized his own approach, but indirectly clarified his ambivalent position with regard to Hilbert's theory. While evidently fascinated by the perspective of unifying gravitation and electromagnetism, he now recognized that, at least in Hilbert's case, this involved the risk of neglecting the sound foundation of the new theory of gravitation in the classical theory.

4.2 What Hilbert Could Learn from Einstein

Hilbert must have seen Einstein's letter of 12 November, announcing publication of new insights into a fundamental role of gravitation in the constitution of matter, as a threat to his priority.[110] At any rate, Hilbert hastened public presentation of his results. His response of 13 November gave a brief sketch of his theory and announced a 16th November seminar on it:

> Actually, I wanted first to think of a quite palpable application for physicists, namely valid relations between physical constants, before obliging with my axiomatic solution to your great problem. But since you are so interested, I would like to develop my th[eory] in very complete detail on the coming Tuesday, that is, the day after the day after tomorrow (the 16th of this mo.). I find it ideally beautiful math[ematically], and also insofar as calculations that are not completely transparent do not occur at all, and absolutely compelling in accordance with the axiom[atic] meth[od] and therefore rely on its reality. As a result of a gen. math. theorem, the (generalized Maxwellian) electrody. eqs. appear as a math. consequence of the gravitation eqs., so that gravitation and electrodynamics are actually not at all different. Furthermore, my energy concept forms the basis: $E = \Sigma(e_s t^s + e_{ih} t^{ih})$, [the t^s corresponds to p^s in Hilbert's papers, etc.] which is likewise a general invariant [see (56)], and from this then also follow from a very simple

109 "schwer war es, zu erkennen, dass diese Gleichungen eine Verallgemeinerung, und zwar eine einfache und natürliche Verallgemeinerung des Newton'schen Gesetzes bilden. Dies gelang mir erst in den letzten Wochen (meine erste Mitteilung habe ich Ihnen geschickt), während ich die einzig möglichen allgemein kovarianten Gleichungen, [die] sich jetzt als die richtigen erweisen, schon vor 3 Jahren mit meinem Freunde Grossmann in Erwägung gezogen hatte. Nur schweren Herzens trennten wir uns davon, weil mir die physikalische Diskussion scheinbar ihre Unvereinbarkeit mit Newtons Gesetz ergeben hatte."

110 This aspect of the Hilbert-Einstein relationship was first discussed in (Sauer 1999), where the chronology of events is carefully reconstructed.

> axiom the 4 still-missing "spacetime equations" $e_s = 0$. I derived most pleasure in the discovery already discussed with Sommerfeld that the usual electrical energy results when a certain absolute invariant is differentiated with respect to the gravitation potentials and then g are set = 0,1.[111]

This letter presents the essential elements of Hilbert's theory as presented in the Proofs. His reference to "the missing spacetime equations" suggests that he saw these equations and their relation to the energy concept as an issue common to his theory and Einstein's.

Einstein responded on 15 November 1915, declining the invitation to come to Göttingen on grounds of health.[112] Instead, he asked Hilbert for the proofs of his paper. As mentioned above, by 18 November Hilbert had fulfilled Einstein's request. He could not have sent the typeset Proofs, which are dated 6 December, so he must have sent a manuscript on 20 November, presumably corresponding to his talk. Since the Proofs are also dated 20 November, this manuscript may well have presented practically the same version of his theory. On 19 November, a day after Einstein announced his successful perihelion calculation to Hilbert, the latter sent his congratulations, making clear once more that the physical problems facing Hilbert's theory were of a rather different nature:

> Many thanks for your postcard and cordial congratulations on conquering perihelion motion. If I could calculate as rapidly as you, in my equations the electron would correspondingly have to capitulate, and simultaneously the hydrogen atom would have to produce its note of apology about why it does not radiate.
> I would be grateful if you were to continue to keep me up-to-date on your latest advances.[113]

111 "Ich wollte eigentlich erst nur für die Physiker eine ganz handgreifliche Anwendung nämlich treue Beziehungen zwischen den physikalischen Konstanten überlegen, ehe ich meine axiomatische Lösung ihres grossen Problems zum Besten gebe. Da Sie aber so interessiert sind, so möchte ich am kommenden Dienstag also über-über morgen (d. 16 d. M.) meine Th. ganz ausführlich entwickeln. Ich halte sie für math. ideal schön auch insofern, als Rechnungen, die nicht ganz durchsichtig sind, garnicht vorkommen. und absolut zwingend nach axiom. Meth., und baue deshalb auf ihre Wirklichkeit. In Folge eines allgem. math. Satzes erscheinen die elektrody. Gl. (verallgemeinerte Maxwellsche) als math. Folge der Gravitationsgl., so dass Gravitation u. Elektrodynamik eigentlich garnichts verschiedenes sind. Desweiteren bildet mein Energiebegriff die Grundlage: $E = \Sigma(e_s t^s + e_{ih} t^{ih})$, die ebenfalls eine allgemeine Invariante ist, und daraus folgen dann aus einem sehr einfachen Axiom die noch fehlenden 4 "Raum-Zeitgleichungen" $e_s = 0$. Hauptvergnügen war für mich die schon mit Sommerfeld besprochene Entdeckung, dass die gewöhnliche elektrische Energie herauskommt, wenn man eine gewisse absolute Invariante mit den Gravitationspotentialen differenziert und dann $g = 0, 1$ setzt." David Hilbert to Einstein, 13 November 1915, (CPAE 8, 195).

112 Einstein to David Hilbert, 15 November 1915, (CPAE 8, 199).

113 "Vielen Dank für Ihre Karte und herzlichste Gratulation zu der Ueberwältigung der Perihelbewegung. Wenn ich so rasch rechnen könnte, wie Sie, müsste bei meinen Gleichg entsprechend das Elektron kapituliren und zugleich das Wasserstoffatom sein Entschuldigungszettel aufzeigen, warum es nicht strahlt. Ich werde Ihnen auch ferner dankbar sein, wenn Sie mich über Ihre neuesten Fortschritte auf dem Laufenden halten." David Hilbert to Einstein, 19 November 1915, (CPAE 8, 202).

No doubt Einstein fulfilled this request to keep Hilbert up to date. His definitive paper on the field equations, submitted 25 November and published 2 December, must have been on Hilbert's desk within a day or two. In contrast to all earlier versions of his theory, Einstein now showed that energy-momentum conservation does not imply additional coordinate restrictions on the field equations (89). He also made clear that these field equations fulfill the requirement of having a Newtonian limit and allow derivation of the perihelion shift of Mercury.

Our analysis of the Proofs suggests that neither the astronomical implications of Einstein's theory nor the latter's treatment of the Newtonian limit directly affected Hilbert's theory since they lay outside its scope, as Hilbert then perceived it. But Einstein's insight that energy-momentum conservation does not lead to a restriction on admissible coordinate systems was of crucial significance for Hilbert. As we have seen, in Hilbert's theory the entire complex of results on energy-momentum conservation was structured by a logic paralleling that of Einstein's earlier non-covariant theory. Moreover, Theorem I, Hilbert's *Leitmotiv*, was motivated by Einstein's hole argument that generally-covariant field equations cannot have unique solutions. His definitive paper of 25 November did not explicitly mention the hole argument, but simply took it for granted that his new generally-covariant field equations avoid such difficulties.[114] Hilbert may well have checked that Einstein's definitive field equations were actually compatible[115] with the equations that follow from Hilbert's variational principle, which he had not explicitly calculated—or at least not included in the Proofs, and this compatibility would certainly have been reassuring for Hilbert. But the fact that the hole argument evidently no longer troubled Einstein must have led Hilbert to question his *Leitmotiv*, with its double role of motivating coordinate restrictions and providing the link between gravitation and electromagnetism.

Thus, Einstein's paper of 25 November 1915 represented a major challenge for Hilbert's theory. As we shall see when discussing the published version of Hilbert's paper, while Einstein temporarily took over Hilbert's physical perspective, Hilbert appears to have accepted the mathematical implications of Einstein's rejection of the hole argument.

4.3 Cooperation in the Form of Competition

In a situation such as we have described, in which the interaction between two people working on closely related problems changes the way in which each of them proceeds, it is not easy for the individuals to assess their own contributions. While Einstein was happy to have found in Hilbert one of the few colleagues, if not the only one, who appreciated and understood the nature of his work on gravitation, he also

---

114 The fact that these equations were supported by Einstein's successful calculation of the perihelion shift made it impossible for Hilbert simply to disregard them.

115 Compatible, but not the same, because of the trace term, and because of the different treatment of the stress-energy tensor, as discussed elsewhere in this paper.

resented the way in which Hilbert took over some of his results without, as Einstein saw it, giving him due credit. Einstein wrote to his friend Heinrich Zangger on 26 November 1915 with regard to his newly-completed theory:

> The theory is beautiful beyond comparison. However, only *one* colleague has really understood it, and he is seeking to "partake" [*nostrifizieren*] in it (Abraham's expression) in a clever way. In my personal experience I have hardly come to know the wretchedness of mankind better than as a result of this theory and everything connected to it. But it does not bother me.[116]

Einstein's reaction becomes particularly understandable in the light of his prior positive experience of collaboration with his friend, the mathematician Marcel Grossmann. Grossmann had restricted himself to putting his superior mathematical competence at Einstein's service.[117] What Hilbert offered was not cooperation but competition. Hilbert may well have been upset by Einstein's anticipation in print, in his paper of 11 November, of what Hilbert felt to be his idea of a close link between gravitation and the structure of matter. Even more disturbing may have been the fact that, contrary to Hilbert's assertion in the Proofs, Einstein's final formulation of his theory required no restriction on general covariance. But it is not clear exactly when Hilbert abandoned all non-covariant elements of his program, in particular his approach to the energy problem and consequent restriction to a preferred class of coordinate systems.[118]

Hilbert evidently learned of Einstein's resentment over lack of recognition by Hilbert, possibly as a result of Einstein's letter of 18 November pointing out his priority in setting up generally-covariant field equations. In any case, he began to introduce changes in his Proofs on or after 6 December, documented by handwritten marginalia, changes which not only acknowledge Einstein's priority but attempt to placate him. Hilbert's revision also provides an indication of the content of Einstein's complaints. He revised the programmatic statement in the introduction of his paper (his insertion is rendered in italics):

> In the following — in the sense of the axiomatic method — I would like to develop, *essentially* from three simple axioms a ~~new~~ system of basic equations of physics, of ideal beauty, containing, I believe, the solution of the problems presented.[119]

---

116 "Die Theorie ist von unvergleichlicher Schönheit. Aber nur *ein* Kollege hat sie wirklich verstanden und der eine sucht sie auf geschickte Weise zu "*nostrifizieren*" (Abraham'scher Ausdruck). Ich habe in meinen persönlichen Erfahrungen kaum je die Jämmerlichkeit der Menschen besser kennen gelernt wie gelegentlich dieser Theorie und was damit zusammenhängt. Es ficht mich aber nicht an." Einstein to Heinrich Zangger, 26 November 1915, (CPAE 8, 205). See the discussion of "nostrification" above.

117 See the editorial note "Einstein on Gravitation and Relativity: The Collaboration with Marcel Grossmann" in (CPAE 4, 294–301).

118 According to (Sauer 1999, 562), Hilbert had found the new energy expression by 25 January 1916.

119 "Ich möchte im Folgenden - im Sinne der axiomatischen Methode - *wesentlich* aus drei einfachen Axiomen ein ~~neues~~ System von Grundgleichungen der Physik aufstellen, die von idealer Schönheit sind, und in denen, wie ich glaube, die Lösung der gestellten Probleme enthalten ist." (Proofs, 1)

The insertion "wesentlich" was presumably motivated by Hilbert's recognition that his theory actually presupposed additional assumptions of substantial content, such as the assumption of a split of the Lagrangian into gravitational and electromagnetic parts and the assumption that the latter does not depend on derivatives of the metric (see section 3). A further assumption was the requirement that the gravitational part of the Lagrangian not involve derivatives of the metric higher than second order. Einstein had justified this requirement by the necessity for the theory to have a Newtonian limit, and it may have been Einstein's argument that drew Hilbert's attention to the fact that his theory was actually based on a much wider array of assumptions than his axiomatic presentation had indicated. More remarkably, in characterizing his system of equations, Hilbert deleted the word "neu," a clear indication that he had read Einstein's 25 November paper and recognized that the equations implied by his own variational principle are formally equivalent (because of where the trace term occurs) to Einstein's if Hilbert's electrodynamic stress-energy tensor is substituted for the unspecified one on the right-hand side of Einstein's field equations.

Hilbert's next change was presumably related to a complaint by Einstein about the lack of proper acknowledgement for what he considered to be one of his fundamental contributions, the introduction of the metric tensor as the mathematical representation of the gravitational potentials. Hilbert had indeed given the impression that Einstein's merit was confined to asking the right questions, while Hilbert provided the answers.

Hilbert's revised description of these gravitational potentials reads (his insertion is again rendered in italics):

> The quantities characterizing the events at w_s shall be:
>
> 1) The ten gravitational potentials *first introduced by Einstein*, $g_{\mu\nu}$ $(\mu, \nu = 1, 2, 3, 4)$ having the character of a symmetric tensor with respect to arbitrary transformation of the world parameter w_s;
>
> 2) The four electrodynamic potentials q_s having the character of a vector in the same sense.[120]

The next change represents an even more far-going recognition that Hilbert could not simply claim the results in his paper as parts of "his theory," as if it had nothing substantial in common with that of Einstein:

> The guiding motive for setting up ~~my~~ *the* theory is given by the following theorem, the proof of which I will present elsewhere.[121]

120 "Die das Geschehen in w_s charakterisierenden Größen seien:
1) die zehn *von Einstein zuerst eingeführten* Gravitationspotentiale $g_{\mu\nu}$ $(\mu, \nu = 1, 2, 3, 4)$ mit symmetrischem Tensorcharakter gegenüber einer beliebigen Transformation der Weltparameter w_s;
2) die vier elektrodynamischen Potentiale q_s mit Vektorcharakter im selben Sinne." (Proofs, 1)

121 "Das Leitmotiv für den Aufbau ~~meiner~~ *der* Theorie liefert der folgende mathematische Satz, dessen Beweis ich an einer anderen Stelle darlegen werde." (Proofs, 2)

Hilbert's final marginal notation consists of just an exclamation mark next to a minor correction of the energy expression (39)—perhaps evidence that he had identified this expression as the central problem in the Proofs. While Hilbert's first annotations were presumably intended as revisions of a text that was going to remain basically unchanged, this exclamation mark signals the abandonment of such an attempt at revision. At this point, perhaps it dawned upon Hilbert that Einstein's results forced him to rethink his entire approach.

Hilbert's recognition of the problematic character of his treatment of energy-momentum conservation appears to have been solely in reaction to Einstein's results and not as a consequence of any internal dynamics (see section 3) of the development of his theory.[122] Indeed, as our analysis of the deductive structure of Hilbert's theory showed, this treatment is well anchored in the remainder of his theory without in turn having much effect on the remainder. Hence, there was no "internal friction" that could have driven a further development of Hilbert's theory. On the contrary, since the link between energy-momentum conservation and coordinate restrictions was motivated by Hilbert's Theorem I, Einstein's abandonment of this link left Hilbert at a loss, as we have argued above. But the way in which energy-momentum conservation was connected to other results of his theory also suggested how to modify it in the direction indicated by Einstein: Hilbert had to find a new energy expression that does not imply a coordinate restriction but is still connected with Mie's energy-momentum tensor. Precisely the decoupling of his energy expression from the physical consequences of Hilbert's theory made such a modification possible. Hilbert gave up immediate publication and began to rework his theory. By early 1916 had he arrived at results that made possible this rewriting of his paper and its submission for publication; by mid-February 1916, Paper 1, which we will discuss in the following section, was in press.[123]

Meanwhile, having emerged triumphant from the exchange of November 1915, Einstein offered a reconciliation to Hilbert:

> There has been a certain ill-feeling between us, the cause of which I do not want to analyze. I have struggled against the feeling of bitterness attached to it, and this with complete success. I think of you again with unmarred friendliness and ask you to try to do the same with me. Objectively it is a shame when two real fellows who have extricated themselves somewhat from this shabby world do not afford each other mutual pleasure.[124]

---

122 For a different view, see (Sauer 1999, 570).

123 For a detailed chronology, see the reconstruction in (Sauer 1999, 560–565).

124 "Es ist zwischen uns eine gewisse Verstimmung gewesen, deren Ursache ich nicht analysieren will. Gegen das damit verbundene Gefühl der Bitterkeit habe ich gekämpft, und zwar mit vollständigem Erfolge. Ich gedenke Ihrer wieder in ungetrübter Freundlichkeit, und bitte Sie, dasselbe bei mir zu versuchen. Es ist objektiv schade, wenn sich zwei wirkliche Kerle, die sich aus dieser schäbigen Welt etwas herausgearbeitet haben, nicht gegenseitig zur Freude gereichen." Einstein to David Hilbert, 20 December 1915, (CPAE 8, 222). The "schäbige [.] Welt" probably refers to World War I—given Einstein and Hilbert's critical attitude to the war.

5. HILBERT'S ASSIMILATION OF EINSTEIN'S RESULTS: THE THREE PUBLISHED VERSIONS OF HIS FIRST PAPER

5.1 The New Energy Concept—An Intermediary Solution

As we have seen, modification of Hilbert's treatment of energy-momentum conservation was the most urgent step necessitated by Einstein's results of 25 November 1915. First of all, the energy-momentum conservation law should not involve coordinate restrictions but be an invariant equation. Second, the modified energy expression should still involve Mie's energy-momentum tensor; otherwise the link between gravitation and electromagnetism, fundamental to Hilbert's program, would be endangered. Third, to accord with Hilbert's understanding of energy-momentum conservation, the new energy concept must still satisfy a divergence equation. As we shall show, Hilbert's modification of his energy expression was guided by these criteria, but its relation to a physical interpretation remained as tenuous as ever.[125] The next section concerns the effect of the new energy concept on the deductive structure of Hilbert's theory.

In the introductory discussion of energy, Paper 1 emphasizes that only axioms I and II are required:

> The most important aim is now the formulation of the concept of energy, and the derivation of the energy theorem solely on the basis of the two axioms I and II.[126]

This emphasis is in contrast with the treatment in the Proofs, in which the energy concept is closely related to axiom III, which was dropped in Paper 1. Hilbert then proceeds exactly as in the Proofs, introducing a polarization of the Lagrangian with respect to the gravitational variables (see the definition of P_g, (20)):

$$P_g(\sqrt{g}H) = \sum_{\mu,\nu,k,l}\left(\frac{\partial\sqrt{g}H}{\partial g^{\mu\nu}}p^{\mu\nu} + \frac{\partial\sqrt{g}H}{\partial g_k^{\mu\nu}}p_k^{\mu\nu} + \frac{\partial\sqrt{g}H}{\partial g_{kl}^{\mu\nu}}p_{kl}^{\mu\nu}\right). \tag{90}$$

In contrast to (37), however, Hilbert polarizes $\sqrt{g}H$ instead of H. Clearly, his aim was to formulate an equation analogous to (45), but with only a divergence term on the right-hand side. Indeed, since:

$$P(\sqrt{g}H) = \sqrt{g}PH + H\sum_{\mu,\nu}\frac{\partial\sqrt{g}}{\partial g^{\mu\nu}}p^{\mu\nu}, \tag{91}$$

use of $\sqrt{g}H$ eliminates the first term of the right-hand side of (45), giving:

---

125 For a discussion of Hilbert's concept of energy, see also (Sauer 1999, 548–550), which stresses the mathematical roots of this concept.

126 "Das wichtigste Ziel ist nunmehr die Aufstellung des Begriffes der Energie und die Herleitung des Energiesatzes allein auf Grund der beiden Axiome I und II." (Hilbert 1916, 400)

$$P_g(\sqrt{g}H) - \sum_l \frac{\partial\sqrt{g}(a^l + b^l)}{\partial w_l} = \sum_{\mu,\nu} [\sqrt{g}H]_{\mu\nu} p^{\mu\nu}. \tag{92}$$

Since the right-hand side vanishes due to the field equations, this equation is of just the desired form.

The way in which Hilbert obtained (92) closely parallels that used in the Proofs, i.e. by splitting off divergence terms. He starts out by noting that:

$$a^l = \sum_{\mu,\nu,k} \frac{\partial H}{\partial g^{\mu\nu}_{kl}} A^{\mu\nu}_k, \tag{93}$$

where $A^{\mu\nu}_k$ is the covariant derivative of $p^{\mu\nu}$, is a contravariant vector.

Then he observes that:

$$P_g(\sqrt{g}H) - \sum_l \frac{\partial\sqrt{g}a^l}{\partial w_l} \tag{94}$$

no longer contains the second derivatives of $p^{\mu\nu}$, and hence can be written:

$$\sqrt{g} \sum_{\mu,\nu,k} (B_{\mu\nu} p^{\mu\nu} + B^k_{\mu\nu} p^{\mu\nu}_k), \tag{95}$$

where $B^k_{\mu\nu}$ is a tensor. Finally, Hilbert forms the vector:

$$b^l = \sum_{\mu,\nu} B^l_{\mu\nu} p^{\mu\nu}, \tag{96}$$

obtaining (92).

He next forms the expression for the electromagnetic variables analogous to (92) (see the definition of P_q, (20) above):

$$P_q(\sqrt{g}H) - \sum_l \frac{\partial\sqrt{g}c^l}{dw_l} = \sum_k [\sqrt{g}H]_k p_k, \tag{97}$$

with:

$$c^l = \sum_k \frac{\partial H}{\partial q_{kl}} p_k. \tag{98}$$

Adding (92) and (97), and taking account of the field equations, Hilbert could thus write:

$$P(\sqrt{g}H) = \sum_l \frac{\partial\sqrt{g}(a^l + b^l + c^l)}{\partial w_l}. \tag{99}$$

The final step consists in also rewriting the left-hand side of this equation as a divergence, using (91), which is expanded as:

$$P(\sqrt{g}H) \;=\; \sqrt{g}PH + H\sum_s\Big(\frac{\partial\sqrt{g}}{\partial w_s}p^s + \sqrt{g}p_s^s\Big); \tag{100}$$

using Theorem II (see (22)),[127] he then obtained:

$$P(\sqrt{g}H) \;=\; \sqrt{g}\sum_z\frac{\partial H}{\partial w_s}p^s + H\sum_s\Big(\frac{\partial\sqrt{g}}{\partial w_s}p^s + \sqrt{g}p_s^s\Big) \;=\; \sum_s\frac{\partial\sqrt{g}Hp^s}{\partial w_s}, \tag{101}$$

and, in view of (99),

$$\sum_l\frac{\partial}{\partial w_l}\sqrt{g}(Hp^l - a^l - b^l - c^l) \;=\; 0. \tag{102}$$

This equation could have been interpreted as giving the energy expression since, being an invariant divergence, it satisfies two of the three criteria mentioned above. But it is not related to Mie's energy-momentum tensor. So Hilbert adds yet another term $-d^l$ to the expression in the parenthesis in (102):

$$d^l \;=\; \frac{1}{2\sqrt{g}}\sum_{k,s}\frac{\partial}{\partial w_k}\Big\{\Big(\frac{\partial\sqrt{g}H}{\partial q_{lk}} - \frac{\partial\sqrt{g}H}{\partial q_{kl}}\Big)p^s q_s\Big\}, \tag{103}$$

which does not alter its character since d^l is a contravariant vector (because:

$$\frac{\partial H}{\partial q_{lk}} - \frac{\partial H}{\partial q_{kl}} \tag{104}$$

is an antisymmetric tensor) that satisfies the identity:

$$\sum_l\frac{\partial\sqrt{g}d^l}{\partial w_l} \;=\; 0. \tag{105}$$

Hilbert concluded:

> Let us now define
>
> $$e^l \;=\; Hp^l - a^l - b^l - c^l - d^l \tag{106}$$
>
> as the *energy vector, then the energy vector is a contravariant vector, which moreover depends linearly on the arbitrarily chosen vector* p^s, *and satisfies identically for that choice of this vector* p^s *the invariant energy equation*
>
> $$\sum_l\frac{\partial\sqrt{g}e^l}{\partial w_l} \;=\; 0.\ ^{128} \tag{107}$$

127 In Paper 1, this is the only purpose for which this form of Theorem II is explicitly introduced. However, (23) presumably already had been derived from it.

While Hilbert did not explicitly introduce the condition that his energy vector be related to Mie's energy-momentum tensor, it seems to be the guiding principle of his calculation. Apparently, he wanted this connection to appear to be the result of an independently-justified definition of this vector.

In effect, starting from (106) and taking into account definitions (98) and (103), Hilbert obtained for the contribution to the energy originating from the electromagnetic term L in the Lagrangian:

$$Lp^l - \sum_k \frac{\partial L}{\partial q_{kl}} p_k - \frac{1}{2\sqrt{g}} \sum_{k,s} \frac{\partial}{\partial w_k} \left\{ \left(\frac{\partial \sqrt{g} L}{\partial q_{lk}} - \frac{\partial \sqrt{g} L}{\partial q_{kl}} \right) p^s q_s \right\}. \tag{108}$$

Using the field equations and (27), this can be rewritten as:

$$\sum_{s,k} \left(L\delta_s^l - \frac{\partial L}{\partial M_{lk}} M_{sk} - \frac{\partial L}{\partial q_l} q_s \right) p^s, \tag{109}$$

which corresponds to the right-hand side of (36), the generally-covariant generalization of Mie's electromagnetic energy-momentum tensor, contracted with p^s.

In contradistinction to the Proofs, Theorem II and (36) no longer explicitly enter this demonstration. Theorem II enters implicitly by determining the form in which the electromagnetic variables enter the Lagrangian (see (27)). Hilbert still needed Theorem II to derive his "first result," that is, to show that this energy-momentum can be written as the variational derivative of $\sqrt{g}L$ with respect to the gravitational potentials. Furthermore, (36) allows Hilbert to argue that, due to the field equations (see (72)), the electromagnetic energy and energy-vector e^l can be expressed exclusively in terms of K, the gravitational part of the Lagrangian; so that they depend only on the metric tensor and not on the electromagnetic potentials and their derivatives. Whereas, in the Proofs, this result had been an immediate consequence of the definition of the energy and of the field equations (see (49)), now it follows only with the help of Theorem II.

While Hilbert had succeeded in satisfying his heuristic criteria as well as the new challenge of deriving an invariant energy equation, the status of this equation within his theory had become more precarious. An analysis of the deductive structure of Hilbert's theory in Paper 1 (see Fig. 2) shows that it still comprises two main clusters of results: those concerning the implications of gravitation for electromagnetism and those concerning energy conservation. But the latter cluster is now even more isolated

128 "Definieren wir nunmehr [(106); (14) in the original text] als den *Energievektor, so ist der Energievektor ein kontravarianter Vektor, der noch von dem willkürlichen Vektor p^s linear abhängt und identisch für jene Wahl dieses Vektors p^s die invariante Energiegleichung* [(107)] *erfüllt*." (Hilbert 1916, 402)

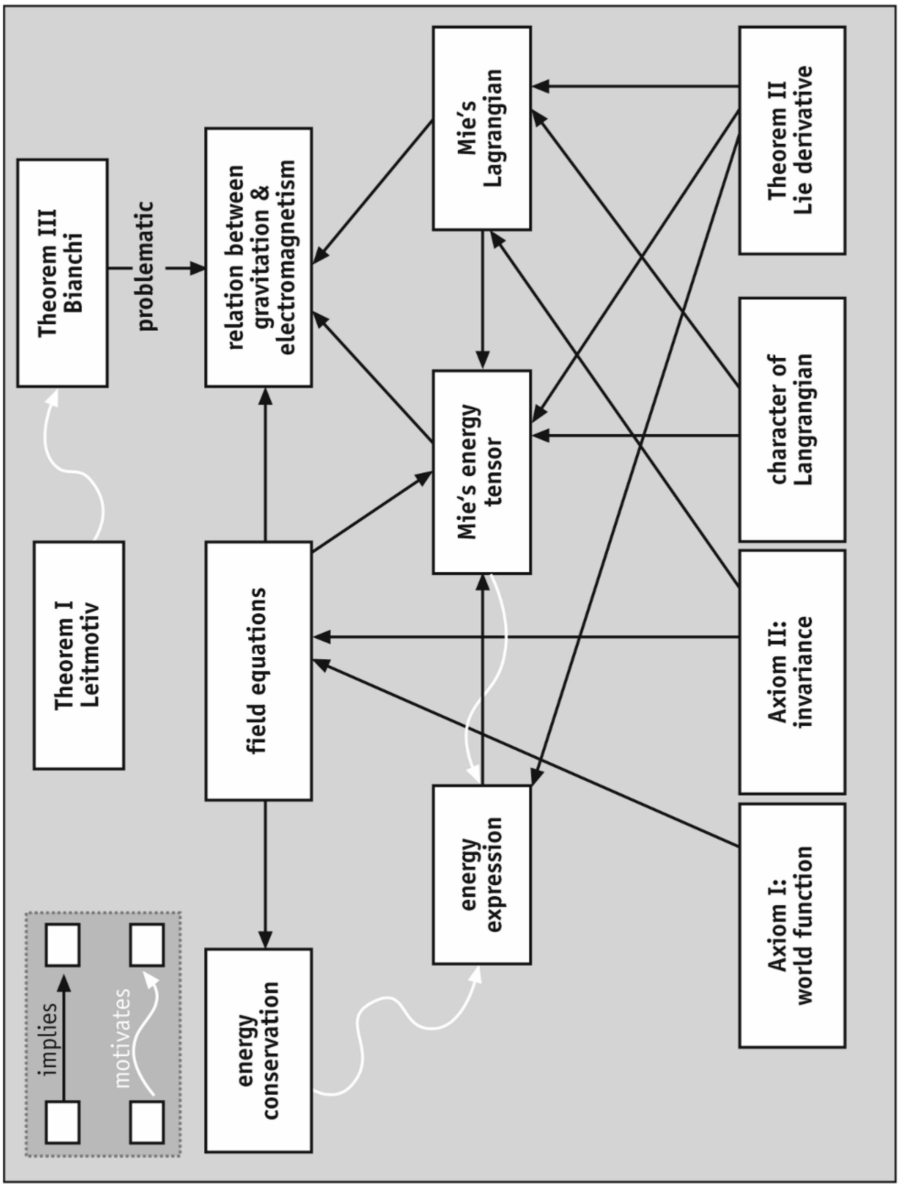


Figure 2: Deductive Structure of Paper 1 (1916)

from the rest of his theory than in the Proofs. Indeed, the new energy concept is no longer motivated by Hilbert's powerful Theorem I, but only by arguments concerning the formal properties of energy-momentum conservation and the link with Mie's energy-momentum tensor. It plays no role in deriving any other results of Hilbert's theory, nor does it serve to integrate this theory with other physical theories, a key function of the energy concept since its formulation in the 19th century. Therefore, it is not surprising that this concept only played a transitional role and was eventually replaced by the understanding of energy-momentum conservation developed by Einstein, Klein, Noether, and others.[129]

In fact, neither the physical significance nor the mathematical status of Hilbert's new energy concept was entirely clear. Physically Hilbert had failed to show that his energy equation (107) gave rise to a familiar expression for energy-momentum conservation in the special-relativistic limit, or to demonstrate that his equation was compatible with the form of energy-momentum conservation in a gravitational field that Einstein had established in 1913 (see (11)). Eventually, Felix Klein succeeded in clarifying the relation between Hilbert's and Einstein's expressions. He decomposed (107) into 140 equations and showed that 136 of these actually have nothing to do with energy-momentum conservation, while the remaining 4 correspond to those given by Einstein.[130] Mathematically, in 1917 Emmy Noether and Felix Klein found that equation (107) actually is an identity, and not a consequence of the field equations, as is the case for conservation equations in classical physics.[131] Similar identities follow for the Lagrangian of any generally-covariant variational problem. As a consequence, Hilbert's counting of equations no longer works: he assumed that his variational principle gives rise to 10 gravitational field equations plus 4 identities, which he identified with the electromagnetic equations; and that energy-momentum conservation is represented by additional equations, originally linked to coordinate restrictions. Einstein's abandonment of coordinate restrictions together with the deeper investigation of energy-momentum conservation by Noether, Klein, Einstein, and others, confronted Hilbert's approach with a severe challenge: They questioned the organization of his theory into two more-or-less independent domains, energy-momentum conservation and the implications of gravitation for electromagnetism. We shall argue that Hilbert responded to this challenge by further adapting his theory to the framework provided by general relativity.

5.2 Hilbert's Reorganization of His Theory in Paper 1

The challenge presented by Einstein's abandonment of coordinate restrictions and adoption of generally-covariant field equations forced Hilbert to reorganize his the-

---

129 For discussion, see (Rowe, 1999).

130 See (Klein 1918a, 179–185).

131 See (Klein 1917; 1918a) and also (Noether 1918). For a thorough discussion of the contemporary research on energy-momentum conservation, see (Rowe, 1999).

ory. As we have seen, he had to demonstrate the compatibility between his variational principle and Einstein's field equations (from which he had succeeded strikingly in deriving Mercury's perihelion shift), and completely rework his treatment of energy conservation. Hilbert treated both issues at the end of Paper 1. Energy conservation was no longer tied to Theorem I and its heuristic consequences as in the Proofs, but was treated along with other results of Hilbert's theory. The structure of Paper 1 is thus:[132]

1. Basic Framework (Hilbert 1916, 395–398)

   Axioms I and II, Theorem I, and the combined field equations of gravitation and electromagnetism for an arbitrary Lagrangian

2. Basic Theorems (Hilbert 1916, 398–400)

   Theorems II and III

3. New Energy Expression and Derivation of the New Energy Equation (Hilbert 1915, 400–402)

4. Implications for the Relation between Electromagnetism and Gravitation (Hilbert 1915, 402–407)

   the split of the Lagrangian into gravitational and the electrodynamical terms, the form of Mie's Lagrangian, its relation to his energy tensor, the explicit form of the gravitational field equations, and the relation between electromagnetic and gravitational field equations.

Apart from the technical and structural revisions necessitated by the new energy expression, practically all other changes concern the relation of his theory to Einstein's. Throughout Paper 1, Hilbert followed the tendency, already manifest in the marginal additions to the Proofs, to put greater emphasis on Einstein's contributions while maintaining his claim to have developed an independent approach. In the opening paragraph, Hilbert changed the order in which he mentioned Mie and Einstein. In the Proofs he wrote:

> The far reaching ideas and the formation of novel concepts by means of which Mie constructs his electrodynamics, and the prodigious problems raised by Einstein, as well as his ingeniously conceived methods of solution, have opened new paths for the investigation into the foundations of physics.[133]

In Paper 1 we read instead:

> The vast problems posed by Einstein as well as his ingeniously conceived methods of solution, and the far-reaching ideas and formation of novel concepts by means of which

132 For a sketch of Hilbert's revisions of Paper 1, see also (Corry 1999a, 517–522).

133 "Die tiefgreifenden Gedanken und originellen Begriffsbildungen vermöge derer Mie seine Elektrodynamik aufbaut, und die gewaltigen Problemstellungen von Einstein sowie dessen scharfsinnige zu ihrer Lösung ersonnenen Methoden haben der Untersuchung über die Grundlagen der Physik neue Wege eröffnet." (Proofs, 1)

> Mie constructs his electrodynamics, have opened new paths for the investigation into the foundations of physics.[134]

A footnote lists all of Einstein's publications on general relativity starting with his major 1914 review, and including the definitive paper submitted on 25 November. Although this makes clear that Hilbert must have revised his paper after that date, he failed to change the dateline of his contribution (as did Felix Klein and Emmy Noether in their contributions to the discussion of Hilbert's work in the same journal[135]). It remained "Vorgelegt in der Sitzung vom 20. November 1915," which creates the erroneous impression that there were no subsequent substantial changes in Paper 1.

The next sentence, while combining this claim with a more explicit recognition of what he considered the achievements of his predecessors, shows that Hilbert had not renounced his claim to having solved the problems posed by Mie and Einstein. In the corrected Proofs this sentence reads:

> In the following—in the sense of the axiomatic method — I would like to develop, /essentially from three simple axioms a ~~new~~ system of basic equations of physics, of ideal beauty, containing, I believe, the solution of the problems presented.[136]

In Paper 1, it reads:

> In the following — in the sense of the axiomatic method — I would like to develop, essentially from two simple axioms, a new system of basic equations of physics, of ideal beauty and containing, I believe, *simultaneously* the solution to the problems of Einstein and of Mie. I reserve for later communications the detailed development and particularly the special application of my basic equations to the fundamental questions of the theory of electricity.[137]

Although in a marginal note in the proofs version he had changed "his theory" to "the theory," he now returned to the original version:

> The guiding motive for constructing my theory is provided by the following theorem, the proof of which I shall present elsewhere.[138]

---

134 "Die gewaltigen Problemstellungen von Einstein sowie dessen scharfsinnige zu ihrer Lösung ersonnenen Methoden und die tiefgreifenden Gedanken und originellen Begriffsbildungen vermöge derer Mie seine Elektrodynamik aufbaut, haben der Untersuchung über die Grundlagen der Physik neue Wege eröffnet." (Hilbert 1916, 395)

135 See (Klein 1918a; Noether 1918).

136 "Ich möchte im Folgenden — im Sinne der axiomatischen Methode —/wesentlich aus drei einfachen Axiomen ein ~~neues~~ System von Grundgleichungen der Physik aufstellen, die von idealer Schönheit sind, und in denen, wie ich glaube, die Lösung der gestellten Probleme enthalten ist." (Proofs, 1)

137 "Ich möchte im Folgenden - im Sinne der axiomatischen Methode - wesentlich aus zwei einfachen Axiomen ein neues System von Grundgleichungen der Physik aufstellen, die von idealer Schönheit sind, und in denen, wie ich glaube, die Lösung der Probleme von Einstein und Mie gleichzeitig enthalten ist. Die genauere Ausführung sowie vor Allem die spezielle Anwendung meiner Grundgleichungen auf die fundamentalen Fragen der Elektrizitätslehre behalte ich späteren Mitteilungen vor." (Hilbert 1916, 395)

138 "Das Leitmotiv für den Aufbau meiner Theorie liefert der folgende mathematische Satz, dessen Beweis ich an einer anderen Stelle darlegen werde." (Hilbert 1916, 396)

Although Hilbert had earlier argued that his *Leitmotiv* suggested the need for four additional non-covariant equations to ensure a unique solution, he now dropped all mention of the subject of coordinate restrictions. He simply did not address the question of why, in spite of Einstein's hole argument against this possibility, it is possible to use generally-covariant field equations unsupplemented by coordinate restrictions. The only remnant in Paper 1 of the entire problem is his newly-introduced designation of the world-parameters as "allgemeinste Raum-Zeit-Koordinaten."

The significant result that Hilbert's variational principle gives rise to gravitational field equations formally equivalent to those of Einstein's 25 November theory is rather hidden in Hilbert's presentation, only appearing as an intermediate step in his demonstration that the electromagnetic field equations are a consequence of the gravitational ones. The newly-introduced passage reads:

> Using the notation introduced earlier for the variational derivatives with respect to the $g^{\mu\nu}$, the gravitational equations, because of (20) [i.e. (16)], take the form
>
> $$[\sqrt{g}K]_{\mu\nu} + \frac{\partial\sqrt{g}L}{\partial g^{\mu\nu}} = 0. \tag{110}$$
>
> The first term on the left hand side becomes
>
> $$[\sqrt{g}K]_{\mu\nu} = \sqrt{g}\Big(K_{\mu\nu} - \frac{1}{2}Kg_{\mu\nu}\Big), \tag{111}$$
>
> as follows easily without calculation from the fact that $K_{\mu\nu}$, apart from $g_{\mu\nu}$, is the only tensor of second rank and K the only invariant, that can be formed using only the $g^{\mu\nu}$ and their first and second differential quotients, $g_k^{\mu\nu}$, $g_{kl}^{\mu\nu}$.
>
> The resulting differential equations of gravitation appear to me to be in agreement with the grand concept of the theory of general relativity established by Einstein in his later treatises.[139]

Hilbert's argument for avoiding explicit calculation of $[\sqrt{g}K]_{\mu\nu}$, which he later withdrew (see below), is indeed untenable; there are many invariants and tensors of second rank that can be constructed from the Riemann tensor. Even if one further requires such tensors and invariants to be linear in the Riemann tensor, the crucial coefficient of the trace term still remains undetermined. The explicit form of the field equations given in Paper 1 and not found in the Proofs, appears to be a direct response to Einstein's publication of 25 November; but a footnote appended to this

139 "Unter Verwendung der vorhin eingeführten Bezeichungsweise für die Variationsableitungen bezüglich der $g^{\mu\nu}$ erhalten die Gravitationsgleichungen wegen (20) [i.e. (16)] die Gestalt [(110); (21) in the original text]. Das erste Glied linker Hand wird [(111)] wie leicht ohne Rechnung aus der Tatsache folgt, daß $K_{\mu\nu}$ außer $g_{\mu\nu}$ der einzige Tensor zweiter Ordnung und K die einzige Invariante ist, die nur mit den $g^{\mu\nu}$ und deren ersten und zweiten Differentialquotienten $g_k^{\mu\nu}, g_{kl}^{\mu\nu}$ gebildet werden kann.
Die so zu Stande kommenden Differentialgleichungen der Gravitation sind, wie mir scheint, mit der von Einstein in seinen späteren Abhandlungen aufgestellten großzügigen Theorie der allgemeinen Relativität im Einklang." (Hilbert 1916, 404–405)

passage gives a generic reference to all four of Einstein's 1915 Academy publications. His cautious reference to the apparent agreement between his results and Einstein's, presumably motivated by their different frameworks, adds to the impression that Hilbert actually arrived independently at the explicit form of the gravitational field equations.

The concluding paragraph of Paper 1 acknowledges Hilbert's debt to Einstein in a more indirect way. The beginning of this paragraph of the Proofs had given the impression that Einstein posed the problems while Hilbert offered the solutions:

> As one can see, the few simple assumptions expressed in axioms I, II, III suffice with appropriate interpretation to establish the theory: through it not only are our views of space, time, and motion fundamentally reshaped in the sense called for by Einstein ...[140]

In Paper 1, Hilbert deleted the reference to axiom III and replaced "in dem von Einstein geforderten Sinne" by "in dem von Einstein dargelegten Sinne":

> As one can see, the few simple assumptions expressed in axioms I and II suffice with appropriate interpretation to establish the theory: through it not only are our views of space, time, and motion fundamentally reshaped in the sense explained by Einstein ...[141]

5.3 Einstein's Energy in Hilbert's 1924 Theory

In 1924 Hilbert published revised versions of Papers 1 and 2 (Hilbert 1924).[142] Meanwhile important developments had taken place, such as the rapid progress of quantum physics, which changed the scientific context of Hilbert's results. But it was undoubtedly the further clarifications of the significance of energy-momentum conservation in general relativity, already mentioned in the preceding sections, that affected his theory most directly. In correspondence between Hilbert and Klein (published in part in 1918),[143] this topic played a central role without, however, leading to an explicit reformulation of Hilbert's theory. Without going into detail about this important strand in the history of general relativity, we shall focus on its effect on Hilbert's 1924 revisions. In spite of the reassertion of his goal of providing foundations for all of physics, his theory was, in effect, transformed into a variation on the themes of general relativity.

---

140 "Wie man sieht, genügen bei sinngemäßer Deutung die wenigen einfachen in den Axiomen I, II, III ausgesprochenen Annahmen zum Aufbau der Theorie: durch dieselbe werden nicht nur unsere Vorstellungen über Raum, Zeit und Bewegung von Grund aus in dem von Einstein geforderten Sinne umgestaltet ..." (Proofs, 13).

141 "Wie man sieht, genügen bei sinngemäßer Deutung die wenigen einfachen in den Axiomen I und II ausgesprochenen Annahmen zum Aufbau der Theorie: durch dieselbe werden nicht nur unsere Vorstellungen über Raum, Zeit und Bewegung von Grund aus in dem von Einstein dargelegten Sinne umgestaltet ..." (Hilbert 1916, 407).

142 In the following, we will refer to the 1924 revision of Paper 1 as "Part 1" and to that of Paper 2 as "Part 2," designations which correspond to Hilbert's own division of his 1924 paper into "Teil 1" (pp. 2–11) and "Teil 2" (pp. 11–32).

143 See (Klein 1917).

On a purely technical level, Hilbert's revisions of Paper 1 appear to be rather modest; the most important one concerns Theorem III (the contracted Bianchi identities), now labelled Theorem 2. Following a suggestion by Klein (Klein 1917, 471–472), Hilbert extended this theorem to include the electromagnetic variables:

> Theorem 2. Let J, as in Theorem 1, be an invariant depending on $g^{\mu\nu}$, $g_l^{\mu\nu}$, $g_{lk}^{\mu\nu}$, q_s, q_{sk}; and as above, let $[\sqrt{g}J]_{\mu\nu}$ denote the variational derivatives of $\sqrt{g}J$ with respect to $g^{\mu\nu}$, and $[\sqrt{g}J]_\mu$, the variational derivative with respect to q_μ. Introduce, furthermore, the abbreviations [(112)]:
>
> $$\begin{aligned} i_s &= \sum_{\mu,\nu}([\sqrt{g}J]_{\mu\nu}g_s^{\mu\nu} + [\sqrt{g}J]_\mu q_{\mu s}), \\ i_s^l &= -2\sum_\mu [\sqrt{g}J]_{\mu s}g^{\mu l} + [\sqrt{g}J]_l q_s, \end{aligned} \tag{112}$$
>
> then the [following] identities hold
>
> $$i_s = \sum_l \frac{\partial i_s^l}{\partial x_l} \qquad (s = 1, 2, 3, 4).^{144} \tag{113}$$

He revised its proof accordingly.

A second, small, but significant change concerns the gravitational field equations. Hilbert now tacitly withdrew his previous claim that no derivation was needed, instead sketching a derivation and writing them, like Einstein, with the energy-momentum tensor as source. As in the earlier versions, he derived (72) but now in the form:[145]

$$[\sqrt{g}K]_{\mu\nu} = -\frac{\partial\sqrt{g}L}{\partial g^{\mu\nu}}. \tag{114}$$

After writing down the electromagnetic field equations, Hilbert proceeded to sketch the following evaluation of the terms in (114):

> To determine the expression for $[\sqrt{g}K]_{\mu\nu}$, first specialize the coordinate system so that at the world point under consideration all the $g_s^{\mu\nu}$ vanish. In this way one finds:
>
> $$[\sqrt{g}K]_{\mu\nu} = \sqrt{g}\Big(K_{\mu\nu} - \frac{1}{2}g_{\mu\nu}K\Big). \tag{115}$$
>
> If, for the tensor
>
> $$-\frac{1}{\sqrt{g}}\frac{\partial\sqrt{g}L}{\partial g^{\mu\nu}} \tag{116}$$
>
> we introduce the symbol $T_{\mu\nu}$, then the gravitational field equations can be written as

144 "Theorem 2. Wenn J, wie im Theorem 1, eine von $g^{\mu\nu}$, $g_l^{\mu\nu}$, $g_{lk}^{\mu\nu}$, q_s, q_{sk} abhängige Invariante ist, und, wie oben, die Variationsableitungen von $\sqrt{g}J$ bez. $g^{\mu\nu}$ mit $[\sqrt{g}J]_{\mu\nu}$, bez. q_μ mit $[\sqrt{g}J]_\mu$ bezeichnet werden, und wenn ferner zur Abkürzung: [(112)] gesetzt wird, so gelten die Identitäten [(113); (7) in the original text]." (Hilbert 1924, 5)

145 See (Hilbert 1924, 7).

$$K_{\mu\nu} - \frac{1}{2} g_{\mu\nu} K = T_{\mu\nu}.^{146} \tag{117}$$

Although the introduction of Einstein's notation for the energy-momentum tensor may appear as no more than an adaptation of Hilbert's notation to the by-then standard usage, it actually effected a major revision in the structure of his theory. The energy-momentum tensor became the central knot binding together the physical implications of Hilbert's theory.

First of all, it served, as Hilbert's energy expressions had previously done, to relate the derivative of Mie's Lagrangian (see (34) or (36)) to Mie's energy-momentum tensor. But, in contrast to Paper 1, Mie's energy-momentum tensor no longer served as a criterion for choosing the energy-expression. The new energy expression, which Hilbert now took over from Einstein, was supported by much more than just this single result. It had emerged from the development of special-relativistic continuum physics by Minkowski, Abraham, Planck, Laue,[147] and others; and been validated by numerous applications to various areas of physics, including general relativity.

By introducing the equation:

$$T_{\mu\nu} = -\frac{\partial \sqrt{g} L}{\partial g^{\mu\nu}}, \tag{118}$$

Hilbert had returned, in a sense, to the approach of the Proofs, establishing a relation between the energy concept and the derivative of the electromagnetic Lagrangian (see (49)). He still did not make clear that this relation does not single out Mie's theory, but actually holds more generally. Introducing the notations:

$$\frac{\partial L}{\partial q_{sk}} = \frac{\partial L}{\partial M_{ks}} = H^{ks}, \tag{119}$$

and:

$$\frac{\partial L}{\partial q_k} = r^k, \tag{120}$$

As in the proofs version, Hilbert again used (35), which he now rewrites as:

$$-\frac{2}{\sqrt{g}} \sum_{\mu} \frac{\partial \sqrt{g} L}{\partial g^{\mu\nu}} g^{\mu m} = L \delta_\nu^m - \sum_s H^{ms} M_{\nu s} - r^m q_\nu, \tag{121}$$

146 "Um den Ausdruck von $[\sqrt{g}K]_{\mu\nu}$ zu bestimmen, spezialisiere man zunächst das Koordinatensystem so, daß für den betrachteten Weltpunkt die $g_s^{\mu\nu}$ sämtlich verschwinden. Man findet auf diese Weise: [(115)]. Führen wir noch für den Tensor [(116)] die Bezeichnung $T_{\mu\nu}$ ein, so lauten die Gravitationsgleichungen [(117)]." See (Hilbert 1924, 7–8).

147 For the first systematic development of relativistic continuum mechanics, see (Laue 1911a; 1911b). For further discussion, see Einstein's "Manuscript on the Special Theory of Relativity" (CPAE 4, Doc. 1, 91–98; Janssen and Mecklenburg 2006).

(see (36)). On the basis of this equation, Hilbert claims, in almost exactly the same words as in the earlier versions, that there is a necessary connection between the theories of Mie and Einstein:

> Hence the [following] representation of $T_{\mu\nu}$ results:
>
> $$T_{\mu\nu} = \sum_{\mu} g_{\mu m} T_{\nu}^{m}$$
> $$T_{\nu}^{m} = \frac{1}{2}\left\{L\delta_{\nu}^{m} - \sum_{s} H^{ms} M_{\nu s} - r^{m} q_{\nu}\right\}. \tag{122}$$
>
> The expression on the right agrees with Mie's electromagnetic energy tensor, and thus we find that Mie's electromagnetic energy tensor is nothing but the generally-invariant tensor resulting from differentiation of the invariant L with respect to the gravitational potentials $g^{\mu\nu}$ — a circumstance which gave me the first hint of the necessary close connection between Einstein's theory of general relativity and Mie's electrodynamics, and which convinced me of the correctness of the theory developed here.[148]

While Hilbert's claim remained unchanged, what he had done actually was to specialize the source term left arbitrary in Einstein's field equations. The nature of this source term can be specified on the level of the Lagrangian or of the energy-momentum tensor, and these two ways are obviously equivalent if a Lagrangian exists—but this relation is in no way peculiar to Mie's theory. The fact that the energy expression in Paper 1 was specifically chosen to produce Mie's energy-momentum tensor had obscured this circumstance, now made rather obvious by the introduction of Einstein's arbitrary energy-momentum tensor. It was no doubt difficult for Hilbert to draw this conclusion because it contradicted his program, according to which electromagnetism should arise as an effect of gravitation.

The situation was similar for Hilbert's second important application of Einstein's energy-momentum tensor, the derivation of a relation between the gravitational and electromagnetic field equations. After recognition of the close relation between the contracted Bianchi identities and energy-momentum conservation in general relativity, it was necessary for Hilbert to reconsider the link he believed he had established between the two groups of field equations. Energy-momentum conservation now played a central role in his approach, turning the link between gravitation and electromagnetism into a mere by-product. It existed, not because of any deep intrinsic connection between these two areas of physics, but due to the introduction of electromagnetic potentials into the variational principle. With the same logic, one

148 "Demnach ergibt sich für $T_{\mu\nu}$ die Darstellung: [(122)]. Der Ausdruck rechts stimmt überein mit dem Mie'schen elektromagnetischen Energietensor, und wir finden also, daß der Mie'sche elektromagnetische Energietensor ist nichts anderes als der durch Differentiation der Invariante L nach den Gravitationspotentialen $g^{\mu\nu}$ entstehende allgemein invariante Tensor—ein Umstand, der mich zum ersten Mal auf den notwendigen engen Zusammenhang zwischen der Einsteinschen allgemeinen Relativitätstheorie und der Mie'schen Elektrodynamik hingewiesen und mir die Überzeugung von der Richtigkeit der hier entwickelten Theorie gegeben hat." (Hilbert 1924, 9)

could argue that *any* form of matter giving rise to a stress-energy tensor derivable from a Lagrangian involving the metric tensor is an effect of gravitation.

This weakened link is reflected in Hilbert's new way of obtaining the desired link between gravitation and electromagnetism. Following Klein's suggestion, in Part 1 Hilbert treated the contracted Bianchi identities in parallel for both the gravitational and the electromagnetic terms in the Lagrangian:

> The application of Theorem 2 to the invariant K yields:
>
> $$\sum_{\mu\nu}[\sqrt{g}K]_{\mu\nu}g_s^{\mu\nu} + 2\sum_m \frac{\partial}{\partial x_m}\Big(\sum_\mu [\sqrt{g}K]_{\mu s}g^{\mu m}\Big) = 0. \tag{123}$$
>
> Its application to L yields:[149]
>
> $$\begin{aligned}&\sum_{\mu\nu}(-\sqrt{g}T_{\mu\nu})g_s^{\mu\nu} + 2\sum_m \frac{\partial}{\partial x_m}(-\sqrt{g}T_s^{\ m})\\ &+\sum_\mu [\sqrt{g}L]_\mu q_{\mu s} - \sum_\mu \frac{\partial}{\partial x_\mu}([\sqrt{g}L]_\mu q_s) = 0 \qquad (s = 1, 2, 3, 4).\end{aligned} \tag{124}$$

Previously, he had derived only the first set of identities and made use of them in order to derive (83). Now Hilbert showed that both sets of identities yield the equations for energy-momentum conservation that had been central to Einstein's work since 1912. Following the work of Einstein and others, Hilbert also made clear that these equations are related to the equations of motion for the sources of the stress-energy tensor,[150] and represent a generalization of energy-momentum conservation laws in special relativity:

> As a consequence of the basic equations of electrodynamics, we obtain from this:
>
> $$\sum_{\mu\nu}\sqrt{g}T_{\mu\nu}g_s^{\mu\nu} + 2\sum_m \frac{\partial}{\partial x_m}\sqrt{g}T_s^{\ m} = 0. \tag{125}$$
>
> These equations also result as a consequence of the gravitational equations due to (15a) [i.e. (123)]. Their interpretation is that they are the basic equations of mechanics. In the case of special relativity, when the $g_{\mu\nu}$ are constants, they reduce to the equations
>
> $$\sum \frac{\partial T_s^{\ m}}{\partial x_m} = 0, \tag{126}$$
>
> which express the conservation of energy and momentum.[151]

---

149 "Die Anwendung des Theorems 2 auf die Invariante K liefert: [(123); (15a) in the original text.] Die Anwendung auf L ergibt: [(124); (15b) in the original text.]" (Hilbert 1924, 9–10)

150 See (Havas 1989, Klein 1917; 1918a; 1918b).

151 "Als Folge der elektrodynamischen Grundgleichungen erhalten wir hieraus: [(125); (16) in the original text.] Diese Gleichungen ergeben sich auch als Folge der Gravitationsgleichungen, auf Grund von (15a) [i.e. (123)]. Sie haben die Bedeutung der mechanischen Grundgleichungen. Im Falle der speziellen Relativität, wenn die $g_{\mu\nu}$ Konstante sind, gehen sie über in die Gleichungen [(126)] welche die Erhaltung von Energie und Impuls ausdrücken." (Hilbert 1924, 10)

Hilbert thus anchored his theory in the same physical foundation that had provided Einstein's search for general relativity with a stable point of reference. Only after having done this did Hilbert turn to his original goal, the link between gravitation and electromagnetism, the problematic character of which we have discussed above:

> From the identities (15b) [i.e. (124)], there follow from the equations (16) [i.e. (125)]:
>
> $$\sum_{\mu}[\sqrt{g}L]_{\mu}q_{\mu s}-\sum_{\mu}\frac{\partial}{\partial x_{\mu}}([\sqrt{g}L]_{\mu}q_{s}) = 0 \tag{127}$$
>
> or
>
> $$\sum_{\mu}\left\{M_{\mu s}[\sqrt{g}L]_{\mu}+q_{s}\frac{\partial}{\partial x_{\mu}}[\sqrt{g}L]_{\mu}\right\} = 0; \tag{128}$$
>
> i.e., four independent linear relations between the basic equations of electrodynamics (5) and their first derivations follow from the gravitational equations (4). This is the precise mathematical expression of the connection between gravitation and electrodynamics, which dominates the entire theory.[152]

The deductive structure of Part 1 shows the fundamental changes with respect to Paper 1 (see Fig. 3) and the central role of Einstein's energy-momentum tensor in this reorganization. In fact, this tensor suggested the particular form in which Hilbert rewrote the gravitational field equations, established the link between gravitation and electromagnetism (in terms of the choice of a specific source), and, of course, was fundamental to Hilbert's new formulation of energy-momentum conservation.

This revised deductive structure has a kernel, consisting of the variational principle, field equations, and energy-momentum conservation, that is—both from a formal and a physical perspective —fully equivalent to the kernel of Einstein's formulation of general relativity. Clearly, Hilbert's deductive presentation places greater emphasis on a variational principle than does Einstein; and the mathematically more elegant formulation of the variational principle, based on the Ricci scalar, contributes to this emphasis. Therefore, this variational formulation of general relativity is today rightly associated with Hilbert's name. On the other hand, Hilbert's original aim, the derivation of electromagnetism as an effect of gravitation, plays only a marginal role in Part 1 and still suffers from the problems indicated above. The links between the main components that had substantiated Hilbert's claim of a special relation between Mie's theory and Einstein's have been weakened, being held together only by the choice of a specific source. This link is thus no longer central to an approach presenting an alternative to that of Einstein, being little more than an attempt to supplement

152 "Aus den Gleichungen (16) [i.e. (125)] folgt auf Grund der Identitäten (15b) [i.e. (124)]: [(127)] oder [(128); (17) in the original text] d.h. aus den Gravitationsgleichungen (4) folgen vier voneinander unabhängige lineare Relationen zwischen den elektrodynamischen Grundgleichungen (5) und ihren ersten Ableitungen. Dies ist der genaue mathematische Ausdruck für den Zusammenhang zwischen Gravitation und Elektrodynamik, der die ganze Theorie beherrscht." See the comments on (83), (Hilbert 1924, 10).

Einstein's general framework with a specific physical content, Mie's electrodynamics—an attempt that is now based on the firm foundations of general relativity.

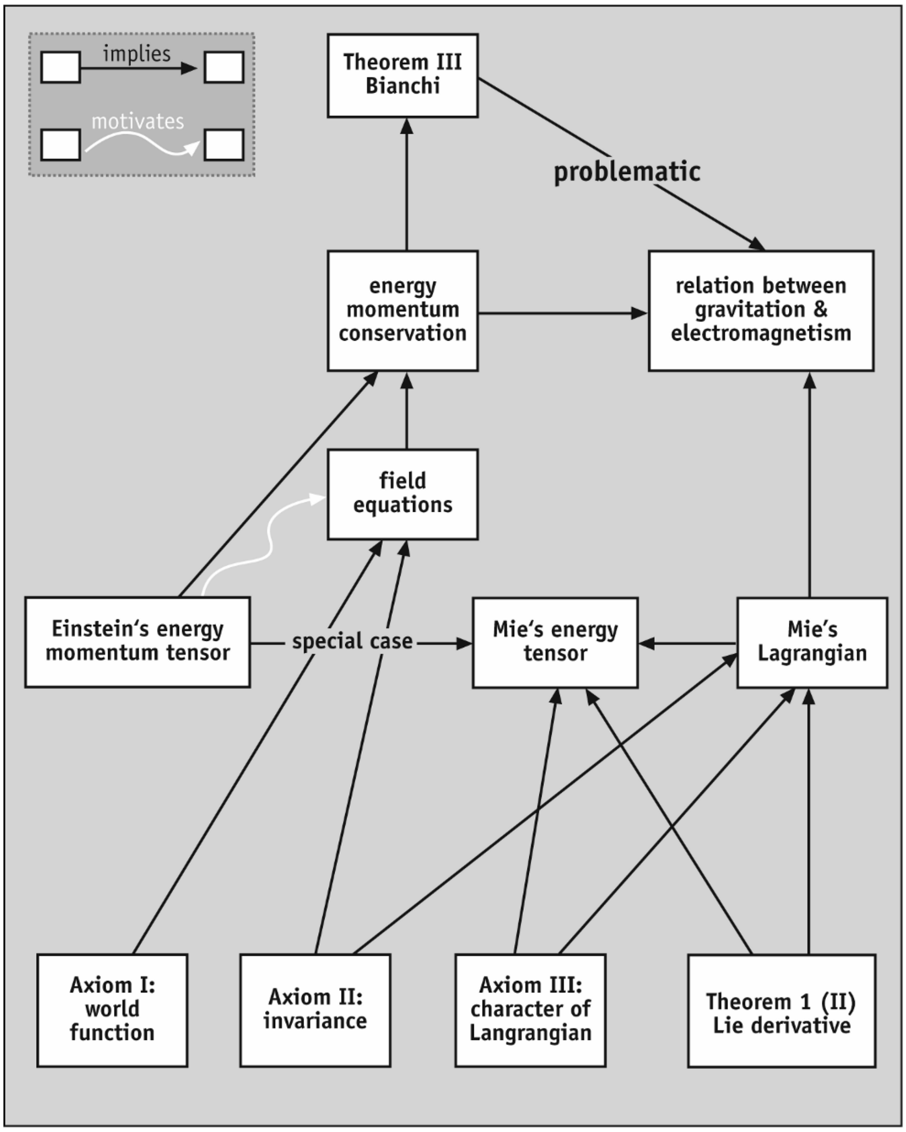


Figure 3: Deductive Structure of Part 1 (1924)

5.4 A Scientist's History

Scientists rarely investigate carefully the often only small and gradual conceptual transformations that their insights undergo in the course of historical development, often at the hands of others. Instead of undertaking such a demanding enterprise with little promise of new scientific results, they rather tend to hold onto their insights, reinterpreting them in the light of their present and prospective uses rather than in the light of past achievements, let alone failures. As we shall see, this tendency was inescapable for Hilbert, who understood the progress of physics in terms of an elaboration of the apparently universal and immutable concepts of classical physics.

Indeed, Hilbert described the 1924 Part 1 version of his theory not as a revision of his 1916 Paper 1 version, including major conceptual adjustments and a reorganization of its deductive structure, but essentially as a reprint of his earlier work:

> What follows is essentially a reprint of both of my earlier communications on the *Grundlagen der Physik*, and my comments on them, which were published by F. Klein in his communication *Zu Hilberts erster Note über die Grundlagen der Physik*, with only minor editorial differences and transpositions in order to facilitate their understanding.[153]

Indeed, the organization of Part 1 has not undergone major changes as compared to Paper 1, but seems to represent simply a tightening up; it can be subdivided into the following sections:

1. General Introduction (Hilbert 1924, 1–2)
2. Basic Setting (Hilbert 1924, 2–4)
   Axioms I and II, field equations of electromagnetism and gravitation
3. Basic Theorems (Hilbert 1924, 4–7)
   Theorems 1 (previously II) and 2 (previously III), the theorem earlier designated as Theorem I (now without numbering)
4. Implications for Electromagnetism, Gravitational Field Equations, and Energy-momentum Conservation (Hilbert 1924, 7–11)
   The character of the gravitational part of the Lagrangian, Axiom III (the split of the Lagrangian and the character of the electrodynamical part of the Lagrangian), the gravitational field equations, the form of Mie's Lagrangian, the relation between Mie's energy tensor and Mie's Lagrangian, energy-momentum conservation, and the relation between electromagnetic and gravitational field equations.

The most noteworthy changes in the order of presentation are: a new introductory section and the integration of the treatment of energy-momentum conservation with other results of Hilbert's theory towards the end. Another conspicuous change is that

153 "Das Nachfolgende ist im wesentlichen ein Abdruck der beiden älteren Mitteilungen von mir über die *Grundlagen der Physik* und meiner Bemerkungen dazu, die F. Klein in seiner Mitteilung *Zu Hilberts erster Note über die Grundlagen der Physik* veröffentlicht hat—mit nur geringfügigen redaktionellen Abweichungen und Umstellungen, die das Verständnis erleichtern sollen." (Hilbert 1924, 1)

Hilbert's *Leitmotiv*, Theorem I of Paper 1, has now lost its central place despite meanwhile having been proven by Emmy Noether. As we have seen, even in Paper 1 it no longer played the key heuristic role for Hilbert that it had originally in the Proofs. As the preceding discussion made clear, the rather unchanged form of its presentation hides major changes in the substance of his theory.

These changes are reflected in the introductory section, in a way that again downplays them.

While earlier Hilbert had introduced his own contribution as a solution to the problems raised by Mie and Einstein (Proofs) or Einstein and Mie (Paper 1), he now characterized his results as providing a simple and natural representation of Einstein's general theory of relativity, completed in formal aspects:

> The vast complex of problems and conceptual structures of Einstein's general theory of relativity now find, as I explained in my first communication, their simplest and most natural expression and, in its formal aspect, a systematic supplementation and completion by following the route trodden by Mie.[154]

In view of the overwhelming contemporary impact of Einstein's theory, Mie's role was downplayed in Hilbert's new version. Mie is no longer portrayed as posing problems of a similar profundity to those of Einstein, but as inspiring Hilbert's "simplest and most natural" presentation of general relativity, as well as "a systematic supplementation and completion in its formal aspect."

Instead of attributing a specific role in contemporary scientific discussions to Mie, Hilbert elevates him to the role of one of the founding fathers of a unified-field theoretical worldview:

> The mechanistic ideal of unity in physics, as created by the great researchers of the previous generation and still adhered to during the reign of classical electrodynamics, now must be definitively abandoned. Through the creation and development of the field concept, a new possibility for the comprehension of the physical world has gradually taken shape. Mie was the first to show a way that makes accessible to general mathematical treatment this newly risen 'field theoretical ideal of unity' as I would like to call it.[155]

Curiously neither Einstein nor Minkowski are mentioned in Hilbert's discussion of the spacetime continuum as the "foundation" of "the new field-theoretical ideal":

154 "Die gewaltigen Problemstellungen und Gedankenbildungen der allgemeinen Relativitätstheorie von Einstein finden nun, wie ich in meiner ersten Mitteilung ausgeführt habe, auf dem von Mie betretenen Wege ihren einfachsten und natürlichsten Ausdruck und zugleich in formaler Hinsicht eine systematische Ergänzung und Abrundung." (Hilbert 1924, 1–2) The changes in Hilbert's theory were accompanied by a change in his attitude to Einstein's achievement, by which he was increasingly impressed: see (Corry 1999a, 522–525).

155 "Das mechanistische Einheitsideal in der Physik, wie es von den großen Forschern der vorangegangenen Generation geschaffen und noch während der Herrschaft der klassischen Elektrodynamik festgehalten worden war, muß heute endgültig aufgegeben werden. Durch die Aufstellung und Entwickelung des Feldbegriffes bildete sich allmählich eine neue Möglichkeit für die Auffassung der physkalischen Welt aus. Mie zeigte als der erste einen Weg, auf dem dieses neuenstandene "feldtheoretische Einheitsideal", wie ich es nennen möchte, der allgemeinen mathematischen Behandlung zugänglich gemacht werden kann." (Hilbert 1924, 1)

> While the old mechanistic conception takes matter itself as a direct starting point and assumes it to be determined by a finite range of discrete parameters; a physical continuum, the so-called spacetime manifold, rather serves as the foundation of the new field-theoretical ideal. While previously universal laws took the form of [ordinary] differential equations with one independent variable, now partial differential equations are their necessary form of expression.[156]

Mie was exalted to the otherwise rather empty heaven of the founding fathers, leaving room for Hilbert's attempts at a unified theory of gravitation and electromagnetism. He generously mentioned other contemporary efforts as off-springs of his own contribution, a view hardly shared by his contemporaries (see below):

> Since the publication of my first communication, significant papers on this subject have appeared: I mention only Weyl's magnificent and profound investigations, and Einstein's communications, filled with ever new approaches and ideas. In the meantime, even Weyl took a turn in his development that led him too to arrive at just the equations I formulated; and on the other hand Einstein also, although starting repeatedly from divergent approaches, differing among themselves, ultimately returns, in his latest publication, to precisely the equations of my theory.[157]

This passage from Hilbert leaves unspecified to which of his equations he is referring. Given his references to Weyl and Einstein, he must mean the two sets of field equations (51) and (52), which are rather obvious ingredients of any attempted unification of gravitation and electromagnetism. The unique feature of his approach, the specific connection he introduced between these two sets of equations (see (83)) constituting the mathematical expression of electrodynamics as a phenomenon following from gravitation, had become highly problematic and was not adopted by either Weyl or Einstein.

Indeed, it was already problematic whether Weyl's and Einstein's attempts at unification were any more fortunate than Hilbert's. In his concluding paragraph, Hilbert himself expressed his doubts, which were based on the rapid progress of quantum physics, on the one hand, and the lack of any concrete physical results of such theories, on the other:

> Whether the pure field theoretical ideal of unity is indeed definitive, and what possible supplements and modifications of it are necessary to enable in particular the theoretical foundation for the existence of negative and positive electrons, as well as the consistent

---

156 "Während die alte mechanistische Auffassung unmittelbar die Materie selbst als Ausgang nimmt und diese durch eine endliche Auswahl diskreter Parameter bestimmt ansetzt, dient vielmehr dem neuen feldtheoretischen Ideal das physikalische Kontinuum, die sogenannte Raum-Zeit-Mannigfaltigkeit, als Fundament. Waren früher Differenzialgleichungen mit einer unabhängigen Variablen die Form der Weltgesetze, so sind jetzt notwendig partielle Differenzialgleichungen ihre Ausdrucksform." (Hilbert 1924, 1)

157 "Seit der Veröffentlichung meiner ersten Mitteilung sind bedeutsame Abhandlungen über diesen Gegenstand erschienen: ich erwähne nur die glänzenden und tiefsinnigen Untersuchungen von Weyl und die an immer neuen Ansätzen und Gedanken reichen Mitteilungen von Einstein. Indes sowohl Weyl gibt späterhin seinem Entwicklungsgange eine solche Wendung, daß er auf die von mir aufgestellten Gleichungen ebenfalls gelangt, und andererseits auch Einstein, obwohl wiederholt von abweichenden und unter sich verschiedenen Ansätzen ausgehend, kehrt schließlich in seinen letzten Publikationen geradewegs zu den Gleichungen meiner Theorie zurück." (Hilbert 1924, 2)

> development of the laws holding in the interior of the atom—to answer this is the task for the future.[158]

In spite of his doubts, Hilbert was convinced that "his theory" would endure, (see the preceding paragraph), expressing the belief that it was of programmatic significance for future developments. Even if not, at least philosophical benefit could be drawn from it:

> I am convinced that the theory I have developed here contains an enduring core and creates a framework within which there is sufficient scope for the future development of physics in the sense of a field theoretical ideal of unity. In any case, it is also of epistemological interest to see how the few, simple assumptions I put forth in Axioms I, II, III, and IV suffice for the construction of the entire theory.[159]

The fact that his theory is not based exclusively on these axioms, but also depends rather crucially on other physical concepts, such as energy, and that his theory might change in content as well structure if these concepts changes their meaning,—all of this evidently remained outside of Hilbert's epistemological scope.

6. HILBERT'S ADOPTION OF EINSTEIN'S PROGRAM: THE SECOND PAPER AND ITS REVISIONS

6.1 From Paper 1 to Paper 2

When Hilbert published his Paper 1 in early 1916, he still hoped that his unification of electromagnetism and gravitation would provide the basis for solving the riddles of microphysics. He opened his paper announcing:

> I reserve for later communications the detailed development and particularly the special application of my basic equations to the fundamental questions of the theory of electricity.[160]

and concluding:

> ... I am also convinced that through the basic equations established here the most intimate, presently hidden processes in the interior of the atom will receive an explanation,

158 "Ob freilich das reine feldtheoretische Einheitsideal ein definitives ist, evtl. welche Ergänzungen und Modifikationen desselben nötig sind, um insbesondere die theoretische Begründung für die Existenz des negativen und des positiven Elektrons, sowie den widerspruchsfreien Aufbau der im Atominneren geltenden Gesetze zu ermöglichen,—dies zu beantworten, ist die Aufgabe der Zukunft." (Hilbert 1924, 2)

159 "Ich glaube sicher, daß die hier von mir entwickelte Theorie einen bleibenden Kern enthält und einen Rahmen schafft, innerhalb dessen für den künftigen Aufbau der Physik im Sinne eines feldtheoretischen Einheitsideals genügender Spielraum da ist. Auch ist es auf jeden Fall von erkenntnistheoretischem Interesse, zu sehen, wie die wenigen einfachen in den Axiomen I, II, III, IV von mir ausgesprochenen Annahmen zum Aufbau der ganzen Theorie genügend sind." (Hilbert 1924, 2)

160 "Die genauere Ausführung sowie vor Allem die spezielle Anwendung meiner Grundgleichungen auf die fundamentalen Fragen der Elektrizitätslehre behalte ich späteren Mitteilungen vor." (Hilbert 1916, 395)

> and in particular that generally a reduction of all physical constants to mathematical constants must be possible ...[161]

Clearly, he intended to dedicate a second communication to the physical consequences of his theory. By March 1916 he had submitted a second installment, which was then withdrawn, no trace remaining.[162] What does remain are the notes of Hilbert's SS 1916 and WS 1916/17 Lectures, and his related Causality Lecture. The WS 1916/17 Lectures offer hints of how his theory would lead to a modification of Maxwell's equations near the sources. While this part is clearly still related to Hilbert's original project, the bulk of these notes testify to his careful study of current work by Einstein and others on general relativity, as well as containing original contributions to that project. In the second communication to the Göttingen Academy submitted at the end of December 1916 (hereafter referred to as "Paper 2"), work on general relativity occupied the entire paper (Hilbert 1917). Hilbert's lecture notes are important for understanding the transition from his original aims to Paper 2, as well as the contents of this paper.[163] One of the most remarkable features of these notes is the openness and informality with which Hilbert shares unsolved problems with his students, later explicitly stating that this was a central goal of his lectures:

> In lectures, and above all in seminars, my guiding principle was not to present material in a standard and as smooth as possible way, just to help the students to maintain ordered notebooks. Above all, I tried to illuminate the problems and difficulties and offer a bridge leading to currently open questions. It often happened that in the course of a semester the program of an advanced lecture was completely changed because I wanted to discuss issues in which I was currently involved as a researcher and which had not yet by any means attained their definite formulation.[164]

6.2 The Causality Quandary

The lecture notes make it clear that Hilbert was still in a quandary over the treatment of causality because his Proofs argument against general covariance seemed to remain valid. The bulk of the typescript notes of his SS 1916 Lectures deal with special relativity (which he calls "die kleine Relativität"): kinematics, and vector and

---

161 "... ich bin auch der Überzeugung, daß durch die hier aufgestellten Grundgleichungen die intimsten, bisher verborgenen Vorgänge innerhalb des Atoms Aufklärung erhalten werden und insbesondere allgemein eine Zurückführung aller physikalischen Konstanten auf mathematische Konstanten möglich sein muß ..." (Hilbert 1916, 407)

162 See the discussion in (Sauer 1999, 560 n. 129).

163 The importance of Hilbert's lectures has been emphasized by Leo Corry. See (Corry 2004).

164 "Es war mein Grundsatz, in den Vorlesungen und erst recht in den Seminaren nicht einen eingefahrenen und so glatt wie möglich polierten Wissensstoff, der den Studenten das Führen sauberer Kolleghefte erleichtert, vorzutragen. Ich habe vielmehr immer versucht, die Probleme und Schwierigkeiten zu beleuchten und die Brücke zu den aktuellen Fragen zu schlagen. Nicht selten kam es vor, daß im Verlauf eines Semesters das stoffliche Programm einer höheren Vorlesung wesentlich abgeändert wurde, weil ich Dinge behandeln wollte, die mich gerade als Forscher beschäftigten und die noch keineswegs eine endgültige Gestalt gewonnen hatten." (Reidemeister 1971, 79) Translation by Leo Corry.

tensor analysis (pp. 1–66); dynamics (pp. 66–70 and 76–82); and Maxwell's electrodynamics (pp. 70–76 and 84–89). Hilbert then discusses Mie's theory in its original, special-relativistic form (pp. 90–102), and the need to combine it with "Einstein's concept of the general relativity of events" ("des Einstein'schen Gedankens von der allgemeinen Relativität des Geschehens," p. 103). After introducing the metric tensor, he develops the field equations for gravitation and electromagnetism (pp. 103–111). Discussing these equations, he notes that the causality problem remains unsolved:

> These are 14 equations for the 14 unknown functions $g^{\mu\nu}$ and q_h $(\mu, \nu = 1 \ldots 4)$. The causality principle may or may not be satisfied (the theory has not yet clarified this point). In any event, unlike the case of Mie's theory, the validity of this principle cannot be inferred from simple considerations. Of these 14 equations, 4 (e.g., the 4 Maxwell equations) are a consequence of the remaining 10 (e.g., the gravitational equations). Indeed, the remarkable theorem holds that the number of equations following from Hamilton's principle always corresponds to the number of unknown functions, except in the case occurring here, that the integral is an ["a general" added by hand] invariant.[165]

He still had not resolved the causality problem when he continued the lectures during the winter semester. Among other things, the WS 1916/17 Lecture notes contain much raw material for Paper 2. For example, the discussion of causal relations between events in a given spacetime very much resembles the treatment in that paper.[166] Yet the notes do not discuss the causality question for the field equations.

The same answer to this problem presented in Paper 2 is given in the typescript (unfortunately undated) of his Causality Lecture. From its contents, it is reasonable to conjecture that this is Hilbert's first exposition of his newly-found solution. After discussing the problem for his generally-covariant system of equations and constructing an example to illustrate its nature (pp. 1–5), he comments:

> Einstein's old theory now amounts to the addition of 4 non-invariant equations. But this too is mathematically incorrect. Causality cannot be saved in this way.[167]

---

165 "Dies sind 14 Gleichungen für die 14 unbekannten Funktionen $g^{\mu\nu}$ und q_h $(\mu, \nu, h = 1 \ldots 4)$. Das Kausalitätsprinzip kann erfüllt sein, oder nicht (Die Theorie hat diesen Punkt noch nicht aufgeklärt). Jedenfalls lässt sich auf die Gültigkeit dieses Prinzips nicht wie im Falle der Mie'schen Theorie durch einfache Ueberlegungen schliessen. Von diesen 14 Gleichungen sind nämlich 4 (z.B. die 4 Maxwellschen) eine Folge der 10 übrigen (z.B. der Gravitationsgleichungen). Es gilt nämlich der merkwürdige Satz, dass die Zahl der aus dem Hamiltonschen Prinzip fliessenden Gleichungen immer mit der Zahl der unbekannten Funktionen übereinstimmt, ausser in dem hier eintretenden Fall, das unter dem Integral ["eine allgemeine" added by hand] Invariante steht." (SS 1916 Lectures, 110)

166 See Chapter XIII of the notes, *Einiges über das Kausalitätsprinzip in der Physik,* 97–103, and pp. 57–59 of Paper 2, both of which are discussed below.

167 "Die alte Theorie von Einstein läuft nun darauf hinaus, 4 nicht invariante Gleichungen hinzuzufügen. Aber auch dies ist mathematisch falsch. Auf diesem Wege kann die Kausalität nicht gerettet werden" (p. 5). As discussed above, in his *Entwurf* theory Einstein did not first set up a system of generally-covariant equations and then supplement them by non-invariant conditions; but started from non-generally-covariant field equations. But he had considered the possibility described by Hilbert that these equations have a generally-covariant counterpart, from which they could be obtained by imposing non-invariant conditions.

A similar comment appears in Paper 2:

> In his original theory, now abandoned, A. Einstein (*Sitzungsberichte der Akad. zu Berlin*, 1914, p. 1067) had indeed postulated certain 4 non-invariant equations for the $g_{\mu\nu}$, in order to save the causality principle in its old form.[168]

Neither here nor in any later publication does Hilbert repeat the claim in the lecture notes that this procedure (which he himself had followed in the Proofs) is "mathematisch falsch," which strongly suggests that the notes precede Paper 2.

This suggested temporal sequence is confirmed by another pair of passages: In his lecture, Hilbert compares the problem created by general covariance of a system of partial differential equations and that created by parameter invariance in the calculus of variations:

> The difficulty of having to distinguish between a meaningful and a meaningless assertion is also encountered in Weierstrass's calculus of variations. There the curve to be varied is assumed to be given in parametric form, and one then obtains a differential equation for two unknown functions. One then considers only those assertions that remain invariant when the parameter p is replaced by an arbitrary function of p. [169]

This comparison may well have played a significant role in his solution of the causality problem. The corresponding passage in Paper 2 generalizes this comparison:

> In the theory of curves and surfaces, where a statement in a chosen parametrization of the curve or surface has no geometrical meaning for the curve or surface itself, if this statement does not remain invariant under an arbitrary transformation of the parameters or cannot be brought to invariant form; so also in physics we must characterize a statement that does not remain invariant under any arbitrary transformation of the coordinate system as *physically meaningless*.[170]

This argument is so much more general that it is hard to believe that, once he had hit upon it, Hilbert would have reverted to its restricted application to extremalization of curves. So we shall assume the priority of the Causality Lecture notes.

In these notes, Hilbert asserts that the causality quandary can be resolved by an appropriate understanding of physically meaningful statements:

168 "In seiner ursprünglichen, nunmehr verlassenen Theorie hatte A. Einstein (Sitzungsberichte der Akad. zu Berlin. 1914 S. 1067) in der Tat, um das Kausalitätsprinzip in der alten Fassung zu retten, gewisse 4 nicht invariante Gleichungen für die $g_{\mu\nu}$ besonders postuliert." (Hilbert 1917, 61)

169 "Auf die Schwierigkeit, zwischen einer sinnvollen und einer sinnlosen Behauptung unterscheiden zu müssen, stösst man übrigens auch in der Weierstrass'schen Variationsrechnung. Dort wird die zu variierende Kurve als in Parametergestalt gegeben angenommen, und man erhält dann eine Differentialgleichung für zwei unbekannte Funktionen. Man betrachtet dann nur solche Aussagen, die invariant bleiben, wenn man den Parameter p durch eine willkürliche Funktion von p ersetzt." (Causality Lecture, 8)

170 "Gerade so wie in der Kurven- und Flächentheorie eine Aussage, für die die Parameterdarstellung der Kurve oder Fläche gewählt ist, für die Kurve oder Fläche selbst keinen geometrischen Sinn hat, wenn nicht die Aussage gegenüber einer beliebigen Transformation der Parameter invariant bleibt oder sich in eine invariante Form bringen läßt, so müssen wir auch in der Physik eine Aussage, die nicht gegenüber jeder beliebigen Transformation des Koordinatensystems invariant bleibt, als *physikalisch sinnlos* bezeichnen." (Hilbert 1917, 61)

> We obtain the explanation of this paradox by attempting to more rigorously grasp the concept of relativity. It does not suffice to say that the laws of the world are independent of the frame of reference, but rather every single assertion about an event or a concurrence of events only then takes on a physical meaning if it is independent of its designation, i.e. when it is invariant.[171]

In the last clause, one hears distant echoes of Einstein's assertion in his expository paper *Die Grundlage der allgemeinen Relativitätstheorie*:

> We allot to the universe four spacetime variables x_1, x_2, x_3, x_4 in such a way that for every point-event there is a corresponding system of values of the variables $x_1 \ldots x_4$. To two coincident point-events there corresponds one system of values of the variables $x_1 \ldots x_4$, i.e. coincidence is characterized by the identity of the co-ordinates. ... As all our physical experience can be ultimately reduced to such coincidences, there is no immediate reason for preferring certain systems of coordinates to others, that is to say, we arrive at the requirement of general co-variance.[172]

Perusal of this paper, published on 11 May 1916 and cited in Hilbert's WS 1916/17 Lectures,[173] may well have contributed to his new understanding of the causality problem.

However, Hilbert's interpretation of a physically meaningful statement actually differs from that of Einstein. Einstein had turned the uniqueness problem for solutions of generally-covariant field equations into an argument against the physical significance of coordinate systems. Hilbert attempted to turn the problem into its own solution by *defining* physically meaningful statements as those for which no such ambiguities arise, whether such statements employ coordinate systems or not. In his Causality Lecture, Hilbert claims to demonstrate the validity of the "causality principle," formulated in terms of physically meaningful statements:

> We would like to prove that the causality principle formulated as follows: "All meaningful assertions are a necessary consequence of the preceding ones [see the citation above]" is valid. Only this theorem is logically necessary and, for physics, also completely sufficient.[174]

To establish this principle, he considers an arbitrary set of generally-covariant field equations (which he calls "ein System invarianter Gleichungen") involving the

---

171 "Die Aufklärung dieses Paradoxons erhalten wir, wenn wir nun den Begriff der Relativität schärfer zu erfassen suchen. Man muss nämlich nicht nur sagen, dass die Weltgesetze vom Bezugssystem unabhängig sind, es hat vielmehr jede einzelne Behauptung über eine Begebenheit oder ein Zusammentreffen von Begebenheiten physikalisch nur dann einen Sinn, wenn sie von der Benennung unabhängig, d.h. wenn sie invariant ist." (Causality Lecture, 5–6)

172 "Man ordnet der Welt vier zeiträumliche Variable x_1, x_2, x_3, x_4 zu, derart, dass jedem Punktereignis ein Wertsystem der Variablen $x_1 \ldots x_4$ entspricht. Zwei koinzidierenden Punktereignissen entspricht dasselbe Wertsystem der Variablen $x_1 \ldots x_4$; d. h. die Koinzidenz ist durch die Übereinstimmung der Koordinaten charakterisiert. .... Da sich alle unsere physikalischen Erfahrungen letzten Endes auf solche Koinzidenzen zurückführen lassen, ist zunächst kein Grund vorhanden, gewisse Koordinatensysteme vor anderen zu bevorzugen, d.h. wir gelangen zu der Forderung der allgemeinen Kovarianz." (Einstein 1916a, 776–777)

173 See (WS 1916/17 Lectures, 112).

metric tensor, the electromagnetic potentials, and their derivatives.[175] He specifies the values of these fields and their derivatives on the space-like hypersurface $t = 0$, which he calls "the present" ("die Gegenwart"); and considers coordinate transformations that do not change the coordinates on this hypersurface, but are otherwise arbitrary (except for continuity and differentiability) off the hypersurface ("die Transformation soll die Gegenwart ungeändert lassen"). He then defines a physically meaningful statement as one that is uniquely determined by Cauchy data, intending to thus establish, at the same time, his principle of causality in terms of what one might call "a mathematical response" to the problem of uniqueness in a generally-covariant field theory:

> Only such a [meaningful assertion] is unequivocally determined by the initial values of $g_{\mu\nu}$, q_μ and their derivatives, and in fact these initial values are to be understood as Cauchy boundary-value conditions. It must be accepted that one can prescribe these boundary values arbitrarily, or that one can proceed to a place in the world at the moment in time when the state characterized by these values prevails. The observer of nature is also considered as standing outside these physical laws; otherwise one would arrive at the antinomies of free will.[176]

As this passage makes clear, Hilbert's proposed definition of physically meaningful statements and clarification of the problem of causality is flawed by the still-unrecognized intricacies of the Cauchy problem in general relativity. He evidently failed to realize that the classical notion of freely-choosable initial values no longer works for generally-covariant field equations since some of them function as constraints on the data that can be given on an initial hypersurface, rather than as evolution equations for that data off this surface. The next section discusses Hilbert's treatment of the problem of causality in Paper 2, including further evidence of his failure to fully grasp Einstein's insight that, in general relativity, coordinate systems have no physical significance of their own.

---

174 "Wir wollen beweisen, dass das so formulierte Kausalitätsprinzip: "Alle sinnvollen Behauptungen sind eine notwendige Folge der vorangegangenen [see the citation above]" gültig ist. Dieser Satz allein ist logisch notwendig und er ist auch für die Physik vollkommen ausreichend." (Causality Lecture, 5–6)

175 The original typescript had specified first and second derivatives of the metric and first derivatives of the electromagnetic potentials, but by hand Hilbert added "beliebig hohen" in the first case and deleted "ersten" in the second.

176 "Nur eine solche [sinnvolle Behauptung] ist durch die Anfangswerte der $g_{\mu\nu}$, q_μ und ihrer Ableitungen eindeutig festgelegt und zwar sind diese Anfangswerte als Cauchy'sche Randbedingungen zu verstehen. Dass man diese Randwerte beliebig vorgeben kann, oder dass man sich an eine Stelle der Welt hinbegeben kann, wo der durch diese Werte charakterisierte Zustand in diesem Zeitmoment herrscht, muss hingenommen werden. Der die Natur beobachtende Mensch wird eben als ausserhalb dieser physikalischen Gesetze stehend betrachtet; sonst käme man zu den Antinomien der Willensfreiheit." (Causality Lecture, 6–7)

6.3 Hilbert at Work on General Relativity

Paper 2 shows that Hilbert's original goal of developing a unified gravito-electromagnetic theory, with the aim of explaining the structure of the electron and the Bohr atom, has been modified in the light of the successes of Einstein's purely gravitational program. Hilbert's shift of emphasis in Paper 1 to the primacy of the gravitational field equations must have facilitated his shift to the consideration of the "empty-space" field equations. From Hilbert's perspective, they are just that subclass of solutions to his fourteen "unified" field equations, for which the electromagnetic potentials vanish. This makes them formally equivalent to the sub-class of solutions to Einstein's field equations with a stress-energy tensor that either vanishes everywhere, or at least outside of some finite world-tube containing the sources of the field. This formal equivalence no doubt contributed to the ease with which contemporary mathematicians and physicists assimilated Hilbert's program to Einstein's, treating Paper 2 as a contribution to the development of the general theory of relativity. This is how Hilbert's contribution came to be assimilated to the relativistic tradition, as we shall discuss in more detail below.

Let us now take a look at the six major topics Hilbert treated in Paper 2:

1. measurement of the components of the metric tensor (Hilbert 1917, 53–55);
2. characteristics and bicharacteristics of the Hamilton-Jacobi equation corresponding to the metric tensor (Hilbert 1917, 56–57);
3. causal relation between events in a spacetime with given metric (Hilbert 1917, 57–59);
4. the causality problem for the field equations determining the metric tensor (Hilbert 1917, 59–63);
5. Euclidean geometry as a solution to the field equations—in particular, the investigation of conditions that characterize it as a unique solution (Hilbert 1917, 63–66 and 70); and
6. the Schwarzschild solution, its derivation (Hilbert 1917, 67–70), and determination of the paths of (freely-falling) particles and light rays in it (Hilbert 1917, 70–76).

1) The metric tensor and its measurement: First of all, Hilbert dropped his previous use of one imaginary coordinate, perhaps influenced by Einstein's use of real coordinates, and emphasized that the $g_{\mu\nu}$, now all real, provide the "Massbestimmung einer Pseudogeometrie" (Hilbert 1917, 54). He classified the elements ("Stücke") of all curves: time-like elements measure proper time; space-like elements measure length; and null elements are segments of a light path. He introduced two ideal measuring instruments: a measuring tape ("Maßfaden") for lengths, and a light clock ("Lichtuhr") for proper times. He makes a comment that suggests, in spite of his remarks in Paper 1 and the Causality Lecture (see above), a lingering belief in some objective significance to the choice of a coordinate system, independently of the metric tensor:

> First we show that each of the two instruments suffices to compute with its aid the values of the $g_{\mu\nu}$ as functions of x_s, just as soon as a definite spacetime coordinate system x_s has been introduced.[177]

He ends with some comments on a possible axiomatic construction ("Aufbau") of the pseudogeometry, suggesting the need for two axioms:

> first an axiom should be established, from which it follows that length resp. proper time must be integrals whose integrand is only a function of the x_s and their first derivatives with respect to the parameter [p, where $x_s = x_s(p)$ is the parametric representation of a curve]; ...
>
> Secondly an axiom is needed whereby the theorems of the pseudo-Euclidean geometry, that is the old principle of relativity, shall be valid in infinitesimal regions;[178]

2) Characteristics and bicharacteristics: Hilbert defined the null cone at each point, and pointed out that the Monge differential equation (Hilbert 1917, 56):

$$g_{\mu\nu}\frac{dx_\mu}{dp}\frac{dx_\nu}{dp} = 0, \tag{129}$$

and the corresponding Hamilton-Jacobi partial differential equation:

$$g^{\mu\nu}\frac{\partial f}{\partial x_\mu}\frac{\partial f}{\partial x_\nu} = 0, \tag{130}$$

determine the resulting null cone field, the geodesic null lines being the characteristics of the first and the bicharacteristics of the second of these equations. The null geodesics emanating from any world point form the null conoid ("Zeitscheide;" many current texts apply the term "null cone" to non flat spacetimes, but we prefer the term "conoid") emanating from that point. He points out that the equation for these conoids are integral surfaces of the Hamilton-Jacobi equation; and that all timelike world lines emanating from a world point lie inside its conoid, which forms their boundary.

These topics, rather briefly discussed in Paper 2, are treated much more extensively in Hilbert's WS 1916/17 Lectures. In many ways Hilbert's discussion in

---

177 "Zunächst zeigen wir, daß jedes der beiden Instrumente ausreicht, um mit seiner Hülfe die Werte der $g_{\mu\nu}$ als Funktion von x_s zu berechnen, sobald nur ein bestimmtes Raum-Zeit-Koordinatensystem x_s eingeführt worden ist." (Hilbert 1917, 55)

178 "erstens ist ein Axiom aufzustellen, auf Grund dessen folgt, daß Länge bez. Eigenzeit Integrale sein müssen, deren Integrand lediglich eine Funktion der x_s und ihrer ersten Ableitungen nach dem Parameter ist; ...
Zweitens ist ein Axiom erforderlich, wonach die Sätze der pseudo-Euklidischen Geometrie d.h. das alte Relativitätsprinzip im Unendlichkleinen gelten soll;" (Hilbert 1917, 56)

Paper 2 reads like a précis of these notes; it becomes much more intelligible if they are consulted. Chapter IX (pp. 69–80) entitled "Die Monge'sche Differentialgleichung" also treats the Hamilton-Jacobi equation and the theory of characteristics, emphasizing their relation to the Cauchy problem, and the reciprocal relation between integral surfaces of the Hamilton-Jacobi equation (the null conoids are called "transzendentale Kegelfläche") and null curves. Chapters X (pp. 80–82, "Die vierdimensionale eigentliche u. Pseudogeometrie") and XI (pp. 82–97, "Zusammenhang mit der Wirklichkeit") cover the material in the first section of Paper 2: the measuring tape ("Massfaden") is discussed in section 38 (pp. 85–86 and pp. 91–92), and the light clock, already introduced in the context of special relativity (see the SS 1916 Lectures, 6–10), is reintroduced in section 44 (pp. 93–94, "Axiomatische Definition der Lichtuhr"). Both instruments are used to determine the components of the metric tensor as functions of the coordinates, "sobald nur ein bestimmtes Raum-Zeit Koordinatensystem x_i eingeführt worden ist" (p. 95).

3) Causal relation between events: [179] In accord with the implicit requirement that three of the coordinates be space-like and one time-like, Hilbert imposes corresponding conditions on the components of the metric tensor. But he has a unique way of motivating them:

> Up to now all coordinate systems x_s that result from any one by arbitrary transformation have been regarded as equally valid. This arbitrariness must be restricted when we want to realize the concept that two world points on the same time line can be related as cause and effect, and that it should then no longer be possible to transform such world points to be simultaneous. In declaring x_4 as the *true* time coordinate we adopt the following definition:
>
> ...
>
> So we see that the concepts of cause and effect, which underlie the principle of causality, also do not lead to any inner contradictions whatever in the new physics, if we only take the inequalities (31) always to be part of our basic equations, that is if we confine ourselves to using *true* spacetime coordinates.[180]

Again, he seems to believe that there is some residual physical significance in the choice of a coordinate system: it must reflect the relations of cause and effect between events on the same time-like world line. He defines a proper ("eigentliches") coordinate system as one, in which (in effect) the first three coordinates are space-like and the fourth time-like in nature; transformations between such proper coordinate systems are also called proper. Given Hilbert's stated goal of restricting the choice of coordinates to those that reflect the causal order on all time-like world lines, his con-

179 This section also includes material from Hilbert's WS 1916/17 Lectures: Chapter XII, *Einiges über das Kausalitätsprinzip in der Physik,* (pp. 97–104) covers the same ground as, but in no more detail than, the text of Paper 2.

ditions are sufficient but not necessary since they exclude retarded null coordinates, which also preserve this causal order.

4) Causality problem for the field equations: As noted, Hilbert's analysis follows his Causality Lecture. In Paper 2 he writes:

> Concerning the principle of causality, let the physical quantities and their time derivatives be known at the present in some given coordinate system: then a statement will only have physical meaning if it is invariant under all those transformations, for which the coordinates just used for the present remain unchanged; I maintain that statements of this type for the future are all uniquely determined, that is, *the principle of causality holds in this form*:
>
> *From present knowledge of the 14 physical potentials* $g_{\mu\nu}$, q_s *all statements about them for the future follow necessarily and uniquely provided they are physically meaningful*.[181]

A hasty reading might suggest that Hilbert is asserting the independence of all physically meaningful statements from the choice of a coordinate system, and he has often been so interpreted; but this is not what he actually says. His very definition of physically meaningful ("physikalisch Sinn haben") involves the class of coordinate systems that leave the coordinates on the initial hypersurface ("die Gegenwart") unchanged. Secondly, Hilbert uses a Gaussian coordinate system, introduced earlier,[182] in order to prove his assertion about the causality principle.[183] Finally, if his words were so interpreted, they would stand in flagrant contradiction to his earlier statements (cited above)

---

180 "Bisher haben wir alle Koordinatensysteme x_s die aus irgend einem durch eine willkürliche Transformation hervorgehen, als gleichberechtigt angesehen. Diese Willkür muß eingeschränkt werden, sobald wir die Auffassung zur Geltung bringen wollen, daß zwei auf der nämlichen Zeitlinie gelegene Weltpunkte im Verhältnis von Ursache und Wirkung zu einander stehen können und daß es daher nicht möglich sein soll, solche Weltpunkte auf gleichzeitig zu transformieren.

...

So sehen wir, daß die dem Kausalitätsprinzip zu Grunde liegenden Begriffe von Ursache und Wirkung auch in der neuen Physik zu keinerlei inneren Widersprüche führen, sobald wir nur stets die Ungleichungen (31) [the conditions Hilbert imposes on the metric tensor] zu unseren Grundgleichungen hinzunehmen d.h. uns auf den Gebrauch *eigentlicher* Raum-zeitkoordinaten beschränken." (Hilbert 1917, 57 and 58)

181 "Was nun das Kausalitätsprinzip betrifft, so mögen für die Gegenwart in irgend einem gegebenen Koordinatensystem die physikalischen Größen und ihre zeitlichen Ableitungen bekannt sein: dann wird eine Aussage nur physikalisch Sinn haben, wenn sie gegenüber allen denjenigen Transformationen invariant ist, bei denen eben die für die Gegenwart benutzten Koordinaten unverändert bleiben; ich behaupte, daß die Aussagen dieser Art für die Zukunft sämtlich eindeutig bestimmt sind d.h. das Kausalitätsprinzip gilt in dieser Fassung:

Aus der Kenntnis der 14 physikalischen Potentiale $g_{\mu\nu}$, q_s *in der Gegenwart folgen alle Aussagen über dieselben für die Zukunft notwendig und eindeutig, sofern sie physikalischen Sinn haben*." (Hilbert 1917, 61)

182 See (Hilbert 1917, 58–59).

183 See (Hilbert 1917, 61–62).

about the measurement of the metric and the causal relation between events which presuppose attaching some residual physical meaning to the choice of coordinates.

His proof consists of a brief discussion of the Cauchy problem for the field equations in a Gaussian coordinate system. One of us has discussed this aspect of his work elsewhere (Stachel 1992), so we shall be brief here. He only considers the ten gravitational field equations (51) since he interprets Theorem I of Paper 1 as showing that the other four (52) follow from them. Gaussian coordinates eliminate four of the 14 field quantities, the $g_{0\mu}$, leaving only ten (the six g_{ab}, $a, b = 1, 2, 3$, and the four q_s), so he concludes that the resulting system of equations is in Cauchy normal form. This treatment is erroneous on several counts, but we postpone discussion of this question until the next section. More relevant to the present topic is Hilbert's statement:

> Since the Gaussian coordinate system itself is uniquely determined, therefore also all statements about those potentials (34) [the ten potentials mentioned above] with respect to these coordinates are of invariant character.[184]

He never discusses the behavior of the initial data under coordinate transformations on the initial hypersurface (three-dimensional hypersurface diffeomorphisms in modern terminology), confirming that his treatment is still tied to the use of particular coordinate systems rather than being based on coordinate-invariant quantities.

Finally, his discussion of how to implement the requirement of physically meaningful assertions depends heavily on the choice of a coordinate system. He remarks:

> The forms, in which physically meaningful, i.e. invariant, statements can be expressed mathematically are of great variety.[185]

and proceeds to discuss three ways:

> *First*. This can be done by means of an invariant coordinate system. ...
>
> *Second*. The statement, *according to which a coordinate system can be found* in which the 14 potentials $g_{\mu\nu}$, q_s have certain definite values in the future, or fulfill certain definite conditions, is always an invariant and therefore a physically meaningful one. ...
>
> *Third*. A statement is also invariant and thus has physical meaning if it is supposed to be valid in any arbitrary coordinate system.[186]

---

184 "Da das Gaußische Koordinatsystem selbst eindeutig festgelegt ist, so sind auch alle auf dieses Koordinatensystem bezogenen Aussagen über jene Potentiale (34) von invariantem Charakter." (Hilbert 1917, 62)

185 "Die Formen in denen physikalisch sinnvolle d.h. invariante Aussagen mathematisch zum Ausdruck gebracht werden können, sind sehr mannigfaltig." (Hilbert 1917, 62)

186 "Erstens. Dies kann mittelst eines invarianten Koordinatensystem geschehen. ...
Zweitens. Die Aussage, wonach sich ein Koordinatensystem finden läßt, in welchem die 14 Potentiale $g_{\mu n}$, q_s für die Zukunft gewisse bestimmte Werte haben oder gewisse Beziehungen erfüllen, ist stets eine invariante und daher physikalisch sinnvoll. ...
Drittens. Auch ist eine Aussage invariant und hat daher stets physikalisch Sinn, wenn sie für jedes beliebige Koordinatensystem gültig sein soll." (Hilbert 1917, 62–63)

The first two ways explicitly depend on the choice of a coordinate system, which is not necessarily unique. As examples of the first way, he cites Gaussian and Riemannian coordinates. It is true that, discussing the second, he notes:

> The mathematically invariant expression for such a statement is obtained by eliminating the coordinates from those relations.[187]

But he does not give an example, nor does he suggest the most obvious way of realizing his goal, if indeed it was a coordinate-independent solution to the problem: the use of invariants as coordinates. As Kretschmann noted a few years later, in matter- and field-free regions the four non-vanishing invariants of the Riemann tensor may be used as coordinates. If the metric is then expressed as a function of these coordinates, its components themselves become invariants.[188] The use of such coordinates was taken up again by Arthur Komar in the 1960s, and today they are often called Kretschmann-Komar coordinates.[189]

One might think that Hilbert had in mind something like this in his third suggested way. However, the example he cites makes it clear that he meant something else:

> An example of this are Einstein's energy-momentum equations having divergence character. For, although Einstein's energy [that is, the gravitational energy-momentum pseudotensor] does not have the property of invariance, and the differential equations he put down for its components are by no means covariant as a system of equations, nevertheless the assertion contained in them, that they shall be satisfied in any coordinate system, is an invariant demand and therefore it carries physical meaning.[190]

Rather than invariant quantities, evidently he had in mind non-tensorial entities and sets of equations, which nevertheless take the same form in every coordinate system.

In summary, Hilbert's treatment in Paper 2 of the problem of causality in general relativity still suffers from many of the flaws in his original approach. In particular, physical significance is still ascribed to coordinate systems, and the claim is maintained that the identities following from Theorem I represent a coupling between the two sets of field equations. On the other hand, his efforts to explore the solutions of the gravitational field equations from the perspective of a mathematician produced significant contributions to general relativity, to be discussed later.

---

187 "Der mathematische invariante Ausdruck für eine solche Aussage wird durch Elimination der Koordinaten aus jenen Beziehungen erhalten." (Hilbert 1917, 62–63)

188 See (Kretschmann 1917).

189 See (Komar 1958).

190 "Ein Beispiel dafür sind die Einsteinschen Impuls-Energiegleichungen vom Divergenz Character. Obwohl nämlich die Einsteinsche Energie die Invarianteneigenschaft nicht besitzt und die von ihm aufgestellten Differentialgleichungen für ihre Komponenten auch als Gleichungssystem keineswegs kovariant sind, so ist doch die in ihnen enthaltene Aussage, daß sie für jedes beliebige Koordinatensystem erfüllt sein sollen, eine invariante Forderung und hat demnach einen physikalischen Sinn." (Hilbert 1917, 63)

5) Euclidean geometry: This section opens with some extremely interesting general comments contrasting the role of geometry in what Hilbert calls the old and the new physics:

> The old physics with the concept of absolute time took over the theorems of Euclidean geometry and without question put them at the basis of every physical theory. ...
>
> The new physics of Einstein's principle of general relativity takes a totally different position vis-à-vis geometry. It takes neither Euclid's nor any other particular geometry *a priori* as basic, in order to deduce from it the proper laws of physics, but, as I showed in my first communication, the new physics provides at one fell swoop through one and the same Hamilton's principle the geometrical and the physical laws, namely the basic equations (4) and (5) [the ten gravitational and four electromagnetic field equations], which tell us how the metric $g_{\mu\nu}$ —at the same time the mathematical expression of the phenomenon of gravitation—is connected with the values q_s of the electrodynamic potentials.[191]

Hilbert declares:

> With this understanding, an old geometrical question becomes ripe for solution, namely whether and in what sense Euclidean geometry—about which we know from mathematics only that it is a logical structure free from contradictions—also possesses validity in the real world.[192]

He later formulates this question more precisely:

> The geometrical question mentioned above amounts to the investigation, whether and under what conditions the four-dimensional Euclidean pseudo-geometry [i.e., the Minkowski metric] ... is a solution, or even the only regular solution, of the basic physical equations.[193]

Hilbert thus takes up a problem that emerged with the development of non-Euclidean geometry in the 19th century and considered by such eminent mathematicians as Gauss and Riemann: the question of the relation between geometry and physical real-

---

191 "Die alte Physik mit dem absoluten Zeitbegriff übernahm die Sätze der Euklidische Geometrie und legte sie vorweg einer jeden speziellen physikalischen Theorie zugrunde. ...
Die neue Physik des Einsteinschen allgemeinen Relativitätsprinzips nimmt gegenüber der Geometrie eine völlig andere Stellung ein. Sie legt weder die Euklidische noch irgend eine andere bestimmte Geometrie vorweg zu Grunde, um daraus die eigentlichen physikalischen Gesetze zu deduzieren, sondern die neue Theorie der Physik liefert, wie ich in meiner ersten Mitteilung gezeigt habe, mit einem Schlage durch ein und dasselbe Hamiltonsche Prinzip die geometrischen und die physikalischen Gesetze nämlich die Grundgleichungen (4) und (5), welche lehren, wie die Maßbestimmungen $g_{\mu\nu}$ — zugleich der mathematischen Ausdruck der physikalischen Erscheinung der Gravitation — mit den Werten q_s der elektrodynamischen Potentiale verkettet ist." (Hilbert 1917, 63–64)

192 "Mit dieser Erkenntnis wird nun eine alte geometrische Frage zur Lösung reif, die Frage nämlich, ob und in welchem Sinne die Euklidische Geometrie — von der wir aus der Mathematik nur wissen, daß sie ein logisch widerspruchsfreier Bau ist — auch in der Wirklichkeit Gültigkeit besitzt." (Hilbert 1917, 63)

193 "Die oben genannte geometrische Frage läuft darauf hinaus, zu untersuchen, ob und unter welchen Voraussetzungen die vierdimensionale Euklidische Pseudogeometrie ... eine Lösung der physikalischen Grundgleichungen bez. die einzige reguläre Lösung derselben ist." (Hilbert 1917, 64)

ity. For a number of reasons, this question was not central to Einstein's heuristic. He had never addressed the question posed by Hilbert: the conditions under which Minkowski spacetime is a unique solution to the gravitational field equations. To Einstein, the question of the Newtonian limit, and hence the incorporation of Newton's theory into his new theory of gravitation, was much more important than the question of the existence of matter-free solutions to his equations. Indeed, this question was a rather embarrassing one for Einstein since such solutions display inertial properties of test particles even in the absence of matter, a feature that he had difficulty in accepting because of his Machian conviction that all inertial effects must be due to interaction of masses.[194] By establishing a connection between general relativity and the mathematical tradition questioning the geometry of physical space, Hilbert made a significant contribution to the foundations of general relativity.

In attempting to answer the question of the relation between Minkowski spacetime and his equations, Hilbert first of all notes that, if the electrodynamic potentials vanish, then the Minkowski metric is a solution of the resulting equations, i.e., of the vanishing of what we now call the Einstein tensor.[195] He then poses the converse question: under what conditions is the Minkowski metric the *only* regular solution to these equations? He considers small perturbations of the Minkowski metric (a technique that Einstein had already introduced) and shows that, if these perturbations are time independent (curiously, here reverting to use of an imaginary time coordinate) and fall off sufficiently rapidly and regularly at infinity, then they must vanish everywhere. In the next section of the paper, he proves another relevant result, which we shall discuss below.

This section of Paper 2 is a condensation of material covered in his WS 1916/17 Lectures:

- in the table of contents (p. 197), pp. 104–106 are entitled: "Der Sinn der Frage: Gilt die Euklidische Geometrie?"
- pp. 109–111 are headed "Gilt die Euklidische Geometrie in der Physik?" in the typescript, with the handwritten title "Die Grundgleichungen beim Fehlen von Materie" added in the margin, and entitled "Aufstellung der Grundgleichungen beim Fehlen der Materie" in the table of contents; and
- pp. 111–112, bear the handwritten title "Zwei Sätze über die Gültigkeit der Euklidischen Geometrie" in the margin, and "Zwei noch unbewiesene Sätze über die Gültigkeit der Pseudoeuklidischen Geometrie in der Physik" in the table of contents.

The lecture notes make much clearer than Paper 2 Hilbert's motivation for a discussion of the empty-space field equations in general, and of the Schwarzschild metric in particular. In the notes, Hilbert introduces the field equations in section 51 (WS

194 For a historical discussion, see (Renn 1994).

195 "wenn alle Elektrizität entfernt ist, so ist die pseudo-Euklidische Geometrie möglich" See (Hilbert 1917, 64).

1916/17 Lectures, 106–109),[196] sandwiched between discussions of his motivation for raising the question of the validity of Euclidean geometry and his attempts to answer it. At the end of the previous section he points out:

> We would like to anticipate the results of our calculation: in general our basic physical equations have no solutions at all. In my opinion, this is a positive result of the theory: since in no way are we able to impose Euclidean geometry on nature through a different interpretation of experiments. Assuming namely that my basic physical equations to be developed are really correct, then no other physics is possible, i.e., reality cannot be understood in a different way.[197]

Hilbert evidently thought he had found a powerful argument against geometric conventionalism—presumably, he had Poincaré in mind here. He continues:

> On the other hand we shall see that under certain very specialized assumptions—perhaps the absence of matter throughout space is sufficient for this—the only solution to the differential equations are $g_{\mu\nu} = \delta_{\mu\nu}$ [the Minkowski metric].[198]

At this point, the problem of the status of geometry is broadened from three-dimensional geometry to four dimensional pseudo-geometry—and in particular the question of the status of Euclidean geometry is broadened to that of four-dimensional Minkowski pseudo-geometry. In this form, it plays a central role in Hilbert's thinking about his program. This problem, rooted as it was in a mathematical tradition going back to Gauss, led him naturally to consider what we call the empty-space Einstein field equations. He hoped that the absence of matter and non-gravitational fields might suffice to uniquely single out the Minkowski metric as a solution to his field equations (which are identical to Einstein's in this case):

> It is possible that the following theorem is correct:
> Theorem: If one removes all electricity from the world (i.e. $q_i \equiv 0$) and demands absolute regularity—i.e. the possibility of expansion in a power series—of the gravitational potentials $g_{\mu\nu}$ (a requirement that in our opinion must always be fulfilled, even in the general case), then Euclidean geometry prevails in the world, i.e. the 10 equations (3) [equation number in the original; the vanishing of the Einstein tensor] have $g_{\mu\nu} = \delta_{\mu\nu}$ as their only solution.[199]

(He explains what he means by "regular" in his discussion of the Schwarzschild metric, considered below.) Of course, Hilbert was *not* able to establish this theorem, since it is not true, as Einstein's work on gravitational waves might already have sug-

196 Page 107 is missing from the typescript.

197 "Wir wollen das Resultat unserer Rechnung vorwegnehmen: unsere physikalischen Grundgleichungen haben im allgemeinen keineswegs Lösungen. Dies ist meiner Meinung nach ein positives Resultat der Theorie: denn wir können der Natur die Euklidischen Geometrie durch andere Deutung der Experimente durchaus nicht aufzwingen. Vorausgesetzt nämlich, dass meine zu entwickelnden physikalischen Grundgleichungen wirklich richtig sind, so ist auch keine andere Physik möglich, d.h., die Wirklichkeit kann nicht anders aufgefasst werden." (WS 1916/17 Lectures, 106)

198 "Andererseits werden wir sehen, dass unter gewissen sehr spezialisierenden Voraussetzungen—vielleicht ist das Fehlen von Materie im ganzen Raum dazu schon hinreichend—die einzige Lösungen der Diffentialgleichungen $g_{\mu\nu} = \delta_{\mu\nu}$ [the Minkowski metric] sind."

gested (Einstein 1916c). Nor was he able to find any other set of necessary and sufficient conditions for the uniqueness of the Minkowski metric; but he did almost establish one set of sufficient conditions and proved another:

> I consider the following theorem to be very probably correct: If one removes all electricity from the world and demands for the gravitational potential, apart from the self-evident requirement of regularity, that $g_{\mu\nu}$ is independent of t, i.e. that gravitation is static, and finally [one demands] also regular behavior at infinity, then $g_{\mu\nu} = \delta_{\mu\nu}$ are the only solutions to the gravitation equations (3)[equation number in the original].
> I can now already prove this much of the theorem, that in the neighborhood of Euclidean geometry there are certainly no solutions to these equations.[200]

This is, of course, the result that he did prove in Paper 2 (see above). The proof of this result for the full, non-linear field equations hung fire for a long time with several proofs for the case of static metrics being given over the years; the proof for stationary metrics was finally given by André Lichnerowicz in 1946.[201]

6) The Schwarzschild solution: The Schwarzschild solution had already been published (Schwarzschild 1916) and Hilbert dedicates considerable space to it, both in his lecture notes and in Paper 2. He uses it in the course of his effort to exploit the new tools of general relativity for addressing the foundational questions of geometry raised in the mathematical tradition. In his lecture notes, he introduces a number of assumptions on the metric tensor in order to prove a theorem on the uniqueness of Euclidean geometry:

> 1) Let $g_{\mu\nu}$ again be independent of t.
>
> 2) Let $(g_{\nu 4} = 0)$ $(\nu = 1, 2, 3)$ [interpolated by hand: "i.e. Gaussian coordinate system, which can always be introduced by a transformation"] (Orthogonality of the t- axis to the x_1, x_2, x_3 -space, the so-called metric space.)
>
> 3) There is a distinguished point in the world, with respect to which central symmetry holds, i.e. the rotation of the coordinate system around this point is a transformation of the world onto itself.

---

199 "Es ist möglich, dass folgender Satz richtig ist:
Satz: Nimmt man alle Elektrizität aus der Welt hinweg (d.h. $q_i \equiv 0$) und verlangt man absolute Regularität—d.h. Möglichkeit der Entwicklung in eine Potenzreihe—der Gravitationspotentiale $g_{\mu\nu}$ (eine Forderung, die nach unserer Auffassung auch im allgemeinen Fall immer erfüllt sein muss), so herrscht in der Welt die Euklidische Geometrie, d.h. die 10 Gleichungen (3) haben $g_{\mu\nu} = \delta_{\mu\nu}$ als einzige Lösung." (WS 1916/17 Lectures, 111–112)

200 "Für sehr wahrscheinlich richtig halte ich folgenden Satz:
Nimmt man alle Elektrizität aus der Welt fort und verlangt von den Gravitationspotentialen ausser der selbstverständlichen Forderung der Regularität noch, dass $g_{\mu\nu}$ von t unabhängig ist, d.h. dass die Gravitation stille steht, und schliesslich noch reguläres Verhalten im Unendlichen, so sind $g_{\mu\nu} = \delta_{\mu\nu}$ die einzigen Lösungen der Gravitationsgleichungen (3).
Von diesem Satz kann ich schon jetzt so viel beweisen, dass in der Nachbarschaft der Euklidischen Geometrie sicher keine Lösung dieser Gleichungen vorhanden sind." (WS 1916/17 Lectures, 112)

201 See (Lichnerowicz 1946).

> Now the following theorem holds: If the gravitational potentials fulfill conditions 1–3, then Euclidean geometry is the only solution to the basic physical equations.[202]

The proof of this theorem leads him to consider the problem of spherically-symmetric solutions to the empty-space Einstein field equations, a problem that Hilbert notes had previously been treated by Einstein (in the linear approximation) and Schwarzschild (exactly). He claims for his own calculations only that, compared to those of others, they are "auf ein Minimum reduziert" (WS 1916/17 Lectures, 113) by working from his variational principle for the field equations (see above). Hermann Weyl gave a similar variational derivation in 1917 (Weyl 1917); the section of his book *Raum-Zeit-Materie* on the Schwarzschild metric includes a reference to Hilbert's Paper 2, which reproduces Hilbert's variational derivation, (Weyl 1918a; 1918b, 230 n.9; 1923, 250 n.19). But Pauli's magisterial survey of the theory of relativity mentions only Weyl's paper, this probably contributing to the neglect of Hilbert's contribution in most later discussions (Pauli 1921).

In Paper 2, Hilbert derives the Schwarzschild metric from the same three assumptions as in the lecture notes, emphasizing that:

> In the following I present for this case a procedure that makes no assumptions about the gravitational potentials $g_{\mu\nu}$ at infinity, and which moreover offers advantages for my later investigations.[203]

In spite of this, many later derivations of the Schwarzschild metric still continue to impose unnecessary boundary conditions. But Hilbert did not show that the assumption of time-independence is also unnecessary, as proved by Birkhoff in 1923. (The assertion that the Schwarzschild solution is the only spherically symmetric solution to the empty-space Einstein equations is known as Birkhoff's theorem.)[204]

Hilbert's discussion of the Schwarzschild solution also raises the problem of its singularities and their relation to Hilbert's theory of matter. In his lecture notes, after establishing the Schwarzschild metric, he writes:

---

202 "1) Es sei wieder $g_{\mu\nu}$ unabhängig von t.
2) Es sei $g_{\nu 4} = 0 \qquad \nu = 1, 2, 3$ [interpolated by hand: "d.h. Gauss'sches Koordinatensystem, das durch Transformation immer eingeführt werden kann"] (Orthogonalität der t-Achse auf dem x_1, x_2, x_3 -Raum, dem sogenannten Streckenraum.)
3) Es gebe einen ausgezeichneten Punkt in der Welt, in Bezug auf welchen zentrische Symmetrie vorhanden sein soll, d.h. die Drehung des Koordinatensystems um diesen Punkt ist eine Transformation der Welt in sich.
Nun gilt folgender Satz:
Erfüllen die Gravitationspotentiale die Bedingungen 1–3, so ist die Euklidische Geometrie die einzige Lösung der physikalischen Grundgleichungen." (WS 1916/17 Lectures, 113)

203 "Ich gebe im Folgenden für diesen Fall einen Weg an, der über die Gravitationspotentiale $g_{\mu\nu}$ im Unendlichen keinerlei Voraussetzungen macht und ausserdem für meine späteren Untersuchungen Vorteile bietet." (Hilbert 1917, 67) For the derivation, see pp. 67–70.

204 See (Birkhoff 1923, 253–256).

> According to our conception of the nature of matter, we can only consider those $g_{\mu\nu}$ to be physically viable solutions to the differential equations $K_{\mu\nu} = 0$ [the Einstein equations] that are regular and singularity free.
> We call a gravitational field or a metric "regular"—this definition had to be added—when it is possible to introduce a coordinate system, such that the functions $g_{\mu\nu}$ are regular and have a non-zero determinant at every point in the world. Furthermore, we describe a single function as being regular if it and all its derivatives are finite and continuous. This is incidentally always the definition of regularity in physics, whereas in mathematics a regular function is required to be analytic.[205]

It is curious that Hilbert identifies physical regularity with *infinite* differentiability and continuity of all derivatives. Either of these requirements is much too strong: each precludes gravitational radiation carrying new information, for example gravitational shock waves.[206] But at least Hilbert attempted to define a singularity of the gravitational field. In his understanding, the Schwarzschild solution has singularities at $r = 0$ and at the Schwarzschild radius. But we now know the first singularity is real, while the second can be removed by a coordinate transformation. He remarks:

> When we consider that these singularities are due to the presence of a mass, then it also seems plausible that they cannot be eliminated by coordinate transformations. However, we will give a rigorous proof of this later by examining the behavior of geodesic lines in the vicinity of this point.[207]

Hilbert then returns to his original motif: the Schwarzschild solution as a tool for discussing foundational problems of geometry:

> In order to obtain singularity-free solutions, we must assume that a [i.e., the mass parameter] = 0. [This leads to the Minkowski metric.] ... This proves the ... theorem: In the absence of matter, under the stated assumptions 1–3 [see above], the pseudo-Euclidean geometry of the little relativity principle [i.e., special relativity] actually holds in physics; and for t = const Euclidean geometry is in fact realized in the world.[208]

---

205 "Nach unserer Auffassung vom Wesen der Materie können wir als physikalisch realisierbare Lösungen $g_{\mu\nu}$ der Differentialgleichungen $K_{\mu\nu} = 0$ [the Einstein equations] nur diejenigen ansehen, welche regulär und singularitätenfrei sind.
"Regulär" nennen wir ein Gravitationsfeld oder eine Massbestimmung,— diese Definition war noch nachzutragen—wenn es möglich ist, ein solches Koordinatensystem einzuführen, dass die Funktionen $g_{\mu\nu}$ an jeder Stelle der Welt regulär sind und eine von null verschiedenen Determinante haben. Wir bezeichnen ferner eine einzelne Funktion als regulär, wenn sie mit allen ihren Ableitungen endlich und stetig ist. Dies ist übrigens immer die Definition der Regularität in der Physik, während in der Mathematik von einer regulären Funktion verlangt wird, dass sie analytisch ist." (WS 1916/17 Lectures, 118)

206 See, e.g., (Papapetrou 1974, 169–177).

207 "Wenn wir bedenken, dass diese Singularitäten von der Anwesenheit einer Masse herrühren, so erscheint es auch plausibel, dass dieselben durch Koordinatentransformation nicht zu beseitigen sind. Einen strengen Beweis dafür werden wir aber erst weiter unten geben, indem wir den Verlauf der geodätischen Linien in der Umgebung dieser Punkt untersuchen." (WS 1916/17 Lectures, 118–119)

208 "Wir müssen also, um singularitätenfreie Lösungen zu erhalten, a [i.e., the mass parameter] = 0 annehmen. Wir haben damit den ... Satz bewiesen: Bei Abwesenheit von Materie ($q_i = 0$) existiert unter den ... genannten Voraussetzungen 1–3 [see above] die pseudoeuklidischen Geometrie des kleinen Relativitätsprinzips in der Physik tatsächlich, und für t = const ist in der Welt die Euklidische Geometrie wirklich realisiert." (WS 1916/17 Lectures, 119)

In the sequel, Hilbert explores its physical significance for describing the behavior of matter in space and time. His conception of matter, based on Mie's theory, plays no significant role in this discussion, its role being taken instead by assumptions that Hilbert assimilated from Einstein's work, such as the geodesic postulate for the motion of free particles.

He then turns to the justification for considering the case $a \neq 0$:

> Then we are acting against our own prescription that we shall regard only singularity-free gravitational fields as realizable in nature. Hence we must justify the assumption $a \neq 0$.[209]

He emphasizes the extraordinary difficulty of integrating the 14 field equations, even for "the simple special case when they go over to $K_{\mu\nu} = 0$":

> Mathematical difficulties already hinder us, for example, from constructing a single neutral mass point. If we were able to construct such a neutral mass, and if its behavior in the neighborhood of this point were known, then, if we let the neutral mass degenerate increasingly to a mass point, the $g_{\mu\nu}$ at this point would display a singularity. Such a singularity we would have to regard as being allowed in the sense that the $g_{\mu\nu}$ outside the immediate neighborhood of the singularity correctly describes the course actually realized in nature. In [the Schwarzschild line element] we must now have this kind of singularity at hand. Incidentally, we can now state that the construction of a neutral mass point, even if this is possible later, will prove to be so complicated that for purposes, in which one does not look at the immediate neighborhood of the mass point, one will be able to calculate the approximately correct gravitational potentials containing a singularity with sufficient precision.
>
> We now maintain the following: If we could actually carry out the mathematical expansion leading to construction of a neutral massive particle, we would probably find laws that, for the time being, still must be formulated axiomatically; but which later will emerge as consequences of our general theory, consequences that admittedly only can be proven categorically by means of a broad-ranging theory and complex calculations. These axioms, which thus have only provisional significance, we formulate as follows:
>
> Axiom I.: The motion of a mass point in the gravitational field is represented by a geodesic line that is a time-like.
>
> Axiom II: The motion of light in the gravitational field is represented by a null geodesic curve.
>
> Axiom III.: A singular point of the metric is equivalent to a gravitational center.[210]

Hilbert calls the first two axioms, taken from Einstein's work, a "rational generalization" of the behavior of massive particles and light rays in the "old physics," in which the metric tensor takes the limiting Minkowski values. He states that the Newtonian law of gravitational attraction and the resulting Keplerian laws of planetary

---

209 "Dann handeln wir zwar entgegen unserer eigenen Vorschrift, dass wir nur singularitätenfreie Gravitationsfelder als in der Natur realisierbar ansehen wollen. Daher müssen wir die Annahme $a \neq 0$ rechtfertigen." (WS 1916/17 Lectures, 120)

motion follow from these axioms "in the first approximation." In this way, Hilbert integrated into his theory the essential physical elements, on which Einstein's path to general relativity was based. Even his epistemological justification for the superiority of the new theory now makes use of an argument for the integration of knowledge. Remarkably, from Hilbert's perspective, this integration not only involves knowledge of classical physics such as Newton's law of gravitation, but also of Euclidean geometry as a physical interpretation of space:

> In principle, however, this new Einsteinian law has no similarity to the Newtonian. It is infinitely more complicated than the latter. If we nevertheless prefer it to the Newtonian, this is because this law satisfies a profound philosophical principle—that of general invariance—and that it contains as special cases two such heterogeneous things as on the one hand, Newton's law and on the other, the actual validity of Euclidian geometry in physics under certain simple conditions; so that we do not have to, as was the case until now, first assume the validity of Euclidian geometry and then put together a law of attraction.[211]

Thus we see that Hilbert considers his results on the conditions of validity of Euclidean geometry on a par in importance with, and logically prior to, Einstein's and Schwarzschild's results on the Newtonian limit of general relativity.

In accord with the physical interpretation they are given in Axioms I and II, Hilbert then goes on to study the time-like and null geodesics of the Schwarzschild metric, leading to discussions of two general-relativistic effects that Einstein had already considered: the planetary perihelion precession and the deflection of light due to the Sun's gravitational field. This discussion occupies almost all of the rest of this chapter of his lecture notes (WS 1916/17 Lectures, 122–156). After a short discussion of the

---

210 "Die mathematischen Schwierigkeiten hindern uns z.B. schon an der Konstruktion eines einzigen neutralen Massenpunktes. Könnten wir eine solche neutrale Masse konstruieren, und würden wir den Verlauf in der Umgebung dieser Stelle kennen, so würden die $g_{\mu\nu}$ wenn wir die neutrale Masse immer mehr gegen einen Massenpunkt hin degenerieren lassen, in diesem Punkte eine Singularität aufweisen. Eine solche müssten wir als erlaubt ansehen in dem Sinne, dass die $g_{\mu\nu}$ ausserhalb der nächsten Umgebung der Singularität den in der Natur wirklich realisierten Verlauf richtig wiedergeben. Eine solche Singularität müssen wir nun in [the Schwarzschild line element] vor uns haben. Im übrigen können wir schon jetzt sagen, dass die Konstruktion eines neutralen Massenpunktes, auch wenn sie später möglich sein wird, sich als so kompliziert erweisen wird, dass man für die Zwecke, in denen man nicht die nächste Umgebung des Massenpunktes betrachtet, mit ausreichender Genauigkeit mit den mit einer Singularität behafteten, angenähert richtigen Gravitationspotentialen wird rechnen können.

Wir behaupten nun Folgendes: Wenn wir die mathematische Entwicklung, die zur Konstruktion eines neutralen Massenteilchens führt, wirklich durchführen können, so werden wir dabei vermutlich auf Gesetze stossen, die wir einstweilen noch axiomatisch formulieren müssen, die aber später sich als Folgen unserer allgemeinen Theorie ergeben werden, als Folgen freilich, die bestimmt nur durch eine weitsichtige Theorie und komplizierte Rechnung zu begründen sein werden. Diese Axiome, die also nur provisorische Geltung haben sollen, fassen wir folgendermassen:

Axiom I: Die Bewegung eines Massenpunktes im Gravitationsfeld wird durch eine geodätische Linie dargestellt, welche eine Zeitlinie ist.

Axiom II: Die Lichtbewegung im Gravitationsfeld wird durch eine geodätische Nulllinie dargestellt.

Axiom III: Eine singuläre Stelle der Massbestimmung ist äquivalent einem Gravitationszentrum." (WS 1916/17 Lectures, 120–121) dimensions of various physical quantities (WS 1916/17 Lectures, 156–158), he discusses the behavior of measuring threads and clocks in the Schwarzschild gravitational field (WS 1916/17 Lectures, 159–163), and concludes the chapter with a discussion of the third general-relativistic effect treated by Einstein, the gravitational redshift of spectral lines (WS 1916/17 Lectures, 163–166).

In Paper 2, these topics are treated more briefly if at all: Axioms I and II and their motivations, are discussed on pp. 70–71. The discussion of time-like geodesics occupies pp. 71–75, and the paper closes with a discussion of null geodesics on pp. 75–76. In summary, this paper must be considered a singular hybrid between the blossoming of a rich mathematical tradition that Hilbert brings to bear on the problems of general relativity, and the agony of facing the collapse of his own research program.

6.4 Revisions of Paper 2

Paper 2, like Paper 1, was republished twice: Indeed, the two were combined in the 1924 version, Paper 2 becoming Part 2 of *Die Grundlagen der Physik* (Hilbert 1924, 11–32). We shall refer to this version as "Part 2." The reprint of Hilbert 1924 in the *Gesammelte Abhandlungen* was edited by others, presumably under Hilbert's supervision (Hilbert 1935, 268–289). We shall refer to this version as "Part 2–GA." Compared to Paper 1, Hilbert's additions and corrections to Paper 2 are less substantial, as is to be expected since Paper 2 was written largely within the context of general relativity. Most changes are minor improvements, e.g. in connection with recent literature on the theory. There are three significant changes however. One, introduced by Hilbert at the beginning of Part 2, concerns Hilbert's view of the relation between Papers 1 and 2, the other two by the editors of the *Gesammelte Abhandlungen* in Part 2–GA. The second concerns the Cauchy problem, and the third concerns his understanding of invariant assertions. We shall discuss these revisions, both major and minor.

The first significant change concerns the paper's goal: Paper 2 states that "it seems necessary to discuss some more general questions of a logical as well as physical nature" ("erscheint es nötig, einige allgemeinere Fragen sowohl logischer wie physikalischer Natur zu erörtern" Hilbert 1917, 53). Part 2 states: "now the relation of the theory with experience shall be discussed more closely" ("Es soll nun der Zusammenhang der Theorie mit der Erfahrung näher erörtert werden" Hilbert 1924, 11). This revision confirms our interpretation of Paper 2 as resulting, in its original

211 "Prinzipiell aber hat dieses neue Einsteinsche Gesetz gar keine Ähnlichkeit mit dem Newtonschen. Es ist unmöglich komplizierter als das letztere. Wenn wir es trotzdem dem Newtonschen vorziehen, so ist dies darin begründet, dass dieses Gesetz einem tiefliegenden philosophischen Prinzip — dem der allgemeinen Invarianz — genüge leistet, und dass es zwei so heterogene Dinge, wie das Newtonsche Gesetz einerseits und die tatsächliche Gültigkeit der Euklidischen Geometrie in der Physik unter gewissen einfachen Voraussetzungen andererseits als Spezialfälle enthält, sodass wir also nicht, wie dies bis jetzt der Fall war, zuerst die Gültigkeit der Euklidischen Geometrie voraussetzen, und dann ein Attraktionsgesetz anflicken müssen." (WS 1916/17 Lectures, 122)

version, from the tension between Hilbert's concern about the unsolved problems of his theory, in particular the problem of causality, and his immersion in the challenging applications of general relativity, in particular to astronomy. Since Hilbert's revision of Paper 1 had effectively transformed his theory into a version of general relativity, the revision of Paper 2 could now be presented as relating this theory to its empirical basis, the astronomical problems being addressed by contemporary general relativity.

We shall now discuss the changes, which occur in four of the six topics discussed (see above):

1. The metric tensor and its measurement: Part 2 drops all reference to "Messfaden." The discussion of measurement is based entirely on the "Lichtuhr," but is otherwise parallel to that in Paper 2 (Hilbert 1924, 11–13).
2. The causality problem for the field equations (Hilbert 1924, 16–19): There are several changes in the discussion. The wording, with which Hilbert introduces the problem now reads:

   > Our basic equations of physics [the gravitational and the electromagnetic field equations] in no way take the form characterized above [Cauchy normal form]: rather four of them are, as I have shown, a consequence of the rest ...[212]

   Note that "wie ich gezeigt habe" replaces "nach Theorem I" (see p. 59 of Paper 2). Hilbert says that, if there were 4 additional invariant equations, then the system of equations in Gaussian normal coordinates "ein überbestimmtes System bilden würde" (see p. 16 of Part 2) replacing "untereinander in Widerspruch ständen" (see p. 60 of Paper 2).

   In the discussion of the first way, in which "physically meaningful, i.e., invariant assertions can be expressed mathematically" (Hilbert 1917, 62; 1924, 18), he corrects a number of the equations in his example. His discussion of the third way is shortened considerably, now reading:

   > An assertion is also invariant and is therefore always physically meaningful if it is valid for any arbitrary coordinate system, without the need for the expressions occurring in it to possess a formally invariant character.[213]

   In Paper 2, this sentence had ended with "...gültig sein soll," and the paragraph had given the example of Einstein's gravitational energy-momentum complex.
3. Euclidean geometry: His discussion is the same, except that the discussion of gravitational perturbations drops the use of an imaginary time coordinate and Euclidean metric (Hilbert 1924, 19–23, 26).

---

212 "Unsere Grundgleichungen der Physik sind nun keineswegs von der oben charakterisierten Art; vielmehr sind, wie ich gezeigt habe, vier von ihnen eine Folge der übrigen ..." (Hilbert 1924, 16)

213 "Auch ist eine Aussage invariant und hat daher stets physikalischen Sinn, wenn sie für jedes beliebige Koordinatensystem gültig ist, ohne daß dabei die auftretenden Ausdrücke formal invarianten Charakter zu besitzen brauchen." (Hilbert 1924, 19)

4. The Schwarzschild solution (Hilbert 1924, 23–32): He adds a footnote to the light ray axiom:

> Laue has shown for the special case $L = \alpha Q$ [i.e., for the usual Maxwell Lagrangian] how this theorem can be derived from the electrodynamic equations by considering the limiting case of zero wavelength.[214]

   followed by a reference to Laue's 1920 paper (Laue 1920) showing that Hilbert kept up with the relativity literature. He also dropped a rather trivial footnote to Axiom I (massive particles follow time-like world lines):

> This last restrictive addition [i.e., "Zeitlinie"] is to be found neither in Einstein nor in Schwarzschild.[215]

   He adds a more careful discussion of circular geodesics, the radius of which equals the Schwarzschild radius (Hilbert 1924, 30, compared to 1917, 75), but otherwise the discussion of geodesics remains the same.

When the 1924 version of his two papers was republished in 1935 in his *Gesammelte Abhandlungen*, the editors introduced two extremely significant changes, as well as more trivial ones that we shall not discuss, that retract the last elements of Hilbert's attempt to provide a solution to the causality problem for his theory. These changes in Part 2–GA are footnotes marked "Anm[erkung] d[er] H[erausgeber]". The first occurs in the discussion of the causality principle for generally-covariant field equations (Hilbert 1924, 18–19; 1935, 275–277). The sentence:

> Since the Gaussian coordinate system itself is uniquely determined, therefore also all assertions with respect to these coordinates about those potentials (24) [equation number in the original] are of invariant character.[216]

is dropped; and a lengthy footnote is added (Hilbert 1935, 275–277). This footnote shows that the editors[217] correctly understood the nature of the fourteen field equations. Six of the ten gravitational and three of the four electromagnetic equations contain second time derivatives of the six spatial components of the metric tensor and three spatial components of the electromagnetic potentials. Thus, their values together with those of their first time derivatives on the initial hypersurface determine their evolution off that hypersurface. But these initial values are subject to constraints, set by the remaining four gravitational and one electromagnetic equation, which contain no second time derivative. Due to the differential identities satisfied

214 "Laue hat für den Spezialfall $L = \alpha Q$ [i.e., for the usual Maxwell Lagrangian] gezeigt, wie man diesen Satz aus den elektrodynamischen Gleichungen durch Grenzübergang zur Wellenlänge Null ableiten kann." (Hilbert 1924, 27).

215 "Dieser letzte einschränkende Zusatz findet sich weder bei Einstein noch bei Schwarzschild." (Hilbert 1917, 71)

216 "Da das Gaußische Koordinatensystem selbst eindeutig festgelegt ist, so sind auch alle auf dieses Koordinatensystem bezogenen Aussagen über jene Potentiale (24) von invariantem Charakter." (Hilbert 1924, 18)

217 Paul Bernays, Otto Blumenthal, Ernst Hellinger, Adolf Kratzer, Arnold Schmidt, and Helmut Ulm.

by the field equations, if these constraint equations hold initially, they will continue to hold by virtue of the remaining field equations. This footnote culminates in the statement:

> Thus causal lawfulness does not express the full content of the basic equations; rather, in addition to this lawfulness, these equations also yield *restrictive conditions on the respective initial state*.[218]

The editors also explain that, in the gauge-invariant electromagnetic case, it is only the fields and not the potentials that are determined by the field equations. The editors' addition thus presents a lucid account of the Cauchy problem in general relativity, and shows that Hilbert's attempt to formulate a principle of causality for his theory in terms of the classical notion of initial data (i.e. values that can be freely chosen at any given moment in time, which then determine their future evolution) had not taken into account the existence of constraints on the initial data.

The second footnote occurs in the discussion of how to satisfy the requirement that physically meaningful assertions be invariant by use of an invariant coordinate system (Hilbert 1924, 18–19). The footnote, which actually undermines claims in Hilbert's paper, reads:

> In the case of each of the three types of preferred coordinate systems named here, there is only a partial fixation of the coordinates. The Gaussian nature of a coordinate system is preserved by arbitrary transformations of the space coordinates and by Lorentz transformations, and a coordinate system in which the vector r^k has the components $(0, 0, 0, 1)$, is transformed into another such system by an arbitrary transformation of the spatial coordinates together with a spatially varying shift of the temporal origin.
>
> The characterization of a Gaussian coordinate system by conditions (23) [equation number in the original] and likewise that of the third-named preferred coordinate system through the conditions for r^k is in fact not completely invariant insofar as the specification of the fourth coordinate—introduced through conditions (21) [equation number in the original; the conditions for a "proper" coordinate system]—plays a role in it.[219]

218 "Somit bringt die kausale Gesetzlichkeit nicht den vollen Inhalt der Grundgleichungen zum Ausdruck, diese liefern vielmehr außer jener Gesetzlichkeit noch *einschränkende Bedingungen für den jeweiligen Anfangszustand*." (Hilbert 1935, 277)

219 "Bei den drei hier genannten Arten von ausgezeichneten Koordinatensystemen handelt es sich jedesmal nur um eine partielle Festlegung der Koordinaten. Die Eigenschaft des Gaußischen Koordinatensystems bleibt erhalten bei beliebigen Transformationen der Raumkoordinaten und bei Lorentztransformationen, und ein Koordinatensystem, in welchem der Vektor r^k die Komponenten $(0, 0, 0, 1)$ hat, geht wieder in ein solches über bei einer beliebigen Transformation der Raumkoordinaten nebst einer örtlich variablen Verlegung des zeitlichen Nullpunktes.
Die Charakterisierung des Gaußischen Koordinatensystems durch die Bedingungen (23) und ebenso die des drittgenannten ausgezeichneten Koordinatensystems durch die Bedingungen für r^k ist übrigens insofern nicht völlig invariant, als darin die Auszeichnung der vierten Koordinate zur Geltung kommt, die mit der Aufstellung der Bedingungen (21) eingeführt wurde." (Hilbert 1935, 277)

The editors of Hilbert's papers corrected two major mathematical errors that survived his own revision of Paper 2, and since he was still active when this edition of his papers was published, it can be assumed that these changes were made with his consent, if not participation.

7. THE FADING AWAY OF HILBERT'S POINT OF VIEW AND ITS SUBSUMPTION BY EINSTEIN'S PROGRAM

Early on, Einstein and Weyl set the tone for the way in which Hilbert's papers on the *Foundations of Physics* were integrated into the mainstream of research in physics and mathematics. Not only did the articles by Einstein and Weyl receive immediate attention when first published in the *Sitzungsberichte* of the Prussian Academy of Sciences, but they were soon incorporated into successive editions of *Das Relativitätsprinzip*, then the standard collection of original works on the development of relativity.[220] Three out of four of Einstein's works added to the third edition mention Hilbert, as does Weyl's contribution to the fourth edition—although, as we shall see, the latter's omissions are as significant as his attributions. Translated into French, English and other languages, and in print to this day, countless scholars had their impression of the scope and history of relativity shaped by this book.

First we shall discuss Einstein's two mentions of Hilbert in 1916. (His third in 1919 is related to Weyl's 1918 paper, so we shall discuss it afterwards.) In contrast with Hilbert's need to reorganize his theory in reaction to Einstein's work, Einstein could assimilate Hilbert's results into the framework of general relativity without being bothered by the latter's differing interpretation of them. This assimilation, in turn, assigned Hilbert a place in the history of general relativity.

Einstein's 1916 review paper on general relativity mentions Hilbert in a discussion of the relation between the conservation identities for the gravitational field equations and the field equations for matter:

> Thus the field equations of gravitation contain four conditions [the conservation equations for the energy-momentum tensor of matter] which govern the course of material phenomena. They give the equations of material processes completely of the latter are capable of being characterized by four independent differential equations.[221]

220 See (Blumenthal 1913; 1919; 1923; 1974). All editions were edited by the mathematician Otto Blumenthal. The first edition appeared as the second volume of his series *Fortschritte der Mathematischen Wissenschaften in Monographien* (the first being a collection of Minkowski's papers on electrodynamics), "als eine Sammlung von Urkunden zur Geschichte des Relativitätsprinzips" ("Vorwort" [n.p.]). The third edition in 1919 included additional papers by Einstein on general relativity, the fourth edition added Weyl's first paper on his unified theory of gravitation and electromagnetism. The fifth edition in 1923 is the basis of the editions currently in print, and of the translations into other languages. It would be interesting to know how Blumenthal chose the papers to include in what became the canonical source book on relativity.

A footnote adds a reference to Paper 1.[222] Thus, Einstein subsumed into the general theory of relativity, as a particular case of an important general result, what Hilbert regarded as an outstanding achievement of his theory. Hilbert's interpretation of this result as embodying a unique coupling between gravitation and electromagnetism, is not even mentioned.

In the same year, Einstein published his own derivation of the generally-covariant gravitational field equations from a variational principle. While in the 1916 review paper he had given a non-invariant "Hamiltonian" (= Lagrangian) for the field equations modulo the coordinate condition $\sqrt{-g} = 1$, he now proceeded in a manner reminiscent of Hilbert's in Paper 1. He uses the same gravitational variables (the $g_{\mu\nu}$ and their first and second derivatives), but Einstein's $q_{(\rho)}$ "describe matter (including the electromagnetic field" ("beschreiben die Materie (inklusive elektromagnetisches Feld)") and hence are arbitrary in number and have unspecified tensorial transformation properties. By his straightforward generalization, Einstein transformed Hilbert's variational derivation into a contribution to general relativity, without adopting the latter's perspective on this derivation as providing a synthesis between gravitation and a specific theory of matter. Rather, Einstein's generalization made it possible to regard Hilbert's theory as no more than a special case.

Einstein prefaced his calculations with some observations placing his work in context:

> H. A. Lorentz and D. Hilbert have recently succeeded [footnoted references to Lorentz's four papers of 1915–1916 and Hilbert's Paper 1] in presenting the theory of general relativity in a particularly comprehensive form by deriving its equations from a single variational principle. The same shall be done in this paper. My aim here is to present the fundamental connections as transparently and comprehensively as the principle of general relativity allows. In contrast to Hilbert's presentation, I shall make as few assumptions about the constitution of matter as possible.[223]

---

221 "The Foundation of the General Theory of Relativity" p. 810, in (CPAE 6E, Doc. 30, 187). "Die Feldgleichungen der Gravitation enthalten also gleichzeitig vier Bedingungen [the conservation equations for the energy-momentum tensor of matter], welchen der materielle Vorgang zu genügen hat. Sie liefern die Gleichungen des materiellen Vorganges vollständig, wenn letzterer durch vier voneinander unabhängige Differentialgleichungen charakterisierbar ist." (Einstein 1916a, 810)

222 The reference to "p. 3," is probably to a separately paginated off-print; see the discussion in (Sauer 1999).

223 Einstein, "Hamilton's Principle and the General Theory of Relativity" Sitzungsberichte 1916, 1111–1116, citation from p. 1111, in (CPAE 6E, Doc. 41, 240). "In letzter Zeit ist es H. A. Lorentz und D. Hilbert gelungen [footnoted references to Lorentz's four papers of 1915–1916 and Hilbert's Paper 1], der allgemeinen Relativitätstheorie dadurch eine besonders übersichtliche Gestalt zu geben, daß sie deren Gleichungen aus einem einzigen Variationsprinzipe ableiteten. Dies soll auch in der nachfolgenden Abhandlung geschehen. Dabei ist es mein Ziel, die fundamentalen Zusammenhänge möglichst durchsichtig und so allgemein darzustellen, als es der Gesichtspunkt der allgemeinen Relativität zuläßt. Insbesondere sollen über die Konstitution der Materie möglichst wenig spezialisierende Annahmen gemacht werden, im Gegensatz besonders zur Hilbertschen Darstellung." (Einstein 1916b, 1111)

Thus Einstein both gave Hilbert credit for his accomplishments and circumscribed their nature: Like Lorentz, Hilbert was supposedly looking for a variational derivation of the general-relativistic field equations, but included assumptions about the constitution of matter that were too special. In an earlier, unpublished draft, Einstein's tone was even sharper:

> Hilbert, following the assumption introduced by Mie that the H function depends on the components of a four-vector and their first derivatives, I do not consider very promising.[224]

In private correspondence, he was still more harsh, but also gave his reasons for disregarding Hilbert's point of view:

> Hilbert's assumption about matter appears childish to me, in the sense of a child who knows none of the perfidy of the world outside. [...] At all events, mixing the solid considerations originating from the relativity postulate with such bold, unfounded hypotheses about the structure of the electron or matter cannot be sanctioned. I gladly admit that the search for a suitable hypothesis, or Hamilton function, for the construction of the electron, is one of the most important tasks of theory today. The "axiomatic method" can be of little use here, though.[225]

Evidently, Einstein clearly perceived the diverse status of the physical assumptions underlying general relativity, on the one hand, and Hilbert's theory, on the other. From Einstein's point of view, Hilbert's detailed results, such as his variational derivation of the Schwarschild metric could be—and were—acknowledged as contributions to the development of general relativity, without any need to refer to the grandiose program, within which Hilbert had originally placed them.

In view of his own claims in this regard, one might expect Hilbert's work to have played a prominent role in the developing search for a unified field theory.[226] But his fate was that of a transitional figure, eclipsed by both his predecessors and his successors. His achievements were perceived as individual contributions to general relativity rather than as genuine milestones on the way towards a unified field theory. Evidently, this "mixed score" was the price Hilbert had to pay for being made one of the founding fathers of general relativity.

In his first contribution to unified field theory, Weyl assigned a definite place to Hilbert, if largely by omission. After presenting his generalization of Riemannian

---

224 "Die von Hilbert im Anschluss an Mie eingeführte Voraussetzung, dass sich die Funktion H durch die Komponenten eines Vierervektors q_ρ und dessen erste Ableitungen darstellen lasse, halte ich für wenig aussichtsvoll." See note 3 to Doc. 31 in (CPAE 6, 346).

225 "Der Hilbertsche Ansatz für die Materie erscheint mir kindlich, im Sinne des Kindes, das keine Tükken der Aussenwelt kennt. [...] Jedenfalls ist es nicht zu billigen, wenn die soliden Überlegungen, die aus dem Relativitätspostulat stammen, mit so gewagten, unbegründeten Hypothesen über den Bau des Elektrons bezw. der Materie verquickt werden. Gerne gestehe ich, dass das Aufsuchen der *geeigneten* Hypothese bezw. Hamilton'schen Funktion für die Konstruktion des Elektrons eine der wichtigsten heutigen Aufgaben der Theorie bildet. Aber die "axiomatische Methode" kann dabei wenig nützen." Einstein to Hermann Weyl, 23 November 1916, (CPAE 8, 365–366).

226 For a historical discussion, see (Majer and Sauer 2005; Goenner 2004).

geometry to include what he called "gauge invariance" (Eichinvarianz),[227] Weyl turned to unified field theory:

> Making the transition from geometry to physics, we must assume, in accord with the example of Mie's theory [references to Mie's papers of 1912/13 and Weyl's recently-published *Raum-Zeit-Materie*], that the entire lawfulness of nature is based upon a certain integral invariant, the action
>
> $$\int W d\omega = \int \mathfrak{W} dx \qquad (\mathfrak{W} = W\sqrt{g}),$$
>
> in such a way that the actual world is distinguished from all possible four-dimensional metric spaces, by the fact that the action contained in every region of the world takes an extremal value with respect to those variations of the potentials g_{ij}, ϕ_i that vanish at the boundaries of the region in question.[228]

In spite of its obvious relevance, there is no mention here of Hilbert. The sole mention comes in what we shall refer to as "the litany" since this or a similar list occurs so frequently in the subsequent literature:

> We shall show in fact, in the same way that, according to the investigations of Hilbert, Lorentz, Einstein, Klein and the author [reference follows to Paper 1 for Hilbert], the four conservation laws of matter (of the energy-momentum-tensor) are connected with the invariance of the action under coordinate transformations containing four arbitrary functions; the charge conservation law is linked to a newly introduced "scale-invariance" depending on a fifth arbitrary function.[229]

This passage, (incorrectly) attributing to Hilbert a clarification of energy-momentum conservation in general relativity and disregarding his attempt to create a unified field theory, makes his "mixed score" particularly evident. In a footnote added to the republication of his paper in *Das Relativitätsprinzip,* Weyl notes that:

> The problem of defining all invariants W admissible as actions, while requiring that they contain the derivatives of g_{ij} up to second order at most, and those of ϕ_i only up to first order, was solved by R. Weitzenböck [Weitzenböck 1920], [230]

---

227 This generalization was named a Weyl space by J.A. Schouten (see Schouten 1924).

228 "Von der Geometrie zur Physik übergehend, haben wir nach dem Vorbild der Mieschen Theorie anzunehmen, daß die gesamte Gesetzmäßigkeit der Natur auf einer bestimmten Integralinvariante, der Wirkungsgröße $\int W d\omega = \int \mathfrak{W} dx \qquad (\mathfrak{W} = W\sqrt{g})$ beruht, derart, daß die wirkliche Welt unter allen möglichen vierdimensionalen metrischen Räumen dadurch ausgezeichnet ist, daß für sie die in jedem Weltgebiet enthaltene Wirkungsgröße einen extremalen Wert annimmt gegenüber solchen Variationen der Potentiale g_{ik}, ϕ_i, welche an den Grenzen des betreffenden Weltgebiets verschwinden." (Weyl 1918c, 475)

229 "Wir werden nämlich zeigen: in der gleichen Weise, wie nach Untersuchungen von Hilbert, Lorentz, Einstein, Klein und dem Verf. [reference follows to Paper 1 for Hilbert] die vier Erhaltungssätze der Materie (des Energie-Impuls-Tensors) mit der, vier willkürliche Funktionen enthaltenden Invarianz der Wirkungsgröße gegen Koordinatentransformationen zusammenhängen, ist mit der hier neu hinzutretenden, eine fünfte willkürliche Funktion hereinbringenden "Maßstab-Invarianz" [...] das Gesetz von der Erhaltung der Elektrizität verbunden." (Weyl 1918c, 475)

without mentioning that this is the solution to the problem raised by Hilbert's ansatz for the invariant Lagrangian, first introduced in Paper 1. Little wonder that those whose knowledge of the history of relativity came from *Das Relativitätsprinzip* had no idea of Hilbert's original aims and little more of his achievements.

Hilbert fared a little better in Weyl's *Raum-Zeit-Materie,* the first treatise on general relativity (Weyl 1918a; 1918b; 1919; 1921; 1923).[231] The discussion of the energy-momentum tensor in the first edition (section 27) credits Hilbert with having shown that (Weyl 1918a; 1918b, 184):

> [...] Mie's electrodynamics can be generalized from the assumptions of the special to those of the general theory of relativity. This was done by Hilbert.[232]

Footnote 5 cites Paper 1 and adds (Weyl 1918a; 1918b, 230):

> The connection between Hamilton's function and the energy-momentum tensor is established here, and the gravitational equations articulated almost simultaneously with Einstein, if only within the confines of Mie's theory,[233]

Hilbert's work has already been subsumed under general relativity. Curiously, both textual reference and footnote disappear from all later editions (but see the discussion below of the fifth edition). Presumably because Weyl had already mentioned Hilbert, the latter's name does not appear in the litany in the first edition (footnote 6), listing those who had worked on the derivation of the energy-momentum conservation laws. By the third edition, Hilbert has been added to the litany (Weyl 1919, 266 n. 8), and remained there. In his discussion of causality for generally-covariant field equations in the first edition, Weyl credits Papers I and II (Weyl 1918a; 1918b, 190 and 230, n. 9); again, this note disappears from all later editions. Paper 2 is also cited in the first edition in connection with the Schwarzschild solution (Weyl 1918a; 1918b, 230, n. 15), and the introduction of geodesic normal coordinates (Weyl 1918a; 1918b, 230, n. 21).

The third edition carries over these references to Paper 2 and adds one in connection with linearized gravitational waves (Weyl 1919, 266, n. 14); and the fourth edition includes all these footnotes. Perhaps questions had been raised concerning

---

230 "Die Aufgabe, alle als Wirkungsgrößen zulässigen invarianten *W* zu bestimmen, wenn gefordert ist, daß sie die Ableitungen der g_{ik} höchstens bis zur 2., die der ϕ_i nur bis zur 1. Ordnung enthalten dürfen, wurde von R. Weitzenböck [Weitzenböck 1920] gelöst." (Blumenthal 1974, 159; translation from Lorentz et al. 1923.) This seventh edition from 1974 is an unchanged reprint of the fifth edition of 1923, 159, n. 2. Weitzenböck has his own version of the litany: "Die obersten physikalischen Gesetze: Feldgesetze und Erhaltungsätze werden nach den klassischen Arbeiten von Mie, Hilbert, Einstein, Klein und Weyl aus einem Variationsprinzip [...] hergeleitet"(p. 683). It is not clear why Lorentz is omitted from the litany; perhaps he was too much of a physicist for Weitzenböck.

231 The second edition of 1918 was unchanged, the fourth of 1921 was translated into English and French; the fifth of 1923, being thereafter reprinted without change.

232 "[...] die Miesche Elektrodynamik von den Voraussetzungen der speziellen auf die der allgemeinen Relativitätstheorie übertragen werden [kann]. Dies ist von Hilbert durchgeführt worden."

233 "Hier ist auch der Zusammenhang zwischen Hamiltonscher Funktion and Energie-Impuls-Tensor aufgestellt und wurden, etwa gleichzeitig mit Einstein, wenn auch nur im Rahmen der Mieschen Theorie, die Gravitationsgleichungen ausgesprochen."

Weyl's treatment of Hilbert in the book; at any rate, the footnote to the litany citing Hilbert in the fifth edition again credits him with a contribution to general relativity, rather than to unified field theories:

> In the first communication, Hilbert established the invariant field equations simultaneously with and independently of Einstein, but within the framework of Mie's hypothetical theory of matter.[234]

In short, in none of the editions is Hilbert mentioned in connection with unified field theories.

Pauli's standard 1921 review article on relativity is another major source, still consulted mainly in the English translation of 1958 (with additional notes) by physicists and mathematicians for historical and technical information about relativity and unified field theories (Pauli 1921; 1958). Pauli adopted what we may call the Einstein-Weyl line on Hilbert, considering him a somewhat unfortunate founding father of general relativity. After describing Einstein's work on general relativity culminating in the November 1915 breakthrough, Pauli adds in a footnote (Pauli 1921):[235]

> At the same time as Einstein, and independently, Hilbert formulated the generally covariant field equations [reference to Paper 1]. His presentation, though, would not seem to be acceptable to physicists, for two reasons. First, the existence of a variational principle is introduced as an axiom. Secondly, of more importance, the field equations are not derived for an arbitrary system of matter, but are specifically based on Mie's theory of matter ... .

His discussion of invariant variational principles in section 23 cites the litany: "investigations by Lorentz, Hilbert, Einstein, Weyl and Klein[236] on the role of Hamilton's Principle in the general theory of relativity" (Pauli 1921).[237]

Later (section 56), he discusses the question of causality in "a generally relativistic [i.e, generally-covariant] theory," arguing from general covariance to the existence of 4 identities between the 10 field equations, and concluding (Pauli 1921):[238]

> The contradiction with the causality principle is only apparent, since the many possible solutions of the field equations are only formally different. Physically they are completely equivalent. The situation described here was first recognized by Hilbert.

This passage represents a striking example of erroneously crediting Hilbert with a contribution to general relativity while neglecting his actual achievements. To make matters worse, Pauli's footnote cites Paper 1, rather than Paper 2; after also crediting Mach with a version of this insight, he adds (Pauli 1921):[239]

---

234 "In der 1. Mitteilung stellte Hilbert gleichzeitig und unabhängig von Einstein die invarianten Feldgleichungen auf, aber im Rahmen der hypothetischen Mieschen Theorie der Materie." (Weyl 1923, 329, n. 10)

235 Section 50, cited from translation in (Pauli 1958, 145 n. 277).

236 See Felix Klein to Wolfgang Pauli, 8 May 1921 in (Pauli 1979, 31).

237 Cited from translation in (Pauli 1958, 68).

238 Cited from translation in (Pauli 1958, 160).

239 Cited from translation in (Pauli 1958, 160, n. 315).

> Furthermore it deserves mentioning that Einstein had, for a time, held the erroneous view that one could deduce from the non-uniqueness of the solution that the gravitational equations could not be generally-covariant [reference to (Einstein 1914b)].

Pauli does acknowledge various contributions to general relativity in Paper 2.[240] But his discussion of unified field theories (Part V), like Weyl's, jumps from Mie (section 64) to Weyl (section 65) without mention of Hilbert.

By examining a couple of early treatises on relativity by non-German authors, we can get some idea of the propagation of the Einstein-Weyl line as canonized by Pauli. Jean Becquerel's *Le Principe de la Relativité et la Théorie de la Gravitation* was the first French treatise on general relativity. In Chapter 16 on "Le Principe d'Action Stationnaire," Becquerel asserts:

> Lorentz and Hilbert [references to Papers 1 and 2], and then Einstein succeeded in presenting the general equations of the theory of gravitation as consequences of a unique stationary action principle, …[241]

followed by section 103 on "Méthode de Lorentz et d'Hilbert" (Becquerel 1922, 257–262). Paper 2 is cited in connection with linearized gravitational waves (Becquerel 1922, 216), but there is no mention of Hilbert in Chapter 18 on "Union du Champ de Gravitation et du Champ Électromagnétique. Géometries de Weyl et d'Eddington" (Becquerel 1922, 309–335).

Until recently Eddington's treatise, *The Mathematical Theory of Relativity,* was widely read, cited and studied by students; and was translated into French and German (Eddington 1923; 1924). The two English editions cite Papers 1 and 2 in the bibliography, with a reference to section 61 on "A Property of Invariants,"[242] which demonstrates the theorem:[243]

> The Hamiltonian [i.e, Lagrangian] derivative of any fundamental invariant is a tensor whose divergence vanishes.

Outside the Bibliography, few references are given in the English editions; but Eddington added material to the German translation, including several references to Hilbert (Eddington 1925). On p. 114, footnote 1 credits Hilbert (Paper 2) with realizing that the assumption of asymptotic flatness is not needed in the derivation of the Schwarzschild metric. On p. 116, he credits Paper 2 for an "elegante Methode" for deducing the Christoffel symbols from the geodesic equation; and on p. 183, he credits the same paper for the first strict proof that one can always satisfy the linearized

---

240 See (Pauli 1921), section 13 for Axiom II; section 22 for discussion of the restrictions on coordinate systems if three coordinates are to be space-like and one time-like; and section 60 for the proof that linearized harmonic coordinate conditions may always be imposed.

241 "Lorentz et Hilbert [references to Papers 1 and 2], puis Einstein, ont réussi à presenter les équations générales de la theorie de la gravitation comme des conséquences d'un unique principe d'action stationnaire," (Becquerel 1922, 256).

242 See (Eddington 1924, 264): "wherever possible the subject matter is indicated by references to the sections in this book chiefly concerned."

243 See (Eddington 1924, 140–141).

harmonic coordinate conditions by an infinitesimal coordinate transformation. And that is it.

We see that, by the mid-1920s, and with minor variations within the accepted limits, the Einstein-Weyl line on Hilbert's role was already becoming standard in the literature on relativity.

8. AT THE END OF A ROYAL ROAD

The preceding discussion has shown that Hilbert did not discover a royal road to the field equations of general relativity. In fact, he did not formulate these equations at all but, at the end of 1915, developed a theory of gravitation and electromagnetism that is incompatible with Einstein's general relativity. Nevertheless, this theory can hardly be considered an achievement parallel to that of Einstein's creation of general relativity, to be judged by criteria independent of it. Not only is the dependence of Hilbert's theory on and similarity to Einstein's earlier, non-covariant Entwurf theory of gravitation too striking; but its contemporary reception as a contribution to general relativity and regardless of the extent to which Hilbert accepted the transformation of his theory into such a contribution, this is evidence of the theory's evanescent and heteronomous character. It could thus appear as if our account, in the end, describes a race for the formulation of a relativistic theory of gravitation with a clear winner—Einstein—and a clear loser—Hilbert. In contrast to the legend of Hilbert's royal road, such an account would bring us essentially back to Pauli's sober assessment of Hilbert's work as coming close to the formulation of general relativity but being faulted by its dependence on a specific theory of matter. However, as we have shown, this interpretation ascribes to Hilbert results in general relativity that he neither intended nor achieved, and ignores contributions that lay outside the scope of general relativity but were nevertheless crucial for its development. In view of such conundrums, we therefore propose not to consider the Einstein-Hilbert race as the competition between two individuals and their theories but as an event within a larger, collective process of knowledge integration.

As formulated by Einstein in 1915, general relativity incorporates elements of classical mechanics, electrodynamics, the special theory of relativity, and planetary astronomy, as well as such mathematical traditions as non-Euclidean geometry and the absolute differential calculus. It integrates these elements into a single, coherent conceptual framework centered around new concepts of space, time, inertia and gravitation. Without this enormous body of knowledge as its underpinning, it would be hard to explain the theory's impressive stability and powerful role even in today's physics. This integration was the result of an extended and conflict-laden process, to which not only Einstein but many other scientists contributed. From the point of view of historical epistemology, it was a collective process in an even deeper sense.[244] It involved a substantial, shared knowledge base, structured by fundamental concepts, models, heuristic etc., which were transmitted by social institutions, utilizing material representations, such as textbooks, and appropriated by individual learning processes. While individual thinking is governed to a large degree by these shared resources, it also affects and amplifies them, occasionally even changing these epistemic structures. On the basis of such an epistemology, which takes into account the interplay between shared knowledge resources and individual thinking, the emergence and fading away of a theory such as Hilbert's can be understood as an aspect of the process of integration of knowledge that produced general relativity.

To answer the question of from where alternative solutions (or attempted solutions) to the same problem come, we shall look at some of the shared knowledge of the time available for formulating theories such as those of Einstein and Hilbert. To explain the fading-away of Hilbert's theory, we then discuss the interplay between individual thinking and the knowledge resources that led to the formulation of general relativity and the transformation of Hilbert's theory into a contribution to it. It will become clear that, in both cases, the same mechanism was at work. In the case of general relativity, it integrated the various components of shared knowledge and resulted in the creation of a stable epistemic structure, which represents that integrated knowledge. In the case of Hilbert's theory, the same process disaggregated the various components of shared knowledge that had been brought together in a temporary structure, and rearranged and integrated them into a more stable structure.

The available knowledge offered a limited number of approaches to the problem that occupied both Einstein and Hilbert in late 1915: the formulation of differential equations governing the inertio-gravitational potential represented by the metric tensor. Two fundamentally different models underlying contemporary field theories of electrodynamics embodied the principal alternatives. One, the "monistic model," conceived all physical phenomena, including matter, in terms of fields. The other "fields-with-matter-as-source model" (or "Lorentz model") was based on a dualism of fields and matter. The first model was the basis for attempts to formulate an "electromagnetic world picture," which remained fragmentary and never succeeded in accounting for most contemporary physical knowledge. The second model was the basis for Lorentz's formulation of electron theory, the epitome of classical electrodynamics, in which matter acts as source for electrodynamic fields that, in turn, affect the motion of material bodies. Rather than attempting to reduce classical mechanical concepts to electrodynamic field concepts, the task associated with the electrodynamic world picture, Lorentz's electron theory successfully integrated electromag-

---

244 See (Csikszentmihalyi 1988): "All of the definitions [of creativity] ... of which I am aware assume that the phenomenon exists... either inside the person or in the work produced... After studying creativity for almost a quarter of a century, I have come to the reluctant conclusion that this is not the case. We cannot study creativity by isolating individuals and their works from the social and historical milieu in which their actions are carried out. This is because what we call creative is never the result of individual actions alone; it is the product of three main shaping forces: a set of social institutions or *field*, that selects from the variations produced by individuals those that are worth preserving; a stable cultural *domain* that will preserve and transmit the selected new ideas or forms to the following generations; and finally the *individual*, who brings about some change in the domain, a change that the field will consider to be creative." This concept is further discussed in (Stachel 1994).

netic and classical mechanical phenomena. The first model became the core of Hilbert's approach in an attempt to create a unified field theory, while Einstein's search for gravitational field equations was guided by the second. To a large extent, the difference between the two models accounts for the differences between Hilbert's and Einstein's approaches, including their differing capacity to incorporate available physical knowledge into their theories. The information about matter compatible with Hilbert's theory was essentially only Mie's speculative theory: The source-term in Einstein's gravitational field equations could embody the vast amount of information contained in special-relativistic continuum theory, including energy-momentum conservation, as well as Maxwell's theory.

The information available for solving the problem of gravitation was not exhausted by the two different physical models of the interaction between fields and matter. Contemporary mathematics also provided a reservoir of useful tools. The series of attempts between 1912 and 1915 to formulate a theory of gravitation, including contributions by Abraham, Nordström, and Mie, as well as Einstein and Hilbert, illustrates the range of mathematical formalisms available, from partial differential equations for a scalar field to the absolute differential calculus applied to the metric tensor. As did the physical models, different mathematical formalisms showed varying capacities for integrating the available knowledge about matter and gravitation, such as that embodied in Newtonian gravitation theory or in the observational results on Mercury's perihelion shift. To explore its capacity to integrate knowledge, a formalism needs to be elaborated and its consequences interpreted, if possible, as representations of that knowledge. The degree of such successful elaboration and interpretation, the "exploration depth" of a given formalism, determines its acceptability as a possible solution to the physical problem at hand. In early 1913, believing that the Newtonian limit could not be recovered from generally-covariant field equations, Einstein proposed the non-covariant *Entwurf* theory, from which it could be. At the end of 1915, on the basis of an increased "exploration depth" of the formalism, he decided in favor of generally-covariant equations.

Which physical models and mathematical formalisms are favored in a given historical situation depends on many factors, among them their accessibility and specific epistemological preferences that make some of them appear more attractive to certain groups than others. It was natural for a mathematician of Hilbert's caliber to start from a generally-covariant variational principle based on the metric tensor, while Einstein, ignorant of the appropriate mathematical resources, initially tried to develop his own, "pedestrian" calculus for dealing with the metric tensor.[245] It is clear that the monistic field theory model must have appealed more to Hilbert, a mathematician in search for an axiomatic foundation for all of physics, than the conceptually more clumsy dualistic model. The latter, on the other hand, was a more natural starting point for physicists such as Abraham, Einstein, and Nordström, who were familiar

245 See his calculations in "Einstein's Zurich Notebook," e.g. on p. 08L (in vol. 1 of this series). See also the "Commentary" (in vol. 2).

with the extraordinary successes of this model in the domain of electromagnetism. Images of knowledge also determine decisions on the depth and direction of exploration of a given formalism. While the question of the Newtonian limit was crucial to the physicist Einstein, Hilbert did not deal at all with this problem.

Constructs formulated by individual scientists, such as Hilbert's proposal for an axiomatic foundation of physics, are largely contingent; but their building blocks (concepts, models, techniques) are taken from the reservoir of the socially available knowledge characteristic of a given historical situation. This reservoir of shared background knowledge accounts for more than just the intercommunicability of individual contributions such as those of Hilbert and Einstein. Given that such contributions are integrated into already-shared knowledge by various processes of intellectual communication and assimilation, an equilibration process must take place between the individual constructs and the shared knowledge-reservoir. It is the outcome of this process that decides on whether a research program is progressive or degenerating in the sense of Lakatos but also the fate of an individual contribution, its longevity (the case of general relativity), its mutation, or its rapid fading-away (the case of Hilbert's contribution).

Whatever is individually constructed will be brought into contact with other elements of the shared knowledge-base, and thus integrated into it in multiple ways that, of course, are shaped by the social structures of scientific communication. The fate of an individual construct depends on the establishment of such connections. If individual constructs are not embedded, for whatever reasons, within the structures of socially available knowledge, they effectively disappear; if they are so embedded, they will be transmitted as part of shared knowledge. Usually, individual contributions are not assimilated wholesale to shared knowledge but only in a piecemeal fashion. One finds Hilbert's name associated, for instance, with the variational derivation of the field-equations but not with the program of an axiomatic foundation of physics. The "packaging" of individual contributions as they are eventually transmitted and received by a scientific community is not governed by the individual perspectives of their authors but by the more stable cognitive structures of the shared knowledge. The reception of Hilbert's contribution is thus not different from that of most scientific contributions that become assimilated into the great banquet of shared knowledge. It rarely happens that its basic epistemic structures, such as the concepts of space and time in classical physics, are themselves challenged by the growth of knowledge. Usually, these fundamental structures simply overpower any impact of individual contributions by the sheer mass of integrated knowledge they reflect. Only when individual constructs come with their own power of integrating large chunks of shared knowledge do they have a chance of altering these structures. This, in turn, only happens when the individual contributions themselves result from a process of knowledge integration and its reflection in terms of new epistemic structures.

Einstein's theory of general relativity is the result of such an integration process. Over a period of several years, he had attempted not only to reconcile classical physical knowledge about gravitation with the special-relativistic requirement of the finite propagation speed of physical interactions; but also with insights into the inseparability of gravitation and inertia, and with the special-relativistic generalization of energy-momentum conservation. Each of these building blocks: Newtonian theory, metric structure of space and time, the equivalence principle, and energy-momentum conservation, was associated with a set of possible mathematical representations, more or less well defined by physical requirements. In the case of energy-momentum conservation, for instance, Einstein had quickly arrived at an appropriate mathematical formulation, which stayed fixed throughout his search for the gravitational field equations. The inseparability of gravitation and inertia as expressed by the equivalence principle, on the other hand, could be given various mathematical representations; for Einstein the most natural at the time seemed to be the role of the metric tensor as the potentials for the inertio-gravitational field. The available mathematical representations of Einstein's building blocks were not obviously compatible with each other. In order to develop a theory comprising as much as possible of the knowledge incorporated in these building blocks, Einstein followed a double strategy.[246] On the one hand, he started from those physical principles that embody the vast store of knowledge in classical and special-relativistic physics and explored the consequences of their mathematical representations in terms of the direction of his other building blocks (his "physical strategy"). On the other hand, he started from those building blocks that had not yet been integrated into a physical theory, such as his equivalence principle, chose a mathematical representation, and explored its consequences, in the hope of being able to find a physical interpretation that also would integrate his other building blocks (his "mathematical strategy"). Eventually, he succeeded in formulating a theory that complies with these heterogeneous requirements; but only at the price of having to modify, in a process of reflection on his own premises, some of the original building blocks themselves, with far-going consequences for the structuring of the physical knowledge embodied in these building blocks, e.g. about the meaning of coordinate systems in a physical theory. That such modifications eventually became more than just personal idiosyncrasies and have had a lasting effect on the epistemic structures of physical knowledge is due to the fact that they were stabilized by the knowledge they helped to integrate into general relativity.

Hilbert's theory was clearly not based on a comparable process of knowledge integration and hence shared the fate of most scientific contribution: dissolution and assimilation to the structures of shared knowledge. Even if, in 1915, he had derived the field equations of general relativity, his theory would not have had the same "exploration depth" as that of Einstein's 1915 version, and hence not covered a similarly large domain of knowledge. Hilbert's theory is rather comparable to one of Einstein's early intermediate versions, for instance to that involving the (linearized) Einstein tensor, briefly considered in the Zurich Notebook in the winter of 1912/13. Einstein quickly rejected this candidate because it appeared to him impossible to derive the Newtonian limit from it, while Hilbert intended to publish his version in late 1915, although he had not checked its compatibility with the Newtonian limit.

246 See "Pathways out of Classical Physics …" (in vol. 1 of this series).

This difference in reacting to a similar candidates for solving the problem of the gravitational field equations obviously does not reveal any difference in the epistemic status of Hilbert's theory compared to Einstein's intermediate version but only by a different attitude with regard to a given exploration depth, motivated by the different image of knowledge that Hilbert associated with his endeavor. Such motivations make little difference to the fate of a theory in the life of the scientific community. In fact, the subsequent elaborations, revisions, and transformations of Hilbert's result testify to an equilibration process similar to that also undergone by Einstein's intermediate versions, in which ever new elements of shared knowledge found their way into Hilbert's construct. In the end, as we have seen, his theory comprises the same major building blocks of physical knowledge as those, on which general relativity is based. The exchange with Einstein and others had effectively compensated for Hilbert's original neglect of the need to consider his results in the light of physical knowledge, and thus substituted, in a way, for the "physical strategy" of Einstein's heuristics, constituting a "collective process of reflection." The fact that the equilibration process leading to general relativity essentially went on in private exchanges between Einstein and a few collaborators, while the equilibration process transforming Hilbert's theory of everything into a constituent of general relativity went on in public, as a contest between Einstein and Hilbert, Berlin and Göttingen, physics and mathematics communities, plays an astonishingly small role in the history of knowledge.

9. ACKNOWLEDGEMENTS

The research for this paper was conducted as part of a collaborative research project on the history of general relativity at the Max Planck Institute for the History of Science. We would like to warmly thank our colleagues in this project, in particular Leo Corry, Tilman Sauer, and Matthias Schemmel, for the discussions, support, and criticism, all of it intensive, which have made this collaboration a memorable and fruitful experience for both of us. Without the careful editorial work by Stefan Hajduk, who not only checked references, consistency, and language but also coordinated the contributions of the authors when they were not in one place, the paper would not have reached its present form. The authors are furthermore indebted to Giuseppe Castagnetti, Peter Damerow, Hubert Goenner, Michel Janssen, Ulrich Majer, John Norton, and David Rowe for helpful comments and conversations on the subject of this paper. We are particularly grateful to the library staff of the Max Planck Institute for the History of Science and its head, Urs Schoepflin, for their creative and indefatigable support in all library-related activities of our work. We thank the Niedersächsische Staats- und Universitätsbibliothek Göttingen (*Handschriftenabteilung*) and the library of the Mathematisches Institut, Universität Göttingen, for making unpublished material available to our project. Finally, we thank Ze'ev Rosenkranz and the Einstein Archives, The Hebrew University of Jerusalem, for permission to quote from Einstein's letters.

REFERENCES

Belinfante, F. J. 1939. "Spin of Mesons." *Physica* 6:887–898.

Becquerel, Henri. 1922. *Le Principe de la Relativité et la Théorie de la Gravitation*. Paris: Gauthier-Villars.

Birkhoff, George D., and Rudolph E. Langer. 1923. *Relativity and Modern Physics*. Cambridge, Ma.: Harvard University Press.

Blumenthal, Otto, ed. 1913. *Das Relativitätsprinzip*. 1st ed. Leipzig, Berlin: Teubner.

———. 1919. *Das Relativitätsprinzip*. 3rd ed. Leipzig, Berlin: Teubner.

———. 1923. *Das Relativitätsprinzip*. 5th ed. Leipzig, Berlin: Teubner.

———. 1974. *Das Relativitätsprinzip*. 7th ed. Darmstadt: Wissenschaftliche Buchgesellschaft.

Born, Max. 1914. "Der Impuls-Energie-Satz in der Elektrodynamik von Gustav Mie." *Königliche Gesellschaft der Wissenschaften zu Göttingen. Nachrichten* (1914):23–36.

Born, Max, and Leopold Infeld. 1934. "Foundations of the New Field Theory." *Royal Society of London. Proceedings* A 144:425–451.

Caratheodory, Constantin. 1935. *Variationsrechnung und partielle Differentialgleichungen erster Ordnung*. Leipzig, Berlin: B. G. Teubner.

Corry, Leo. 1997. "David Hilbert and the Axiomatization of Physics (1894–1905)." *Archive for History of Exact Sciences* 51:83–198.

———. 1999a. "David Hilbert between Mechanical and Electromagnetic Reductionism (1910–1915)." *Archive for History of Exact Sciences* 53:489–527.

———. 1999b. "From Mie's Electromagnetic Theory of Matter to Hilbert's Unified Foundations of Physics." *Studies in History and Philosophy of Modern Physics* 30 B (2):159–183.

———. 1999c. "David Hilbert: Geometry and Physics (1900–1915)." In J. J. Gray (ed.), *The Symbolic Universe: Geometry and Physics (1890–1930)*, Oxford: Oxford University Press, 145–188.

———. 2004. *David Hilbert and the Axiomatization of Physics, 1898–1918: From "Grundlagen der Geometrie" to "Grundlagen der Physik"*. Dordrecht: Kluwer.

Corry, Leo, Jürgen Renn, and John Stachel (eds.). 1997. *Belated Decision in the Hilbert-Einstein Priority Dispute*. Vol. 278, Science.

CPAE 4: Martin J. Klein, A. J. Kox, Jürgen Renn, and Robert Schulmann (eds.). 1995. *The Collected Papers of Albert Einstein*. Vol. 4. *The Swiss Years: Writings, 1912–1914*. Princeton: Princeton University Press.

CPAE 6: A. J. Kox, Martin J. Klein, and Robert Schulmann (eds.). 1996. *The Collected Papers of Albert Einstein*. Vol. 6. *The Berlin Years: Writings, 1914–1917*. Princeton: Princeton University Press.

CPAE 6E: *The Collected Papers of Albert Einstein*. Vol. 6. *The Berlin Years: Writings, 1914–1917*. English edition translated by Alfred Engel, consultant Engelbert Schucking. Princeton: Princeton University Press, 1996.

CPAE 8: Robert Schulmann, A. J. Kox, Michel Janssen, and József Illy (eds.). 1998. *The Collected Papers of Albert Einstein*. Vol. 8. *The Berlin Years: Correspondence, 1914–1918*. Princeton: Princeton University Press.

CPAE 8E: *The Collected Papers of Albert Einstein*. Vol. 8. *The Berlin Years: Correspondence, 1914–1918*. English edition translated by Ann M. Hentschel, consultant Klaus Hentschel. Princeton: Princeton University Press, 1998.

Csikszentmihalyi, Mihaly. 1988. "Society, Culture and Person: a Systems View of Creativity." In R. J. Sternberg (ed.), *The Nature of Creativity*. Cambridge: Cambridge University Press.

Earman, John, and Clark Glymour. 1978. "Einstein and Hilbert: Two Months in the History of General Relativity." *Archive for History of Exact Sciences* 19:291–308.

Eddington, Arthur Stanley. 1923. *The Mathematical Theory of Relativity*. Cambridge: The University Press.

———. 1924. *The Mathematical Theory of Relativity*. 2nd ed. Cambridge: The University Press.

———. 1925. *Relativitätstheorie in Mathematischer Behandlung*. Translated by Alexander Ostrowski Harry Schmidt. Berlin: Springer.

Einstein, Albert. 1913. "Zum gegenwärtigen Stande des Gravitationsproblems." *Physikalische Zeitschrift* 14 (25):1249–1262. (English translation in volume 3 of this series.)

———. 1914a. "Prinzipielles zur verallgemeinerten Relativitätstheorie und Gravitationstheorie." *Physikalische Zeitschrift* 15:176–180.

———. 1914b. "Die formale Grundlage der allgemeinen Relativitätstheorie." *Königlich Preußische Akademie der Wissenschaften (Berlin). Sitzungsberichte* (1914) (XLI):1030–1085.

———. 1915a. "Zur allgemeinen Relativitätstheorie." *Königlich Preußische Akademie der Wissenschaften (Berlin). Sitzungsberichte* (1915) (XLIV):778–786.

———.. 1915b. "Zur allgemeinen Relativitätstheorie (Nachtrag)." *Königlich Preußische Akademie der Wissenschaften (Berlin). Sitzungsberichte* (1915) (XLVI):799–801.

———. 1915c. "Erklärung der Perihelbewegung des Merkur aus der allgemeinen Relativitätstheorie." *Königlich Preußische Akademie der Wissenschaften (Berlin). Sitzungsberichte* (1915) (XLVII):831–839.

———. 1915d. [Zusammenfassung der Mitteilung "Erklärung der Perihelbewegung des Merkur aus der allgemeinen Relativitätstheorie."] *Königlich Preußische Akademie der Wissenschaften (Berlin). Sitzungsberichte* (1915) (XLVII):803.

———. 1915e. "Die Feldgleichungen der Gravitation." *Königlich Preußische Akademie der Wissenschaften (Berlin). Sitzungsberichte* (1915) (XLVIII–XLIX):844–847.

———. 1916a. "Die Grundlage der allgemeinen Relativitätstheorie." *Annalen der Physik* 49 (7):769–822.

———. 1916b. "Hamiltonsches Prinzip und allgemeine Relativitätstheorie." *Königlich Preußische Akademie der Wissenschaften (Berlin). Sitzungsberichte* (1916) (XLII):1111–1116.

———. 1916c. "Näherungsweise Integration der Feldgleichungen der Gravitation." *Sitzung der physikalisch-mathematischen Klasse* 668–96. (CPAE 6, Doc. 32, 348–57)

Einstein, Albert, and Marcel Grossmann. 1914. "Kovarianzeigenschaften der Feldgleichungen der auf die verallgemeinerte Relativitätstheorie gegründeten Gravitationstheorie." *Zeitschrift für Mathematik und Physik* 63 (1 / 2):215–225.

Fölsing, Albrecht. 1997. *Albert Einstein: a biography*. New York: Viking.

Frei, Günther, ed. 1985. *Der Briefwechsel David Hilbert-Felix Klein (1886–1918)*. Vol. 19, *Arbeiten aus der Niedersächsischen Staats- und Universitätsbibliothek Göttingen*. Göttingen: Vandenhoeck & Ruprecht.

Goenner, Hubert. 2004. "On the History of Unified Field Theories." *Living Reviews of Relativity* 7 <http://www.livingreviews.org>.

Goenner, Hubert, Jürgen Renn, Jim Ritter, and Tilman Sauer (eds.). 1999. *The Expanding Worlds of General Relativity*. (*Einstein Studies* vol. 7.) Boston: Birkhäuser.

Guth, E. 1970. "Contribution to the History of Einstein's Geometry as a Branch of Physics." In *Relativity,* edited by M. Carmeli et al. New York, London: Plenum Press, 161–207.

Havas, Peter. 1989. "The Early History of the 'Problem of Motion' in General Relativity." In *Einstein and the History of General Relativity*, edited by Don Howard and John Stachel. (*Einstein Studies* vol. 1.) Boston: Birkhäuser, 234–276.

Hilbert, David. 1905. "Logische Prinzipien des mathematischen Denkens." Ms. Vorlesung SS 1905, annotated by E. Hellinger, Bibliothek des Mathematischen Seminars, Göttingen.

———. 1912–13. "Molekulartheorie der Materie." Ms. Vorlesung WS 1912–13, annotated by M. Born, Nachlass Max Born #1817, Stadtbibliothek Berlin.

———. 1913. "Elektronentheorie." Ms. Vorlesung SS 1913, Bibliothek des Mathematischen Seminars, Göttingen.

———. 1916. "Die Grundlagen der Physik. (Erste Mitteilung)." *Königliche Gesellschaft der Wissenschaften zu Göttingen. Mathematisch-physikalische Klasse. Nachrichten* (1915):395–407. (English translation in this volume.)

———. 1917. "Die Grundlagen der Physik (Zweite Mitteilung)." *Königliche Gesellschaft der Wissenschaften zu Göttingen. Mathematisch-physikalische Klasse. Nachrichten* (1917):53–76. (English translation in this volume.)

———. 1924. "Die Grundlagen der Physik." *Mathematische Annalen* 92:1–32.

———, ed. 1935. *Gesammelte Abhandlungen, Band III: Analysis, Grundlagen der Mathematik, Physik, Verschiedenes, Lebensgeschichte*. [1932–35, 3 vols.]. Berlin: Springer.

———. 1971. "Über meine Tätigkeit in Göttingen." In *Hilbert-Gedenkenband*, ed. K. Reidemeister. Berlin, Heidelberg, New York: Springer, 79–82.

Howard, Don, and John D. Norton. 1993. "Out of the Labyrinth? Einstein, Hertz, and the Göttingen Answer to the Hole Argument." In *The Attraction of Gravitation: New Studies in the History of General Relativity,* edited by John Earman, Michel Janssen and John D. Norton. Boston/Basel/Berlin: Birkhäuser, 30–62.

Janssen, Michel and Matthew Mecklenburg. 2006. "Electromagnetic Models of the Electron and the Transition from Classical to Relativistic Mechanics." In *Interactions: Mathematics, Physics and Philosophy, 1860–1930*, edited by V. F. Hendricks et al. *Boston Studies in the Philosophy of Science*, Vol. 251. Dordrecht: Springer, 65–134.

Kerschensteiner, Georg, ed. 1887. *Paul Gordan's Vorlesungen über Invariantentheorie. Zweiter Band: Binäre Formen*. Leipzig: Teubner.

Klein, Felix. 1917. "Zu Hilberts erster Note über die Grundlagen der Physik." *Königliche Gesellschaft der Wissenschaften zu Göttingen. Mathematisch-physikalische Klasse. Nachrichten* (1917):469–482.

———. 1918a. "Über die Differentialgesetze für die Erhaltung von Impuls und Energie in der Einsteinschen Gravitationstheorie." *Königliche Gesellschaft der Wissenschaften zu Göttingen. Mathematisch-physikalische Klasse. Nachrichten* (1918):171–189.

———. 1918b. "Über die Integralform der Erhaltungsätze und die Theorie der räumlich-geschlossenen Welt." *Königliche Gesellschaft der Wissenschaften zu Göttingen. Mathematisch-physikalische Klasse. Nachrichten* (1918):394–423.

———. 1921. "Zu Hilberts erster Note über die Grundlagen der Physik." In *Gesammelte Mathematische Abhandlungen,* edited by R. Fricke and A. Ostrowski. Berlin: Julius Springer, 553–567.

Komar, Arthur. 1958. "Construction of a Complete Set of Independent Observables in the General Theory of Relativity." *Physical Review* 111:1182–1187.

Kretschmann, Erich. 1917. "Über den physikalischen Sinn der Relativitätspostulate, A. Einsteins neue und seine ursprüngliche Relativitätstheorie." *Annalen der Physik* 53 (16):575–614.

Laue, Max. 1911a. "Zur Dynamik der Relativitätstheorie." *Annalen der Physik* 35: 524–542.

———. 1911b. *Das Relativitätsprinzip*. Braunschweig: Friedrich Vieweg und Sohn.

Laue, Max von. 1920. "Theoretisches über neuere optische Beobachtungen zur Relativitätstheorie." *Physikalische Zeitschrift* 21:659–662.

Lichnerowicz, André. 1946. "Sur le caractère euclidien d'espaces-temps extérieurs statiques partout réguliers." *Academie des Sciences* (Paris). *Comptes Rendus* 222:432–436.

Lorentz, Hendrik A., et al. 1923. *The Principle of Relativity.* London: Methuen & Co.

Majer, Ulrich and Tilman Sauer. 2005. "'Hilbert's World Equations' and His Vision of a Unified Science." In *The Universe of General Relativity*, edited by A. Kox and J. Eisenstaedt. (*Einstein Studies*, vol. 11.) Boston: Birkhäuser, 259–276.

Mehra, Jagdish. 1974. *Einstein, Hilbert, and the Theory of Gravitation. Historical Origins of General Relativity Theory.* Dordrecht, Boston: D. Reidel Publishing Company.

Mie, Gustav. 1912a. "Grundlagen einer Theorie der Materie. Erste Mitteilung." *Annalen der Physik* 37:511–534. (English translation of excerpts in this volume.)

———. 1912b. "Grundlagen einer Theorie der Materie. Zweite Mitteilung." *Annalen der Physik* 39:1–40.

———. 1913. "Grundlagen einer Theorie der Materie. Dritte Mitteilung." *Annalen der Physik* 40:1–66. (English translation of excerpts in this volume.)

Noether, Emmy. 1918. "Invariante Variationsprobleme." *Königliche Gesellschaft der Wissenschaften zu Göttingen. Mathematisch-physikalische Klasse. Nachrichten* (1918):235–257.

Norton, John D. 1984. "How Einstein Found His Field Equations, 1912–1915." *Historical Studies in the Physical Sciences* 14:253–316.

Pais, Abraham. 1982. *'Subtle is the Lord ...' The Science and the Life of Albert Einstein*. Oxford, New York, Toronto, Melbourne: Oxford University Press.

Papapetrou, Achille. 1974. *Lectures on General Relativity.* Dordrecht/Boston: D. Reidel.

Pauli, Wolfgang. 1921. "Relativitätstheorie." In *Encyklopädie der mathematischen Wissenschaften, mit Einschluss ihrer Anwendungen*, edited by Arnold Sommerfeld. Leipzig: B. G. Teubner, 539–775.

———. 1958. *Theory of Relativity.* Translated by G. Field. London: Pergamon.

———. 1979. *Scientific Correspondence with Bohr, Einstein, Heisenberg, a.o. Volume 1: 1919–1929.* New York: Springer.

Reidemeister, Kurt, ed. 1971. *Hilbert-Gedenkenband.* Berlin, Heidelberg, New York: Springer,

Renn, Jürgen. 1994. "The Third Way to General Relativity." Preprint n° 9, *Max Planck Institute for the History of Science,* Berlin (http://www.mpiwg-berlin.mpg.de/Preprints/P9.PDF). (Revised edition in vol. 3 of this series.)

Renn, Jürgen, and Tilman Sauer. 1996. "Einsteins Züricher Notizbuch." *Physikalische Blätter* 52:865–872.

———. 1999. "Heuristics and Mathematical Representation in Einstein's Search for a Gravitational Field Equation." In (Goenner et al. 1999, 87–125).

Rosenfeld, Leon. 1940. "Sur le tenseur d'impulsion-énergie." *Mémoires de l'Academie royale de Belgique* 18 (16):1–30.

Rowe, David. 1989. "Klein, Hilbert, and the Göttingen Mathematical Tradition." *Osiris* 5:186–213.

———. 1999. "The Göttingen Response to General Relativity and Emmy Noether's Theorems." In *The Visual World: Geometry and Physics (1890–1930)*, ed., J. J. Gray. Oxford: Oxford University Press.

Sauer, Tilman. 1999. "The Relativity of Discovery: Hilbert's First Note on the Foundations of Physics." *Archive for History of Exact Sciences* 53:529–575.

———. 2002. "Hopes and Disappointments in Hilbert's Axiomatic 'Foundations of Physics.'" In *History of Philosophy and Science: new trends and perspectives*, ed. M. Heidelberger and F. Stadler. Dordrecht: Kluwer, 225–237.

Schaffner, Kenneth K. 1972. *Nineteenth-Century Aether Theories*. Oxford: Pergamon Press.

Schouten, Jan A. 1924. *Der Ricci-Kalkül*, 1st ed. Berlin: Springer-Verlag.

Schouten, Jan A., and Dirk J. Struik. 1935. *Algebra und Übertragungslehre*. Vol. 1, *Einführung in die neueren Methoden der Differentialgeometrie*. Groningen, Batavia: P. Noordhoff.

Schwarzschild, Karl. 1916. "Über das Gravitationsfeld eines Massenpunktes nach der Einsteinschen Theorie." *Königlich Preußische Akademie der Wissenschaften (Berlin). Sitzungsberichte* (1916) (VII):189–196.

Siegmund-Schultze, Reinhard. 1998. *Mathematiker auf der Flucht vor Hitler: Quellen und Studien zur Emigration einer Wissenschaft*. Vol. 10, *Dokumente zur Geschichte der Mathematik*. Braunschweig Wiesbaden: Vieweg.

Slebodzinski, Wladyslaw. 1931. "Sur les equations de Hamilton." *Bulletin de l'Academie royale de Belgique* (5) (17):864–870.

Stachel, John. 1989. "Einstein's Search for General Covariance, 1912–1915." In *Einstein and the History of General Relativity,* edited by Don Howard and John Stachel. Boston/Basel/Berlin: Birkhäuser, 63–100.

———. 1992. "The Cauchy Problem in General Relativity - The Early Years." In *Studies in the History of General Relativity,* edited by Jean Eisenstaedt and A. J. Kox. Boston/Basel/Berlin: Birkhäuser, 407–418.

———. 1994. "Scientific Discoveries as Historical Artifacts." In *Current Trends in the Historiography of Science,* edited by Kostas Gavroglu. Dordrecht, Boston: Reidel, 139–148.

———. 1999. "New Light on the Einstein-Hilbert Priority Question." *Journal of Astrophysics and Astronomy* 20:91–101. Reprinted in (Stachel 2002).

———. 2002. *Einstein from 'B' to 'Z'*. (*Einstein Studies* vol. 9.) Boston: Birkhäuser.

Thorne, Kip S. 1994. *Black Holes and Time Warps: Einstein's Outrageous Legacy.* New York, London: Norton.

Vizgin, Vladimir P. 1989. "Einstein, Hilbert, and Weyl: The Genesis of the Geometrical Unified Field Theory Program." In *Einstein and the History of General Relativity,* edited by Don Howard and John Stachel. Boston/Basel/Berlin: Birkhäuser, 300–314.

———. 1994. *Unified Field Theories in the First Third of the 20th Century.* Translated by Barbour, Julian B. Edited by E. Hiebert and H. Wussing. Vol. 13, *Science Networks, Historical Studies*. Basel, Boston, Berlin: Birkhäuser.

Walter, Scott. 1999. "Minkowski, Mathematicians, and the Mathematical Theory of Relativity." In (Goenner et al. 1999, 45–86).

Weitzenböck, Roland. 1920. "Über die Wirkungsfunktion in der Weyl'schen Physik." *Akademie der Wissenschaften (Vienna). Mathematisch-naturwissenschaftliche Klasse. Sitzungsberichte* 129:683–696.

Weyl, Hermann. 1917. "Zur Gravitationstheorie." *Annalen der Physik* 54:117–145.

———. 1918a. *Raum-Zeit-Materie. Vorlesungen über allgemeine Relativitätstheorie*. 1st ed. Berlin: Julius Springer.

———. 1918b. *Raum-Zeit-Materie. Vorlesungen über allgemeine Relativitätstheorie*. 2nd ed. Berlin: Julius Springer.

———. 1918c. "Gravitation und Elektrizität." *Königlich Preußische Akademie der Wissenschaften (Berlin). Sitzungsberichte* (1918):465–480.

———. 1919. *Raum-Zeit-Materie. Vorlesungen über allgemeine Relativitätstheorie*. 3rd, revised ed. Berlin: Julius Springer.

———. 1921. *Raum-Zeit-Materie. Vorlesungen über allgemeine Relativitätstheorie*. 4th ed. Berlin: Julius Springer.

———. 1923. *Raum-Zeit-Materie. Vorlesungen über allgemeine Relativitätstheorie*. 5th, revised ed. Berlin: Julius Springer.

Whittaker, Edmund Taylor. 1951. *A History of the Theories of Aether and Electricity.* Vol. 1: *The Classical Theories*. London: Nelson.

TILMAN SAUER

EINSTEIN EQUATIONS AND HILBERT ACTION: WHAT IS MISSING ON PAGE 8 OF THE PROOFS FOR HILBERT'S FIRST COMMUNICATION ON THE FOUNDATIONS OF PHYSICS?[1]

1. INTRODUCTION

In contrast to Einstein's discovery of special relativity in 1905, his path towards the theory of general relativity is documented by a rich historical record. Not only did Einstein publish quite a few papers on earlier versions of a generalized theory of relativity, we also have a number of research manuscripts from crucial periods of his search, and we have an extensive correspondence from the relevant years. Hilbert's involvement in the discovery of general relativity is less abundantly documented but also here we have a few key documents that shed light on his work. Compared to other episodes in the history of science, the history of general relativity is very well written, and specifically the competition between Einstein and Hilbert in the final weeks before the publication of generally covariant field equations of gravitation in late 1915 has been commented on extensively.[2] Nevertheless, much of the historical literature on the Einstein-Hilbert competition took sides in what was perceived as a priority debate and it still seems worthwhile to come to a succinct and balanced assessment of the respective contributions of both authors in the final establishment of the general theory of relativity. In this respect, a set of proofs of Hilbert's relevant paper are of some significance and with those proofs the fact that a piece of them is missing. Although the fact that a piece of those proofs is missing is well known and was briefly commented on by several authors, the question naturally arises as to whether that missing part could have contained information that would compel us to reassess the historical account?

1 This paper was first published in *Archive for History of Exact Sciences* 59 (2005) 577–590, and is reprinted here with their kind permission.

2 See (Corry 2004; Corry, Renn, and Stachel 1997; Earman and Glymour 1978; Logunov, Mestvirishvili and Petrov 2004; Mehra 1974; Norton 1984; Pais 1982; Rowe 2001; Sauer 1999; Stachel 1999; Vizgin 2001), and "Hilbert's Foundation of Physics …" (in this volume), as well as further references cited in these works.

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 4
Gravitation in the Twilight of Classical Physics: The Promise of Mathematics.

2. THE CONTEXT

Before focussing on some minor yet significant details of the historical record, let me briefly review the broader historical context. In 1907, Einstein first formulated his equivalence hypothesis according to which no physical experiment can distinguish between the existence of a homogeneous, static gravitational field in a Newtonian inertial frame of reference and a uniformly and rectilinearly accelerated frame of reference that is free of any gravitational field. The hypothesis linked the problem of generalizing the special theory of relativity to accelerated motion with the problem of a relativistic theory of gravitation. In 1912, Einstein realized that such a relativistic theory of gravitation could not be achieved using a scalar gravitational potential but required the introduction of the metric tensor as the crucial mathematical object for a generalized theory of relativity. Together with his mathematician friend Marcel Grossmann, Einstein published an "Outline of a Generalized Theory of Relativity and a Theory of Gravitation" in 1913 (Einstein and Grossmann 1913). The theory of this "Outline" (*Entwurf*) has already many features of the final theory of general relativity except for one "dark spot." Einstein and Grossmann did not succeed in finding gravitational field equations for the components of the metric tensor that were both generally covariant and acceptable from the point of view of Einstein's understanding of the requirements for a satisfactory theory of gravitation.

The final episode of Einstein's path towards General Relativity began in the fall of 1915 when Einstein lost faith in the validity of the field equations of his "Outline" and reverts to a reassessment of the mathematics of general covariance as developed in the work of Riemann, Christoffel, Ricci and Levi-Civita. The final steps were taken in four successive communications to the Prussian Academy of Sciences, all of them presented for publication in the month of November 1915 (Einstein 1915a, b, c, d). On November 4, Einstein advanced field equations that are based on the Ricci tensor but that are not yet generally covariant (Einstein 1915a). Instead, by stipulation of a restrictive condition on the admissible coordinates, he split off a part of the Ricci tensor and equated the remaining part to an unspecified energy-momentum tensor as the source of the gravitational field. In an addendum to this paper, presented a week later on November 11 (Einstein 1915b), Einstein temporarily entertains the speculation that all matter might be of electromagnetic origin. This assumption allowed him to advance a generally covariant field equation of gravitation where the Ricci tensor is directly set proportional to the energy-momentum tensor. Another week later, Einstein presented a paper to the Berlin Academy in which he successfully computed the anomalous advance of the perihelion of Mercury on the basis of his new equations (Einstein 1915c). And yet another week later, Einstein realized that he can add a trace term to the right-hand side of his field equations which turns them into what we now refer to as the Einstein equations (Einstein 1915d).

David Hilbert's path towards general relativity is a rather different one. Half a generation older than Einstein, Hilbert in 1900 formulated his famous 23 problems of mathematical research of the coming century to the International Congress of Mathematicians in Paris. The sixth of these problems asked for an axiomatization of physics. After working on the theory of integral equations in the first decade of the century, Hilbert himself then turned to an intense study of all fields of theoretical physics. In the course of his study of contemporary physics literature he soon became interested in an attempt by the German physicist Gustav Mie to generalize Maxwellian electrodynamics so as to turn it into a theory of matter. Mie's idea was to take Maxwellian electrodynamics in its variational formulation but to search for a generalized Lagrangian entering the action, keeping the requirement of Lorentz covariance but allowing for the Lagrangian to depend explicitly on the electromagnetic vector potential. Mie's hope was to find a modified Lagrangian that would produce modified Maxwell equations which, on microscopic scales, would allow for particle-like solutions. Around that time, Hilbert also became interested in Einstein's recent work on a relativistic theory of gravitation and invited Einstein to give a series of lectures on his new theory to the Göttingen mathematicians and physicists. After Einstein presented his theory in Göttingen in July 1915, Hilbert left Göttingen for his summer vacations and began pondering on Einstein's "Outline" theory. Shortly after coming back to Göttingen at the beginning of the winter term, Hilbert himself then presented a paper to the Göttingen Academy of Sciences. In this communication, Hilbert presented a theory of the "Foundations of Physics" which combined Mie's idea of a generalized electrodynamics with Einstein's idea of a generally covariant theory of gravitation.

The dateline on Hilbert's *First Communication* on the *Foundations of Physics* (Hilbert 1915) says that it was presented to the Göttingen Academy of Sciences on 20 November 1915. The dateline on Einstein's note on *The Field Equations of Gravitation* (Einstein 1915d) says that it was presented to the Berlin Academy of Sciences on 25 November 1915. From a comparison of the two publications, it appears that Hilbert preceded Einstein with the publication of the final gravitational field equations of general relativity by five days, notwithstanding the fact that both authors arrived at these equations along very different routes.

The question as to where the correct field equations of gravitation are first found in print is in need of some qualification. The gravitational field equations of general relativity may be written in two very different yet essentially equivalent ways. Einstein published his final field equations of 25 November (Einstein 1915d, 845),

$$G_{im} = -\kappa\left(T_{im} - \frac{1}{2}g_{im}T\right), \tag{1}$$

as an explicit set of differential equations for the components of the metric tensor g_{im}. Using the Ricci tensor G_{im} as the differential operator acting on the metric and the energy-momentum tensor T_{im} in the source term on the right-hand side made sure that his equations retained its form under arbitrary coordinate transformations, i.e. made them generally covariant. Adding a trace term $-(1/2)g_{im}T$ where $T = \Sigma g^{\rho\sigma}T_{\rho\sigma}$ to the right-hand side of his equations in his last November paper did not violate this feature. Hilbert published the gravitational field equations in implicit form in terms of a variational principle. He axiomatically postulated an action integral (Hilbert 1915, 396)

$$\int H\sqrt{g}d\omega, \tag{2}$$

where $g = |g_{\mu\nu}|$, $d\omega = dw_1 dw_2 dw_3 dw_4$ for spacetime coordinates w_i and required that the Lagrangian H that enters into the action of his variational formulation be invariant under arbitrary coordinate transformations. He also assumed that the Lagrangian splits into the sum of two parts, a gravitational part given by the Riemann curvature scalar and a matter part which he left unspecified except for the postulation that it depend only on the components of the metric and the components of the electromagnetic vector potential and its first derivatives. This specification technically renders Einstein's equations equivalent to Hilbert's action, except for some ambiguity in the assumptions on how the source term is to be specified, i.e. on the fundamental constitution of matter. Both Hilbert and Einstein had left the matter term undetermined to some extent. Einstein had not specified his source term at all. Hilbert had axiomatically required that the source term depend only on the electromagnetic variables and hence that all matter is of electromagnetic origin.

But several years ago it was pointed out (Corry, Renn and Stachel 1997) that a set of proofs for Hilbert's *First Communication* is extant in the Hilbert archives in Göttingen. It bears a printer's stamp of December 6, 1915, and differs in some significant respects from the published version.[3] The main difference pertains to a different treatment of the energy concept that motivated an axiomatic restriction of the general covariance of Hilbert's theory and that was substantially rewritten for the published version. In the published paper, the discussion of the energy concept no longer results in the postulation of a restriction of the general covariance. It was also pointed out that the proofs did not contain the explicit version of the gravitational field equations in terms of the Einstein tensor as does Hilbert's published paper. What we now call the Einstein tensor is obtained by adding a trace term to the Ricci tensor, its covariant divergence vanishes identically, and it is obtained from the explicit variation of the gravitational part of Hilbert's action integral. To be precise, in Einstein's paper of 25 November the trace term was added on the right-hand side of the field equation to the source term and not to the Ricci tensor on the left hand side and strictly speaking his paper does not contain the Einstein tensor explicitly but this difference is a minor detail since both variants are trivially equivalent. In view of the differences between the proofs and the published paper general agreement seems to have been reached[4] about the conclusion that the proofs unequivocally rule out the possibility that Einstein may have taken the clue of adding a trace term to his field equations of 11 November (Einstein 1915b) from Hilbert's paper (1915). No agreement, however, was reached on the question as to the path along which Hilbert arrived at his finally

3 Hilbert's paper was eventually issued only on March 31, 1916, but off-prints of the final version were available to Hilbert already by mid-February (Sauer 1999, note 74). Einstein's November papers were each published a week after their presentation to the Prussian Academy.

4 See (Corry 2004; Rowe 2001; Sauer 1999; Stachel 1999; Vizgin 2001) and also "Hilbert's Foundation of Physics …" (in this volume).

published theory: by taking the main clues from Einstein's paper, as suggested in (Corry, Renn and Stachel 1997), or along an independent logic of discovery, as first advocated in explicit response to this claim in (Sauer 1999). It also remains an open question to what extent Einstein in those weeks of October and November 1915 had heard directly or indirectly about Hilbert's work on his theory and to what extent he may have been influenced by what he heard, e.g. in entertaining temporarily the speculation that all matter is of electromagnetic origin.

To add to the complexity of the issue, it so happens that a portion of one sheet of the extant proofs for Hilbert's *First Communication* is missing.[5] In view of this fact, it seems worthwhile to discuss the question what part of the argument of the proofs is missing and whether an answer to this question may possibly affect our assessment of the Einstein-Hilbert competition in late 1915. In the following, I will argue that an analysis of the internal structure of the text and argument of the proofs and the published version of Hilbert's paper shows that the missing piece in all probability did not contain an explicit version of the Einstein tensor and its trace term. The analysis rather suggests that it contained an explicit form of the Riemann curvature scalar and the Ricci tensor as a specification of the Lagrangian in Hilbert's variational principle.

3. WHAT IS MISSING IN THE PROOFS

Axiom I of Hilbert's *First Communication*, as presented on page 2 of his proofs,[6] introduces an action integral[7]

$$\int H\sqrt{g}d\tau \tag{3}$$

where $g = |g_{\mu\nu}|, d\tau = dw_1 dw_2 dw_3 dw_4$ and H is a Lagrangian density that depends on the components of the metric $g_{\mu\nu}$, its first and second derivatives with respect to the coordinates w_1 of the spacetime manifold, $g_{\mu\nu l} = \partial g_{\mu\nu}/\partial w_l$ and $g_{\mu\nu lk} = \partial^2 g_{\mu\nu}/\partial w_l \partial w_k$, respectively, and also depends on the components of the electromagnetic vector potential q_s and its first derivatives $q_{sl} = \partial q_s/\partial w_l$. Specifically, the axiom demands that the laws of physics be given by the vanishing of the

5 See (Sauer 1999, note 75) and "Hilbert's Foundation of Physics …" note 6 (in this volume).

6 Niedersächsische Staats- und Universitätsbibliothek (NSUB), Handschriftenabteilung, Cod. Ms. Hilbert 634, f.23-29. Facsimile versions of both Hilbert's proofs and of the published version were made available online by the Max Planck Institute for the History of Science, Berlin, on <http://echo.mpiwg-berlin.mpg.de/content/relativityrevolution/hilbert>. A facsimile of the published version is also available online from the website of the *Göttinger Digitalisierungszentrum* of the NSUB, see <http://gdz.sub.uni-goettingen.de/gdz>.

7 The argument being partly one of textual exegesis, I am keeping strictly to Hilbert's notation. He uses an imaginary time-coordinate and, following standard usage of the time, refers to the Lagrangian density as a Hamiltonian function. Contrary to later and current usage, Hilbert and Einstein at the time also consistently wrote contravariant indices of coordinate differentials as subscript indices. Hilbert also uses subscript indices to denote partial coordinate derivatives without, however, indicating this meaning by separating the index with a comma.

variation of the action integral with respect to the fourteen potentials $g_{\mu\nu}$ and q_s for some as yet unspecified function H.

Axiom II, immediately following, then demands that H must be an invariant under all coordinate transformations. Other than that, the Lagrangian H is left undetermined by the axioms.

On page 3, Hilbert writes down the "ten Lagrangian differential equations"[8]

$$\frac{\partial\sqrt{g}H}{\partial g^{\mu\nu}} = \sum_k \frac{\partial}{\partial w_k}\frac{\partial\sqrt{g}H}{\partial g_k^{\mu\nu}} - \sum_{k,l}\frac{\partial^2}{\partial w_k \partial w_l}\frac{\partial\sqrt{g}H}{\partial g_{kl}^{\mu\nu}}, \qquad (\mu, \nu = 1, 2, 3, 4) \tag{4-pr}$$

which he calls the "fundamental equations of gravitation," and the four Lagrangian differential equations

$$\frac{\partial\sqrt{g}H}{\partial q_h} = \sum_k \frac{\partial}{\partial w_k}\frac{\partial\sqrt{g}H}{\partial q_{hk}}, \qquad (h = 1, 2, 3, 4) \tag{5-pr}$$

which he calls the "fundamental equations of electrodynamics or the generalized Maxwell equations." Hilbert then proceeds to discuss the concept of energy in the theory by looking at what we would now call Lie variations of the action, i.e. variations of the metric that arise from pure coordinate transformations. In the course of this discussion he introduces the notational "abbreviation"

$$[\sqrt{g}H]_{\mu\nu} = \frac{\partial\sqrt{g}H}{\partial g^{\mu\nu}} - \sum_k \frac{\partial}{\partial w_k}\frac{\partial\sqrt{g}H}{\partial g_k^{\mu\nu}} + \sum_{k,l}\frac{\partial^2}{\partial w_k \partial w_l}\frac{\partial\sqrt{g}H}{\partial g_{kl}^{\mu\nu}} \tag{4}$$

which he calls "the Lagrangian variational derivative of $\sqrt{g}H$ with respect to $g^{\mu\nu}$." He observes that the fundamental equations of gravitation (4-pr) may now compactly be written as

$$[\sqrt{g}H]_{\mu\nu} = 0. \tag{8-pr}$$

Hilbert's discussion of the energy concept in the proofs does not provide any further specifications of the Lagrangian H, although it does lead to a third axiom that restricts the covariance of the generally covariant equations (4-pr), (5-pr), by demanding that the physically admissible coordinates for the theory obey a set of equations that are not generally covariant.[9]

It is towards the end of the discussion of the problem of the energy concept and the significance of his third axiom, which runs until the bottom of page 7, that we find two passages missing in the proofs, since the top portion of the sheet that contains pages 7and 8 was cut off.[10] Without any further discussion of Hilbert's treatment of

---

8 Hilbert tended to use equation numbers only for those equations that he actually referred to in his text. I will use his own equation numbers whenever an equation was given one and indicate this fact by adding "-pr" resp. "-pu" to the number, depending on whether it is the equation number used in the proofs or the published version, respectively.

the energy concept,[11] I will assume that the missing portion on the top of page 7, i.e. on the verso of the top of page 8, is not in any way relevant to the question under investigation in this note. But what is missing on page 8?

On page 8 of the proofs, immediately following the excised portion, Hilbert asserts: "Since K depends only on $g^{\mu\nu}$, $g^{\mu\nu}_{\ \ k}$, $g^{\mu\nu}_{\ \ lk}$, the ansatz (17-pr) allows us to express the energy E [...] solely as a function of the gravitational potentials $g^{\mu\nu}$ and their derivatives, if only we assume L not to depend on $g^{\mu\nu}_s$, but only on $g^{\mu\nu}$, q_s, q_{sk}." In the next sentence, Hilbert states that he would make that latter assumption in the following.

We observe that the quantities K and L had not been used earlier in the proofs,[12] and we may conclude that K must have been introduced just before as a function of the components of the metric and its derivatives only, and that L must have been introduced just before as a function of the electromagnetic potential, its derivatives as well as of the components of the metric and its first derivatives, although the dependence of L on the derivatives of the metric is immediately assumed away for the rest of the text. We also observe that the previous page has an equation that is numbered (16-pr) and that the next line gives an equation that is numbered (18-pr). The equation with number (17-pr) is referred to a few pages later, on page 11, where Hilbert writes that "because of (17-pr)" the fundamental equations of gravitation (8-pr) take the form

$$[\sqrt{g}K]_{\mu\nu} + \frac{\partial\sqrt{g}L}{\partial g^{\mu\nu}} = 0, \tag{26-pr}$$

and the fundamental equations of electrodynamics take the form

$$[\sqrt{g}L]_h = 0. \tag{27-pr}$$

Spelling out $[\sqrt{g}K]_{\mu\nu}$ in terms of the definition (4), eq. (26-pr) reads

9 Contrary to the discussion in (Logunov, Mestvirishvili and Petrov 2004), this condition is conceptually very different from what we now call a coordinate condition since it pertains to *any* possible application of the field equations. In these volumes, such restricting equations are called "coordinate restrictions" as opposed to "coordinate conditions." Nonetheless, there is a significant difference between Einstein's use of "coordinate restrictions" prior to his final version of the general theory of relativity and Hilbert's third axiom in the proofs. Einstein used "coordinate restrictions" to derive field equations that are covariant only under a correspondingly restricted group of coordinate transformations. Hilbert kept the generally covariant field equations as fundamental field equations and only postulated a limitation of the physically admissible coordinate systems.

10 For a description of the physical appearance of the proofs, see (Sauer 1999, note 75).

11 See (Sauer 1999) and "Hilbert's Foundation of Physics …" (in this volume).

12 The choice of characters seems to have been motivated by alphabetical order. After denoting the generic "Hamiltonian function" as H, some invariant expression is denoted on page 4 as $J^{(h)}$. Later, on page 10, the electromagnetic field tensor is denoted by $M_{ks} = q_{sk} - q_{ks}$.

$$\frac{\partial\sqrt{g}K}{\partial g^{\mu\nu}}+\frac{\partial\sqrt{g}L}{\partial g^{\mu\nu}}-\sum_{k}\frac{\partial}{\partial w_k}\frac{\partial\sqrt{g}K}{\partial g_k^{\mu\nu}}+\sum_{k,l}\frac{\partial^2}{\partial w_k\partial w_l}\frac{\partial\sqrt{g}K}{\partial g_{kl}^{\mu\nu}} = 0. \tag{5}$$

Assuming that the missing piece introduced the quantities K and L by specifying H as some function of these quantities, $H = H(K, L)$, and taking into account that $L = L(g^{\mu\nu}, q_s, q_{sk})$ was assumed not to depend on $g_k^{\mu\nu}$ and $g_{kl}^{\mu\nu}$ we conclude that, in all probability, eq. (17-pr) must have been of the form:

$$H = \zeta(K+L) \tag{6}$$

with some constant ζ that may well have been set equal to 1. Clearly, Eq. (27-pr) is consistent with this conclusion. We also note that later in the text the quantities K and L are referred to as "invariants" (L on page 9 and on page 10, K on page 11).

Taking together these bits of information from the text of the proofs, we can draw the following preliminary conclusions about the content of the missing piece:

1. It must have contained an equation of the form (6) that was given the number (17-pr).
2. The missing piece introduced a quantity K in such a way that the definition or characterization of K, whatever it was, implied that $K = K(g^{\mu\nu}, g_l^{\mu\nu}, g_{kl}^{\mu\nu})$ is an invariant and only depends on the components of the metric and its first and second derivatives.
3. The missing piece introduced a quantity L in such a way that the definition or characterization of L, whatever it was, implied that $L = L(q_s, q_{sl}, g^{\mu\nu}, g_k^{\mu\nu})$ is an invariant and depends on the components of the electromagnetic vector potential and its first derivatives as well as on the metric components and its first derivatives.

It should be noted that these conclusions emerge from looking at the existing text of the proofs alone, without taking recourse to the published version or any other historical source.

4. WHAT IS CONTAINED IN THE PUBLISHED VERSION

Let us now take further account of Hilbert's published version of his *First Communication* (Hilbert 1915). As was indicated above, the published version differs significantly from the proofs in several respects, the main difference being a completely revised discussion of the energy theorem. Specifically, with respect to the gravitational and electrodynamical field equations, however, the differences are not significant, as we will see, apart from the fact that the explicit evaluation of the variational derivative of the gravitational part of the Lagrangian K is found only in the published version and not in the existing part of the proofs. Whether it may have been on the missing part of the proofs will be discussed below.

The formulation of the first two axioms is the same, and in the published version, Hilbert again wrote down the fundamental equations (4-pr), and (5-pr), albeit in a slightly different form as

$$\frac{\partial \sqrt{g}H}{\partial g^{\mu\nu}} - \sum_k \frac{\partial}{\partial w_k}\frac{\partial \sqrt{g}H}{\partial g_k^{\mu\nu}} + \sum_{k,l}\frac{\partial^2}{\partial w_k \partial w_l}\frac{\partial \sqrt{g}H}{\partial g_{kl}^{\mu\nu}} = 0, \tag{4-pu}$$

and

$$\frac{\partial \sqrt{g}H}{\partial q_h} - \sum_k \frac{\partial}{\partial w_k}\frac{\partial \sqrt{g}H}{\partial q_{hk}} = 0. \tag{5-pu}$$

The equivalence of eqs. (4-pr) and (5-pr), with (4-pu) and (5-pu) is, of course, completely trivial but the form (4-pu), (5-pu) allowed Hilbert to introduce the abbreviated notation $[\sqrt{g}H]_{\mu\nu}$ and $[\sqrt{g}H]_h$ already at this point as the left hand sides of the "fundamental equations" (4-pu) and (5-pu).

The specification of the Lagrangian H in terms of a gravitational part K and an electromagnetic part L appears twice in the published version. The first time the relevant equation appears it is in a context that would fit quite naturally into the missing piece of page 8 of the proofs. The relevant passage reads:

> As far as the world function H is concerned, further axioms are needed to determine its choice in a unique way. If the gravitational field equations are to contain only second derivatives of the potentials $g^{\mu\nu}$, then H must have the form
>
> $$H = K + L \tag{7}$$
>
> where K is the invariant that derives from the Riemannian tensor (curvature of the four-dimensional manifold)
>
> $$K = \sum_{\mu\nu} g^{\mu\nu}K_{\mu\nu} \tag{8}$$
>
> $$K_{\mu\nu} = \sum_{\kappa}\left(\frac{\partial}{\partial w_\nu}\begin{Bmatrix}\mu\kappa\\ \kappa\end{Bmatrix} - \frac{\partial}{\partial w_\kappa}\begin{Bmatrix}\mu\nu\\ \kappa\end{Bmatrix}\right) + \sum_{k,\lambda}\left(\begin{Bmatrix}\mu\kappa\\ \lambda\end{Bmatrix}\begin{Bmatrix}\lambda\nu\\ \kappa\end{Bmatrix} - \begin{Bmatrix}\mu\nu\\ \lambda\end{Bmatrix}\begin{Bmatrix}\lambda\kappa\\ \kappa\end{Bmatrix}\right) \tag{9}$$
>
> and where L only depends on $g^{\mu\nu}, g_l^{\mu\nu}, q_s, q_{sk}$. (Hilbert 1915, 402)

Hilbert then adds the following sentence: "Finally, we will, in the following, make the simplifying assumption that L does not depend on $g_l^{\mu\nu}$." The physical size of the missing piece allows for some ten lines of text or the equivalent of some smaller number of lines of text plus a number of displayed equations, taking into account that a displayed equation would take up more than a single line of text.[13] In view of this restriction, the passage in the published version is clearly too long to be inserted into

13 See (Sauer 1999, note 75), the length of the type area seems to vary slightly over the different pages of the proofs.

the missing piece of the proofs. However, we can easily cut down the passage to fit into the size of the missing piece as, e.g., with the following German sentence:

> Wir machen im folgenden den Ansatz
>
> $$H = K + L \tag{10}$$
>
> wo K die aus dem Riemannschen Tensor entspringende Invariante
>
> $$K = \sum_{\mu\nu} g^{\mu\nu} K_{\mu\nu} \tag{11}$$
>
> $$K_{\mu\nu} = \sum_{\kappa}\left(\frac{\partial}{\partial w_\nu}\begin{Bmatrix}\mu\kappa\\ \kappa\end{Bmatrix} - \frac{\partial}{\partial w_\kappa}\begin{Bmatrix}\mu\nu\\ \kappa\end{Bmatrix}\right) + \sum_{k,\lambda}\left(\begin{Bmatrix}\mu\kappa\\ \lambda\end{Bmatrix}\begin{Bmatrix}\lambda\nu\\ \kappa\end{Bmatrix} - \begin{Bmatrix}\mu\nu\\ \lambda\end{Bmatrix}\begin{Bmatrix}\lambda\kappa\\ \kappa\end{Bmatrix}\right) \tag{12}$$
>
> bedeutet und L nur von $g^{\mu\nu}, g_l^{\mu\nu}, q_s, q_{sk}$ abhängt.[14]

It seems perfectly natural to assume that this passage or some very similar variant of it was the missing piece on page 8 of the proofs. And, as already conjectured in (Sauer 1999, note 82), Hilbert himself may have cut out this piece from his proofs, perhaps to paste it into some other unknown manuscript of his, e.g. into the manuscript for his revised version.

As indicated above, the equation $H = K + L$ appears at one other place in the published version of Hilbert's *First Communication*. This passage reads:

> It remains to show directly how with the assumption
>
> $$H = K + L \tag{20-pu}$$
>
> the generalized Maxwell equations (5-pu) put forth above are entailed by the gravitational equations (4-pu).
>
> Using the notation introduced earlier for the variational derivatives with respect to the $g^{\mu\nu}$, the gravitational equations, because of (20-pu), take the form
>
> $$[\sqrt{g}K]_{\mu\nu} + \frac{\partial\sqrt{g}L}{\partial g^{\mu\nu}} = 0. \tag{21-pu}$$
>
> The first term on the left hand side becomes
>
> $$[\sqrt{g}K]_{\mu\nu} = \sqrt{g}\left(K_{\mu\nu} - \frac{1}{2}Kg_{\mu\nu}\right), \tag{13}$$
>
> as follows easily without calculation from the fact that $K_{\mu\nu}$, apart from $g_{\mu\nu}$, is the only tensor of second rank ("Ordnung") and K the only invariant, that can be formed using only the $g^{\mu\nu}$ and their first and second differential quotients $g_k^{\mu\nu}, g_{kl}^{\mu\nu}$. (Hilbert 1915, 404 f.)

14 "We now make the ansatz [...] where K is the invariant that derives from the Riemannian tensor [...] and where L only depends on $g^{\mu\nu}, g_l^{\mu\nu}, q_s, q_{sk}$."

And after this assertion, Hilbert adds the following comment as to the apparent equivalence of his equations to those published by Einstein:

> The resulting differential equations of gravitation are, it seems to me, in agreement with the broad ("großzügigen") theory of general relativity established by Einstein in his later papers.

The reference to Einstein's "later papers" is specified in a footnote by citing all four of Einstein's November memoirs (Einstein 1915a, b, c, d) including the last one that was presented to the Berlin Academy only on 25 November (Einstein 1915d). The question arises whether the missing piece of the proofs could have contained equation (13), i.e. the explicit form of the variational derivative for some gravitational Lagrangian K. Specifically under the assumption that K was defined or characterized as the Riemannian curvature scalar, it would then have displayed what we now call the Einstein tensor with its trace term $-\frac{1}{2}Kg_{\mu\nu}$. This reading would allow revival of a speculation that a version of the theory as laid out in the proofs may then possibly have inspired Einstein to make the transition of his field equations of his second November memoir of 11 November 1915 (Einstein 1915b) to those of his final November paper of 25 November 1915 (Einstein 1915d) by adding a similar trace term to the matter term of his previous equation.

However, from the internal logic and structure of both the argument in the proofs and in the published version, this conjecture seems highly unlikely for the following reasons. In addition to equation (13) or some similar equation displaying the explicit form of the variational derivative of the gravitational part of the Lagrangian, the missing piece must still have contained an equation of the form (6), as in (20-pu), and some kind of characterization of the quantities K and L as discussed above on the basis of the proofs alone. In addition, it must also have contained some kind of characterization of the term $K_{\mu\nu}$ which appears in equation (13) but which had not appeared in the proofs before. In view of the physical size of the missing piece, the explicit form of the Ricci tensor $K_{\mu\nu}$, as in (12), could hardly have fitted on it in addition to equation (20-pu), as well as equation (13). Therefore, the quantity K must then have been defined or characterized without using its explicit form, maybe only with words ("die aus dem Riemannschen Krümmungstensor $K_{\mu\nu}$ entspringende Invariante K").

However, there are at least two arguments against the assumption that the missing piece contained equation (13) in addition to equation (20-pu) and some minimal information needed to introduce K and L.

1. Nowhere in the extant parts of the proofs does Hilbert calculate explicitly the result of the variational derivative or argues on this level. Indeed, in and of itself such an explicit calculation would be at odds with the general thrust of his communication which is to draw quite general conclusions from combining variational calculus and invariant theory. And in the published version, the explicit form of the variational derivative of the gravitational part of the Lagrangian is clearly directly motivated by Hilbert's comment on the presumed equivalence of

his own equations with those of Einstein's November memoirs, specifically as it seems with the final ones of 25 November 1915.

2. The mathematical assertion captured by equation (13), i.e. the assertion that the Einstein tensor $K_{\mu\nu} - \frac{1}{2}Kg_{\mu\nu}$ is obtained by a variation of the Riemann curvature scalar K with respect to the metric $g^{\mu\nu}$, must have been given with even less comments on how this result is obtained and on what assumptions are needed for its validity, as were given in the published version.

To elaborate on the second point, let me finally comment on the derivation of the Einstein tensor from a variation of the Riemann curvature. As pointed out in (Corry, Renn and Stachel 1997), the fact that Hilbert's assertion quoted above about the uniqueness of the Einstein tensor, if taken literally, is wrong, since there are many invariants that are of second rank and "can be formed using only the $g_{\mu\nu}$ and their first and second differential quotients." However, earlier on, Hilbert had also mentioned the condition that second derivatives are to be contained in the gravitational equations only linearly. This additional condition fixes the tensor to the form $K_{\mu\nu} - \alpha K g_{\mu\nu}$ with some undetermined factor α. This factor α is determined to be equal to 1/2 if it is further assumed that the covariant divergence of the expression vanishes, an assumption that is never mentioned explicitly in the published version, although it is implied by the contracted Bianchi identities that follow from Hilbert's proto-version of Noether's second theorem in his published communication (Sauer 1999, note 104 and p. 564; Logunov, Mestvirishvili and Petrov 2004). Corry, Renn and Stachel (1997) also point out that, while Hilbert asserts that the result follows "without calculation," he does give a more explicit derivation of the Einstein tensor in his 1924 republication of his *Communications* on the *Foundations of Physics* (Hilbert 1924).[15] Nevertheless, we have contemporary evidence that may give a meaning to Hilbert's assertion. It is found in a letter by the mathematician Hermann Vermeil to Felix Klein, dated 2 February 1918.[16] In it Vermeil explicitly addressed the question how the result can be obtained "without calculation." The answer that he found goes like this:

Assuming that

$$[\sqrt{g}K]_{\mu\nu} \propto \sqrt{g}(K_{\mu\nu} - \alpha K g_{\mu\nu}) \tag{14}$$

---

15 I disagree with the claim in (Corry, Renn and Stachel 1997) that the 1924 republication was primarily motivated by Hilbert's wish to correct some errors of his 1915 publication. As argued elsewhere (Majer and Sauer 2005), it was on the contrary Hilbert's intention to reaffirm his own priority of the field equations after Einstein in his 1923 papers on Eddington's unified field theory had arrived at equations that were essentially equivalent to the gravitational field equations of 1915 in variational form in the context of the unified field theory program.

16 NSUB Cod. Ms. Klein 22B, f. 28. This letter was discussed extensively at a history of mathematics conference at Oberwolfach in May 2000 in which the Einstein-Hilbert competition was a central topic of discussion. The argument is also presented, apparently without knowledge of Vermeil's letter, in (Logunov, Mestvirishvili and Petrov 2004, 611). For Vermeil's role, see also the discussion in (Rowe 2001, 417f.).

which, as discussed, follows from Hilbert's assumptions if one also demands that second derivatives occur only linearly, Vermeil evaluated $[\sqrt{g}K]_{\mu\nu}$ (see (4) for the scalar $K = \sum_{\rho\sigma} g^{\rho\sigma} K_{\rho\sigma,}$ see (8), and obtained

$$[\sqrt{g}K]_{\mu\nu} = K\frac{\partial\sqrt{g}}{\partial g^{\mu\nu}} + \sqrt{g}K_{\mu\nu} + \sqrt{g}g^{\rho\sigma}\frac{\partial K_{\rho\sigma}}{\partial g^{\mu\nu}} - \sum_k \frac{\partial}{\partial w_k}\frac{\partial\sqrt{g}K}{\partial g_k^{\mu\nu}} + \sum_{k,l}\frac{\partial^2}{\partial w_k \partial w_l}\frac{\partial\sqrt{g}K}{\partial g_{kl}^{\mu\nu}}. \tag{15}$$

Using $dg = -g g_{\mu\nu} dg^{\mu\nu}$, this turns into

$$[\sqrt{g}K]_{\mu\nu} = \sqrt{g}\left(K_{\mu\nu} - \frac{1}{2}K g_{\mu\nu}\right) + \sqrt{g}g^{\rho\sigma}\frac{\partial K_{\rho\sigma}}{\partial g^{\mu\nu}} - \sum_k \frac{\partial}{\partial w_k}\frac{\partial\sqrt{g}K}{\partial g_k^{\mu\nu}} + \sum_{k,l}\frac{\partial^2}{\partial w_k \partial w_l}\frac{\partial\sqrt{g}K}{\partial g_{kl}^{\mu\nu}}. \tag{16}$$

where all terms on the second line do not produce terms of the form (14).

While this derivation shows that Hilbert's claim in the published version about the derivation of the Einstein tensor is correct (granting that the postulate that second derivatives occur only linearly was implied) and credible, the question still remains as to why Hilbert should have done this derivation and included its result into the proofs without elaborating at all about the necessary steps and assumptions. Assuming that Hilbert added the explicit evaluation of $[\sqrt{g}K]_{\mu\nu}$ into the published version after seeing the explicit field equations of Einstein's final November paper, on the other hand, makes good sense. Let us not forget after all, that Hilbert in this context does cite Einstein's paper of 25 November.

5. CONCLUDING REMARKS

What was on the excised piece? Merely requiring continuity with the remaining text constrains the possibilities quite considerably. It is highly unlikely that the missing part contained the explicit result of a variational derivative of the action with respect to the metric and specifically some version of the Einstein tensor. Consistency with the remaining text rather leads virtually uniquely to the conclusion that on the missing piece Hilbert had specified the Lagrangian of his variational principle as a sum of a gravitational part and a matter part, that he had further specified the gravitational part as the Riemann curvature scalar, and that he did so by giving the Ricci tensor in its explicit form.

It still remains true that the proofs of Hilbert's *First Communication* on the *Foundations of Physics* already contain the correct gravitational field equations of general relativity in implicit form, i.e. in terms of a variational principle and the Hilbert action. The variational formulation is fully equivalent to the explicit Einstein equations published by Einstein a few days later, although the theory of Hilbert's proofs was not yet a fully generally covariant theory. It remains an interesting task to spell out in detail a scenario by which Hilbert would have overcome the restriction implied by the third axiom of his proofs following his own heuristics and logic of discovery.

REFERENCES

Corry, Leo. 2004. *David Hilbert and the Axiomatization of Physics. From 'Grundlagen der Geometrie' to 'Grundlagen der Physik'*. Dordrecht/Boston/London: Kluwer.

Corry, Leo, Jürgen Renn and John Stachel. 1997. "Belated Decision in the Hilbert-Einstein Priority Dispute." *Science* 278: 1270–1273.

Earman, J., and C. Glymour. 1978. "Einstein and Hilbert. Two Months in the History of General Relativity." *Archive for History of Exact Sciences* 19: 291–308.

Einstein, Albert. 1915a. "Zur allgemeinen Relativitätstheorie." *Königlich Preußische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte*: 778–786, (CPAE 6, Doc. 21).

———. 1915b. "Zur allgemeinen Relativitätstheorie. (Nachtrag)." *Königlich Preußische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte*: 799–801, (CPAE 6, Doc. 22).

———. 1915c. "Erklärung der Perihelbewegung des Merkur aus der allgemeinen Relativitätstheorie." *Sitzungsberichte der Preußischen Akademie der Wissenschaften: 2. Halbband* XLVII: 831-839, (CPAE 6, Doc. 24).

———. 1915d. "Die Feldgleichungen der Gravitation." *Königlich Preußische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte*: 844–847, (CPAE 6, Doc. 25).

Einstein, Albert, and Marcel Grossmann. 1913. *Entwurf einer verallgemeinerten Relativitätstheorie und einer Theorie der Gravitation*. Leipzig/Berlin: Teubner, (CPAE 4, Doc. 13).

Hilbert, David. 1915. "Die Grundlagen der Physik. (Erste Mitteilung.)" *Königliche Gesellschaft der Wissenschaften zu Göttingen. Mathematisch-physikalische Klasse. Nachrichten*, 395–407. (English translation in this volume.)

———. 1924. "Die Grundlagen der Physik." *Mathematische Annalen* 92: 1–32.

Logunov, A.A., M.A. Mestvirishvili und V.A. Petrov. 2004. "How Were the Hilbert-Einstein Equations Discovered?" *Physics-Uspekhi* 47 (2004) 607–621.

Majer, Ulrich and Sauer, Tilman. 2005. "Hilbert's World Equations and His Vision of a Unified Science." In A. J. Kox and J. Eisenstaedt (eds.) *The Universe of General Relativity*, 259–276.(*Einstein Studies*, vol. 11). Boston/Basel/Berlin: Birkhäuser.

Mehra, Jagdish. 1974. *Einstein, Hilbert, and The Theory of Gravitation*. Dordrecht/Boston: D. Reidel.

Norton, John. 1984. "How Einstein Found His Field Equations: 1912–1915." *Historical Studies in the Physical Sciences* 14 (1984) 253–316. Reprinted in D. Howard and J. Stachel (eds.) *Einstein and the History of General Relativity*. Boston: Birkhäuser, 101–159.

Pais, Abraham. 1982. *'Subtle is the Lord ...' The Science and the Life of Albert Einstein*. Oxford and New York: Clarendon Press and Oxford University Press.

Rowe, David. 2001. "Einstein Meets Hilbert: At the Crossroads of Physics and Mathematics." *Physics in Perspective* 3: 379–424.

Sauer, Tilman. 1999. "The Relativity of Discovery. Hilbert's First Note on the Foundations of Physics." *Archive for History of Exact Sciences* 53: 529–575.

Stachel, John. 1999. "New Light on the Einstein-Hilbert Priority Question." *Journal of Astrophysics and Astronomy* 20: 91-101. Reprinted in: Stachel, John. *Einstein from 'B' to 'Z'*. Boston/Basel/Berlin: Birkhäuser, 2002, 353–364.

Vizgin, V. P. 2001. "On the discovery of the gravitational field equations by Einstein and Hilbert: new materials." *Physics-Uspekhi* 44: 1283–1298.

First proof of my first note.

The Foundations of Physics.

(First communication.)[1]

by

David Hilbert.

Presented in the session of 20 November 1915.

The far reaching ideas and the formation of novel concepts by means of which Mie constructs his electrodynamics, and the prodigious problems raised by Einstein, as well as his ingeniously conceived methods of solution, have opened new paths for the investigation into the foundations of physics.

In the following — in the sense of the axiomatic method — I would like to develop /*essentially* from three simple axioms a ~~new~~ system of basic equations of physics, of ideal beauty, containing, I believe, the solution of the problems presented. I reserve for later communications the detailed development and particularly the special application of my basic equations to the fundamental questions of the theory of electricity.

Let w_s $(s = 1, 2, 3, 4)$ be any coordinates labeling the world's points essentially uniquely — the so-called world parameters. The quantities characterizing the events at w_s shall be:

1) The ten gravitational potentials /*first introduced by Einstein* $g_{\mu\nu}$ $(\mu, \nu = 1, 2, 3, 4)$ having the character of a symmetric tensor with respect to arbitrary transformation of the world parameter w_s;

2) The four electrodynamic potentials q_s having the character of a vector in the same sense.

Physical processes do not proceed in an arbitrary way, rather they are governed by the following two axioms: |

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 4
Gravitation in the Twilight of Classical Physics: The Promise of Mathematics.

[2] Axiom I (Mie's[1] axiom of the world function): *The law governing physical processes is determined through a world function H, that contains the following arguments*:

$$g_{\mu\nu}, \qquad g_{\mu\nu l} = \frac{\partial g_{\mu\nu}}{\partial w_l}, \qquad g_{\mu\nu lk} = \frac{\partial^2 g_{\mu\nu}}{\partial w_l \partial w_k}, \tag{1}$$

$$q_s, \qquad q_{sl} = \frac{\partial q_s}{\partial w_l} \qquad (l, k = 1, 2, 3, 4) \tag{2}$$

where the variation of the integral

$$\int H \sqrt{g} d\tau;$$

$$(g = |g_{\mu\nu}|, \quad d\tau = dw_1 dw_2 dw_3 dw_4)$$

must vanish for each of the 14 potentials $g_{\mu\nu}, q_s$.

Clearly the arguments (1) can be replaced by the arguments

$$g^{\mu\nu}, \qquad g_l^{\mu\nu} = \frac{\partial g^{\mu\nu}}{\partial w_l}, \qquad g_{lk}^{\mu\nu} = \frac{\partial^2 g^{\mu\nu}}{\partial w_l \partial w_k} \tag{3}$$

where $g^{\mu\nu}$ is the subdeterminant of the determinant g with respect to its element $g_{\mu\nu}$, divided by g.

Axiom II[2] (axiom of general invariance): *The world function H is invariant with respect to an arbitrary transformation of the world parameters w_s.*

Axiom II is the simplest mathematical expression of the demand that the interlinking of the potentials $g_{\mu\nu}, q_s$ is by itself entirely independent of the way one chooses to identify the world's points by means of world parameters.

The guiding motive for setting up ~~my~~ *the* theory is given by the following theorem, the proof of which I shall present elsewhere.

Theorem I. If J is an invariant under arbitrary transformations of the four world

[3] parameters, containing n quantities and their derivatives, | and if one forms from

$$\delta \int J \sqrt{g} d\tau = 0$$

1 Mie's world functions do not contain exactly these arguments; in particular the usage of the arguments (2) goes back to Born. However, what is characteristic of Mie's electrodynamics is the introduction and use of such a world function in Hamilton's principle.

2 Orthogonal invariance was already postulated by Mie. In the axiom II established above, Einstein's basic idea fundamental[2] of general covariance finds its simplest expression, even if Hamilton's principle plays only a subsidiary role with Einstein, and his functions H are by no means invariants, and also do not contain the electric potentials.

the n variational equations of Lagrange with respect to each of the n quantities, then in this invariant system of n differential equations for the n quantities there are always four that are a consequence of the remaining $n-4$ — in the sense that, among the n differential equations and their total derivatives there are always four linear and mutually independent combinations that are satisfied identically.

Concerning the differential quotients with respect to $g^{\mu\nu}$, $g_k^{\mu\nu}$, $g_{kl}^{\mu\nu}$ as in (4) and subsequent formulas, let us note once for all that, due to the symmetry in μ, ν on the one hand and in k, l on the other, the differential quotients with respect to $g^{\mu\nu}$, $g_k^{\mu\nu}$ are to be multiplied by 1 resp. $\frac{1}{2}$, according as $\mu = \nu$ resp. $\mu \neq \nu$, further the differential quotients with respect to $g_{kl}^{\mu\nu}$ are to be multiplied by 1 resp. $\frac{1}{2}$ resp. $\frac{1}{4}$, according as $\mu = \nu$ and $k = l$ resp. $\mu = \nu$ and $k \neq l$ or $\mu \neq \nu$ and $k = l$ resp. $\mu \neq \nu$ and $k \neq l$.

Axiom I implies first for the ten gravitational potentials $g^{\mu\nu}$ the ten Lagrangian differential equations

$$\frac{\partial\sqrt{g}H}{\partial g^{\mu\nu}} = \sum_k \frac{\partial}{\partial w_k}\frac{\partial\sqrt{g}H}{\partial g_k^{\mu\nu}} - \sum_{k,l}\frac{\partial^2}{\partial w_k \partial w_l}\frac{\partial\sqrt{g}H}{\partial g_{kl}^{\mu\nu}}, \qquad (\mu, \nu = 1, 2, 3, 4) \qquad (4)$$

and secondly for the four electrodynamic potentials q_s the four Lagrangian differential equations

$$\frac{\partial\sqrt{g}H}{\partial q_h} = \sum_k \frac{\partial}{\partial w_k}\frac{\partial\sqrt{g}H}{\partial q_{hk}}, \qquad (h = 1, 2, 3, 4). \qquad (5)$$

Let us call equations (4) the fundamental equations of gravitation, and equations (5) the fundamental electrodynamic equations, or generalized Maxwell equations. Due to the theorem stated above, the four equations (5) can be viewed as a consequence of equations (4), that is, because of that mathematical theorem we can immediately assert the claim *that in the sense explained above electrodynamic phenomena are effects of gravitation*. I regard this insight as the simple and very surprising solution of the problem of Riemann, who was the first to search for a theoretical connection between gravitation and light.

Since our mathematical theorem shows us that the axioms I and II considered so far can produce only ten essentially independent equations; and since, on the other hand, | if general invariance is maintained, more than ten essentially independent [4] equations for the 14 potentials $g_{\mu\nu}, q_s$ are not at all possible; therefore—provided that we want to retain the determinate character of the basic equation of physics corresponding to Cauchy's theory of differential equations— the demand for four further non-invariant equations in addition to (4) and (5) is imperative. In order to arrive at these equations, I first put up a definition of the concept of energy.

To this end we polarize $g^{\mu\nu}$ in the invariant H by an arbitrary contragredient tensor $h^{\mu\nu}$ and thus form the expression

$$J^{(h)} = \sum_{\mu, \nu} \frac{\partial H}{\partial g^{\mu\nu}} h^{\mu\nu} + \sum_{\mu, \nu, k} \frac{\partial H}{\partial g_k^{\mu\nu}} h_k^{\mu\nu} + \sum_{\mu, \nu, k, l} \frac{\partial H}{\partial g_{kl}^{\mu\nu}} h_{kl}^{\mu\nu},$$

where the abbreviation

$$h_k^{\mu\nu} = \frac{\partial h^{\mu\nu}}{\partial w_k}, \qquad h_{kl}^{\mu\nu} = \frac{\partial^2 h^{\mu\nu}}{\partial w_k \partial w_l}$$

has been used. Since polarization is an invariant process, $J^{(h)}$ is an invariant. Now we treat the expression $\sqrt{g}J^{(h)}$ in the same way as an integrand of a variational problem in the calculus of variations, when one wants to integrate by parts; thus we obtain the following identity:

$$\sqrt{g}J^{(h)} = -\sum_{\mu, \nu} H \frac{\partial \sqrt{g}}{\partial g^{\mu\nu}} h^{\mu\nu} + \sum_{\mu, \nu} [\sqrt{g}H]_{\mu\nu} h^{\mu\nu} + D^{(h)}, \tag{6}$$

where we have put

$$[\sqrt{g}H]_{\mu\nu} = \frac{\partial \sqrt{g}H}{\partial g^{\mu\nu}} - \sum_k \frac{\partial}{\partial w_k} \frac{\partial \sqrt{g}H}{\partial g_k^{\mu\nu}} + \sum_{k, l} \frac{\partial^2}{\partial w_k \partial w_l} \frac{\partial \sqrt{g}H}{\partial g_{kl}^{\mu\nu}}$$

and

$$D^{(h)} = \sum_{\mu, \nu, k} \frac{\partial}{\partial w_k}\left(\frac{\partial \sqrt{g}H}{\partial g_k^{\mu\nu}} h^{\mu\nu}\right) + \sum_{\mu, \nu, k, l} \frac{\partial}{\partial w_k}\left(\frac{\partial \sqrt{g}H}{\partial g_{kl}^{\mu\nu}} h_l^{\mu\nu}\right)$$
$$- \sum_{\mu, \nu, k, l} \frac{\partial}{\partial w_l}\left(h^{\mu\nu} \frac{\partial}{\partial w_k} \frac{\partial \sqrt{g}H}{\partial g_{kl}^{\mu\nu}}\right) \tag{7}$$

as abbreviations. The expression $[\sqrt{g}H]_{\mu\nu}$ is nothing but the Lagrangian variational derivative of $\sqrt{g}H$ with respect to $g_{\mu\nu}$, which yields the gravitational equations (4) when it is put equal to zero,

$$[\sqrt{g}H]_{\mu\nu} = 0 \tag{8}$$

[5] | and the expression $D^{(h)}$ is a sum of differential quotients, so it has the character of a pure divergence.

Now we use the easily proved fact that, if p^j $(j = 1, 2, 3, 4)$ is an arbitrary contravariant vector, the expression

$$p^{\mu\nu} = \sum_s (g_s^{\mu\nu} p^s - g^{\mu s} p_s^{\nu} - g^{\nu s} p_s^{\mu}), \qquad \left(p_s^j = \frac{\partial p^j}{\partial w_s}\right)$$

represents a symmetric contravariant tensor.

If we substitute in the invariant expression $J^{(h)}$ instead of $h^{\mu\nu}$ the special contravariant tensor $p^{\mu\nu}$, there arises again an invariant expression, namely

$$J^{(p)} = \sum_{\mu,\nu} \frac{\partial H}{\partial g^{\mu\nu}} p^{\mu\nu} + \sum_{\mu,\nu,k} \frac{\partial H}{\partial g_k^{\mu\nu}} p_k^{\mu\nu} + \sum_{\mu,\nu,k,l} \frac{\partial H}{\partial g_{kl}^{\mu\nu}} p_{kl}^{\mu\nu},$$

where the abbreviations

$$p_k^{\mu\nu} = \frac{\partial p^{\mu\nu}}{\partial w_k}, \qquad p_{kl}^{\mu\nu} = \frac{\partial^2 p^{\mu\nu}}{\partial w_k \partial w_l}$$

have been used. Now we treat the expression $\sqrt{g}J^{(p)}$ in the same way as an integrand of a variational problem in the calculus of variations, when one wants to integrate by parts — but in such a way that in this procedure the first differential quotient p_s^j of the p^j always remain unchanged, and only the second and third derivatives of the p^j are included in the divergence; and moreover so that the auxiliary expressions become invariant with respect to linear transformation

$$\begin{aligned} E = & \sum\left(H\frac{\partial\sqrt{g}}{\partial g^{\mu\nu}}g_s^{\mu\nu} + \sqrt{g}\frac{\partial H}{\partial g^{\mu\nu}}g_s^{\mu\nu} + \sqrt{g}\frac{\partial H}{\partial g_k^{\mu\nu}}g_{sk}^{\mu\nu} + \sqrt{g}\frac{\partial H}{\partial g_{kl}^{\mu\nu}}g_{skl}^{\mu\nu}\right)p_l^s \\ & -\sum(g^{\mu s}p_s^{\nu} + g^{\nu s}p_s^{\mu})[\sqrt{g}H]_{\mu\nu} \\ & +\sum\left(\frac{\partial\sqrt{g}H}{\partial g_k^{\mu\nu}}-g_s^{\mu\nu} + \frac{\partial\sqrt{g}H}{\partial g_{kl}^{\mu\nu}}-g_{sl}^{\mu\nu} - g_s^{\mu\nu}\frac{\partial}{\partial w_l}\frac{\partial\sqrt{g}H}{\partial g_{kl}^{\mu\nu}}\right)p_k^s; \end{aligned} \tag{9}$$

we thus obtain the following identity:

$$\sqrt{g}J^{(p)} = -\sum_{\mu,\nu} H\frac{\partial\sqrt{g}}{\partial g^{\mu\nu}}p^{\mu\nu} + E + D^{(p)}, \tag{10}$$

where we have put | [6]

$$\begin{aligned} D^{(p)} = \sum\Bigg\{ & -\frac{\partial}{\partial w_k}\left(\sqrt{g}\frac{\partial H}{\partial g_k^{\mu\nu}}(g^{\mu s}p_s^{\nu} + g^{\nu s}p_s^{\mu})\right) \\ & + \frac{\partial}{\partial w_k}\left((p_s^{\nu}g^{\mu s} + p_s^{\mu}g^{\nu s})\frac{\partial}{\partial w_l}\left(\sqrt{g}\frac{\partial H}{\partial g_{kl}^{\mu\nu}}\right)\right) \\ & + \frac{\partial}{\partial w_l}\left(\sqrt{g}\frac{\partial H}{\partial g_{kl}^{\mu\nu}}\left(\frac{\partial p^{\mu\nu}}{\partial w_k} - g_{sk}^{\mu\nu}p^s\right)\right)\Bigg\} \end{aligned}$$

as an abbreviation.[3] The expression E is invariant under linear transformation and with respect to the vector p^j it has the form

$$E = \sum_s e_s p^s + \sum_{s,l} e_s^l p_l^s,$$

where from (10) e_s and e_s^l are well-defined expressions. In particular it turns out, as one can see, that:

$$e_s = \frac{d^{(g)}\sqrt{g}H}{dw_s}; \tag{11}$$

where the differentiation denoted by $d^{(g)}$ is total with respect to w_s, but to be performed in such a way that the electromagnetic potentials q_s remain unaffected.

Call the expression E the energy form. To justify this designation, I prove two properties that the energy form enjoys.

If we substitute the tensor $p^{\mu\nu}$ for $h^{\mu\nu}$ in identity (6) then, together with (9) it follows, provided the gravitational equations (8) are satisfied:

$$E = (D^{(h)})_{h=p} - D^{(p)} \tag{12}$$

or

$$E = \sum\left\{\frac{\partial}{\partial w_k}\left(\sqrt{g}\frac{\partial H}{\partial g_k^{\mu\nu}} g_s^{\mu\nu} p^s\right) - \frac{\partial}{\partial w_k}\left(\frac{\partial}{\partial w_l}\left(\sqrt{g}\frac{\partial H}{\partial g_{kl}^{\mu\nu}}\right) g_s^{\mu\nu} p^s\right) + \frac{\partial}{\partial w_l}\left(\sqrt{g}\frac{\partial H}{\partial g_{kl}^{\mu\nu}} g_{sk}^{\mu\nu} p^s\right)\right\}, \tag{13}$$

that is, we have the proposition:

Proposition 1: In virtue of the gravitational equations the energy form E becomes a sum of differential quotients with respect to w_s, that is, it acquires the character of a divergence.

Had we gone a step further in the above treatment of the expression $\sqrt{g}J^{(p)}$, that led to (9) and converted in the usual way of the variational calculus also the first differential quotient p_s^j of the p^j then the expression containing the p^j alone would l

[eq (14) missing][4] (14)

[7] This theorem shows that the divergence equation corresponding to the energy theorem of the old theory

$$\sum_l \frac{\partial e_s^l}{\partial w_l} = 0 \tag{15}$$

holds if and only if the four quantities e_s vanish, that is if the following equations hold

$$\frac{d^{(g)}\sqrt{g}H}{dw_s} = 0. \tag{16}$$

After these preliminaries I now put down the following axiom:

Axiom III (axiom of space and time). *The spacetime coordinates are those special world parameters for which the energy theorem* (15) *is valid.*

According to this axiom, space and time in reality provide a special labeling of the world's points such that the energy theorem holds.

Axiom III implies the existence of equations (16): these four differential equations (16) complete the gravitational equations (4) to give a system of 14 equations for the 14 potentials $g_{\mu\nu}, q_s$, *the system of fundamental equations of physics*. Because of the agreement in number between equations and potentials to be determined, the principle of causality for physical processes is also guaranteed, revealing to us the closest connection between the energy theorem and the principle of causality, since each presupposes the other. To the transition from one spacetime reference system to another one corresponds the transformation of the energy form from one so-called "normal form"

$$E = \sum_{s,l} e_s^l p_l^s$$

to another normal form. |

$$\text{[eq. (17) missing: } H = K + L.\text{]}^{[4]} \tag{17}$$

Because K depends only on $g^{\mu\nu}$, $g_s^{\mu\nu}$, $g_{kl}^{\mu\nu}$, therefore in ansatz (17), due to (13), [8]
the energy E can be expressed solely as a function of the gravitational potentials $g^{\mu\nu}$ and their derivatives, provided L is assumed to depend not on $g_s^{\mu\nu}$, but only on $g^{\mu\nu}$, q_s, q_{sk}. On this assumption, which we shall always make in the following, the definition of the energy (10) yields the expression

$$E = E^{(g)} + E^{(e)}, \tag{18}$$

where the "gravitational energy" $E^{(g)}$ depends only on $g^{\mu\nu}$ and their derivatives, and the "electrodynamic energy" $E^{(e)}$ takes the form

$$E^{(e)} = \sum_{\mu,\nu,s} \frac{\partial\sqrt{g}L}{\partial g^{\mu\nu}}(g_s^{\mu\nu}p^s - g^{\mu s}p_s^{\nu} - g^{\nu s}p_s^{\mu}), \tag{19}$$

which proves to be *a general invariant multiplied by* $\sqrt{g}$.

To proceed we use two mathematical theorems, which say the following:

Theorem II. If J is an invariant depending on $g^{\mu\nu}$, $g_l^{\mu\nu}$, $g_{kl}^{\mu\nu}$, q_s, q_{sk}, then the following is always identically true in all arguments and for every arbitrary contravariant vector p^s:

$$\sum_{\mu,\nu,l,k}\left(\frac{\partial J}{\partial g^{\mu\nu}}\Delta g^{\mu\nu}+\frac{\partial J}{\partial g_l^{\mu\nu}}\Delta g_l^{\mu\nu}+\frac{\partial J}{\partial g_{kl}^{\mu\nu}}\Delta g_{kl}^{\mu\nu}\right)$$

$$+\sum_{s,k}\left(\frac{\partial J}{\partial q_s}\Delta q_s+\frac{\partial J}{\partial q_{sk}}\Delta q_{sk}\right) = 0;$$

where |

$$\Delta g^{\mu\nu} = \sum_m (g^{\mu m}p_m^{\nu}+g^{\nu m}p_m^{\mu}),$$

$$\Delta g_l^{\mu\nu} = -\sum_m g_m^{\mu\nu}p_l^m+\frac{\partial \Delta g^{\mu\nu}}{\partial w_l},$$

[9] $$\Delta g_{lk}^{\mu\nu} = -\sum_m (g_m^{\mu\nu}p_{lk}^m+g_{lm}^{\mu\nu}p_k^m+g_{km}^{\mu\nu}p_l^m)+\frac{\partial^2 \Delta g^{\mu\nu}}{\partial w_l \partial w_k},$$

$$\Delta q_s = -\sum_m q_m p_s^m,$$

$$\Delta q_{sk} = -\sum_m q_{sm}p_k^m+\frac{\partial \Delta q_s}{\partial w_k}.$$

Theorem III. If J is an invariant depending only on the $g^{\mu\nu}$ and their derivatives and if, as above, the variational derivatives of $\sqrt{g}J$ with respect to $g^{\mu\nu}$ are denoted by $[\sqrt{g}J]_{\mu\nu}$ then the expression — in which $h^{\mu\nu}$ is understood to be any contravariant tensor —

$$\frac{1}{\sqrt{g}}\sum_{\mu,\nu}[\sqrt{g}J]_{\mu\nu}h^{\mu\nu}$$

represents an invariant; if in this sum we substitute in place of $h^{\mu\nu}$ the particular tensor $p^{\mu\nu}$ and write

$$\sum_{\mu,\nu}[\sqrt{g}J]_{\mu\nu}p^{\mu\nu} = \sum_{s,l}(i_s p^s+i_s^l p_l^s),$$

where then the expressions

$$i_s = \sum_{\mu,\nu}[\sqrt{g}J]_{\mu\nu}g_s^{\mu\nu},$$

$$i_s^l = -2\sum_{\mu}[\sqrt{g}J]_{\mu s}g^{\mu l}$$

depend only on the $g^{\mu\nu}$ and their derivatives, then we have

$$i_s = \sum_l \frac{\partial i_s^l}{\partial w_l} \tag{20}$$

in the sense, that this equation is identically fulfilled for all arguments, that is for the $g^{\mu\nu}$ and their derivatives.

Now we apply Theorem II to the invariant L and obtain

$$\sum_{\mu,\nu,m}\frac{\partial L}{\partial g^{\mu\nu}}(g^{\mu m}p^{\nu}_{m}+g^{\nu m}p^{\mu}_{m})-\sum_{s,m}\frac{\partial L}{\partial q_s}q_m p^{m}_{s} \tag{21}$$

$$-\sum_{s,k,m}\frac{\partial L}{\partial q_{sk}}(q_{sm}p^{m}_{k}+q_{mk}p^{m}_{s}+q_m p^{m}_{sk}) = 0.$$

Equating to zero the coefficient of p^{m}_{sk} produces the equation

$$\left(\frac{\partial L}{\partial q_{sk}}+\frac{\partial L}{\partial q_{ks}}\right)q_m = 0$$

| or [10]

$$\frac{\partial L}{\partial q_{sk}}+\frac{\partial L}{\partial q_{ks}} = 0, \tag{22}$$

that is, the derivatives of the electrodynamic potentials q_s occur only in the combinations

$$M_{ks} = q_{sk}-q_{ks}.$$

Thus we learn that under our assumptions the invariant L depends, other than on the potentials $g_{\mu\nu}$, q_s, only on the components of the skew symmetric invariant tensor

$$M = (M_{ks}) = \mathrm{rot}(q_s),$$

that is, of the so-called electromagnetic six vector. *This result here derives essentially as a consequence of the general invariance, that is, on the basis of axiom II.*

If we put the coefficient of p^{ν}_{m} on the left of identity (21) equal to zero, we obtain, using (22)

$$2\sum_{\mu}\frac{\partial L}{\partial g^{\mu\nu}}g^{\mu m}-\frac{\partial L}{\partial q_m}q_\nu-\sum_{s}\frac{\partial L}{\partial M_{ms}}M_{\nu s} = 0, \qquad (\mu = 1, 2, 3, 4). \tag{23}$$

This equation admits an important transformation of the electromagnetic energy. Namely, the part of $E^{(e)}$ multiplied by p^{ν}_{m} in (19) becomes due to (23):

$$-2\sum_{\mu}\frac{\partial\sqrt{g}L}{\partial g^{\mu\nu}}g^{\mu m} = \sqrt{g}\left\{L\delta^{m}_{\nu}-\frac{\partial L}{\partial q_m}q_\nu-\sum_{s}\frac{\partial L}{\partial M_{ms}}M_{\nu s}\right\}, \tag{24}$$

$$(\mu = 1, 2, 3, 4) \quad (\delta^{\mu}_{\nu} = 0,\ \mu\neq\nu, \quad \delta^{\mu}_{\mu} = 1).$$

If one subjects the expression on the right to the limit

$$\begin{aligned} g_{\mu\nu} &= 0, \qquad (\mu \neq \nu) \\ g_{\mu\mu} &= 1, \end{aligned} \tag{25}$$

then this limit agrees exactly with what Mie has established in his electrodynamics: Mie's electromagnetic energy tensor is nothing but the generally invariant tensor that results from differentiation of the invariant L with respect to the gravitational potentials $g^{\mu\nu}$ in the limit (25) — a circumstance that gave me the first hint of the necessary close connection between Einstein's general relativity theory and Mie's electrodynamics, and which convinced me of the correctness of the theory here developed. |

[11] It only remains to show directly from assumption (17) how the generalized Maxwell equations (5) developed above are a consequence of the gravitational equations (4) in the sense given above.

By use of the notation just introduced for the variational derivatives with respect to the $g^{\mu\nu}$ the gravitational equations acquire the form, due to (17)

$$[\sqrt{g}K]_{\mu\nu} + \frac{\partial\sqrt{g}L}{\partial g^{\mu\nu}} = 0. \tag{26}$$

If we further denote in general the variational derivatives of $\sqrt{g}J$ with respect to the electrodynamic potential q_h by

$$[\sqrt{g}J]_h = \frac{\partial\sqrt{g}J}{\partial q_h} - \sum_k \frac{\partial}{\partial w_k}\frac{\partial\sqrt{g}J}{\partial q_{hk}},$$

then the electrodynamic equations take the form, due to (17)

$$[\sqrt{g}L]_h = 0. \tag{27}$$

Since K is an invariant that depends only on $g^{\mu\nu}$ and its derivatives, then by theorem III equation (20) holds identically with

$$i_s = \sum_{\mu,\nu}[\sqrt{g}K]_{\mu\nu}g_s^{\mu\nu} \tag{28}$$

and

$$i_s^l = -2\sum_{\mu}[\sqrt{g}K]_{\mu s}g^{\mu l}, \qquad (\mu = 1, 2, 3, 4). \tag{29}$$

Because of (26) and (29) the left side of (24) equals $-i_\nu^m$. By differentiation with respect to w_m and summation over m we obtain because of (20)

$$i_\nu = \sum_m \frac{\partial}{\partial w_m}\left(-\sqrt{g}L\delta_\nu^m + \frac{\partial\sqrt{g}L}{\partial q_m}q_\nu + \sum_s \frac{\partial\sqrt{g}L}{\partial M_{sm}}M_{s\nu}\right)$$

$$
\begin{aligned}
&= -\frac{\partial\sqrt{g}L}{\partial w_\nu} + \sum_m \left\{ q_\nu \frac{\partial}{\partial w_m}\left([\sqrt{g}L]_m + \sum_s \frac{\partial}{\partial w_s}\frac{\partial\sqrt{g}L}{\partial q_{ms}}\right)\right. \\
&\qquad \left. + q_{\nu m}\left([\sqrt{g}L]_m + \sum_s \frac{\partial}{\partial w_s}\frac{\partial\sqrt{g}L}{\partial q_{ms}}\right)\right\} \\
&\qquad + \sum_s \left([\sqrt{g}L]_s - \frac{\partial\sqrt{g}L}{\partial q_s}\right) M_{s\nu} + \sum_{s,m} \frac{\partial\sqrt{g}L}{\partial M_{sm}}\frac{\partial M_{s\nu}}{\partial w_m},
\end{aligned}
$$

since of course

$$\frac{\partial\sqrt{g}L}{\partial q_m} = [\sqrt{g}L]_m + \sum_s \frac{\partial}{\partial w_s}\frac{\partial\sqrt{g}L}{\partial q_{ms}}$$

| and [12]

$$-\sum_m \frac{\partial}{\partial w_m}\frac{\partial\sqrt{g}L}{\partial q_{sm}} = [\sqrt{g}L]_s - \frac{\partial\sqrt{g}L}{\partial q_s}.$$

Now we take into account that because of (22) we have

$$\sum_{m,s} \frac{\partial^2}{\partial w_m \partial w_s}\frac{\partial\sqrt{g}L}{\partial q_{ms}} = 0,$$

and thus obtain after suitably collecting terms

$$
\begin{aligned}
i_\nu &= -\frac{\partial\sqrt{g}L}{\partial w_\nu} + \sum_m \left(q_\nu \frac{\partial}{\partial w_m}[\sqrt{g}L]_m + M_{m\nu}[\sqrt{g}L]_m \right) \\
&\qquad + \sum_m \frac{\partial\sqrt{g}L}{\partial q_m} q_{m\nu} + \sum_{s,m} \frac{\partial\sqrt{g}L}{\partial M_{sm}}\frac{\partial M_{s\nu}}{\partial w_m}.
\end{aligned} \tag{30}
$$

On the other hand we have

$$-\frac{\partial\sqrt{g}L}{\partial w_\nu} = -\sum_{s,m} \frac{\partial\sqrt{g}L}{\partial g^{sm}} g_\nu^{sm} - \sum_m \frac{\partial\sqrt{g}L}{\partial q_m} q_{m\nu} - \sum_{m,s} \frac{\partial\sqrt{g}L}{\partial q_{ms}}\frac{\partial q_{ms}}{\partial w_\nu}.$$

Due to (26) and (28) the first terms on the right is nothing else but i_ν. The last term on the right proves to be equal and opposite to the last term on the right of (30); for we have

$$\sum_{s,m} \frac{\partial\sqrt{g}L}{\partial M_{sm}}\left(\frac{\partial M_{s\nu}}{\partial w_m} - \frac{\partial q_{ms}}{\partial w_\nu}\right) = 0, \tag{31}$$

since the expression

$$\frac{\partial M_{s\nu}}{\partial w_m} - \frac{\partial q_{ms}}{\partial w_\nu} = \frac{\partial^2 q_\nu}{\partial w_s \partial w_m} - \frac{\partial^2 q_s}{\partial w_\nu \partial w_m} - \frac{\partial^2 q_m}{\partial w_\nu \partial w_s}$$

turns out to be symmetric in s, m, and the first factor under the summation sign in (31) skew symmetric in s, m.

Therefore (30) implies the equations

$$\sum_m \left(M_{m\nu}[\sqrt{g}L]_m + q_\nu \frac{\partial}{\partial w_m}[\sqrt{g}L]_m \right) = 0; \tag{32}$$

that is, from the gravitational equations (4) there follow indeed the four linearly independent combinations (32) of the basic electrodynamic equations (5) and their first derivatives. *This is the entire mathematical expression of the general claim made above about the character of electrodynamics as an epiphenomenon of gravitation.* |

[13] According to our assumption L should not depend on the derivatives of the $g^{\mu\nu}$, therefore L must be a function of certain four general invariants, which correspond to the special orthogonal invariants reported by Mie, and of which the two simplest ones are these:

$$Q = \sum_{k, l, m, n} M_{mn} M_{lk} g^{mk} g^{nl}$$

and

$$q = \sum_{k, l} q_k q_l g^{kl}.$$

The simplest and most straightforward ansatz for L, considering the structure of K, is also that which corresponds to Mie's electrodynamics, namely

$$L = \alpha Q + f(q)$$

or, following Mie even more closely:

$$L = \alpha Q + \beta q^3,$$

where $f(q)$ denotes any function of q, and α, β are constants.

As one can see, the few simple assumptions expressed in axioms I, II, III suffice with appropriate interpretation to establish the theory: through it not only are our views of space, time, and motion fundamentally reshaped in the sense called for by Einstein, but I am also convinced that through the basic equations established here the most intimate, hitherto hidden processes in the interior of atoms will receive an explanation; and in particular that generally a reduction of all physical constants to mathematical constants must be possible—whereby the possibility approaches that physics in principle becomes a science of the type of geometry: surely the highest glory of the axiomatic method, which, as we have seen, here takes into its service the powerful instruments of analysis, namely the calculus of variations and the theory of invariants.

EDITORIAL NOTES

[1] The following is a translation of the proofs of Hilbert's first paper on the foundations of physics, which are preserved at Göttingen in SUB Cod. Ms. 634. These proofs comprise 13 pages and are complete, apart from the fact that roughly the upper quarter of two pages (7 and 8) is cut off. The proofs are dated "submitted on 20 November 1915." The Göttingen copy bears a printer's stamp dated 6 December 1915 and is marked in Hilbert's own hand "First proofs of my first note." In addition, the proofs carry several marginal notes in Hilbert's hand, which are shown here in superscript italics. In contrast to the other source papers in these volumes, this proof version of Hilbert's paper has been formatted as far as possible to recreate the original so that the author's hand-written notes are evident. This paper was later published in a converted version in *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen. Math.-phys. Klasse*. 1915. Issue 8, p 395–407, (1. correction).

[2] The word "fundamental" should appear before "basic". It is written correctly in the printed version.

[3] The superscript $^{\nu}$ in the first occurrence of p_s^{ν} in this equation is missing in the original.

[4] For more detailed information on the missing piece of this document, see "Einstein Equations and Hilbert Actions …" (in this volume).

DAVID HILBERT

THE FOUNDATIONS OF PHYSICS (FIRST COMMUNICATION)

Originally published as "Die Grundlagen der Physik. (Erste Mitteilung)" in Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen. Math.-phys. Klasse. 1916. Issue 8, p. 395–407. Presented in the session of 20 November 1915.

The vast problems posed by Einstein[1] as well as his ingeniously conceived methods of solution, and the far-reaching ideas and formation of novel concepts by means of which Mie[2] constructs his electrodynamics, have opened new paths for the investigation into the foundations of physics.

In the following—in the sense of the axiomatic method—I would like to develop, essentially from two simple axioms, a new system of basic equations of physics, of ideal beauty and containing, I believe, *simultaneously* the solution to the problems of Einstein and of Mie. I reserve for later communications the detailed development and particularly the special application of my basic equations to the fundamental questions of the theory of electricity.

Let w_s $(s = 1, 2, 3, 4)$ be any coordinates labeling the world's points essentially uniquely—the so-called world parameters (most general spacetime coordinates). The quantities characterizing the events at w_s shall be:

1. The ten gravitational potentials $g_{\mu\nu}$ $(\mu, \nu = 1, 2, 3, 4)$ first introduced by Einstein, having the character of a symmetric tensor with respect to an arbitrary transformation of the world parameters w_s;
2. The four electrodynamic potentials q_s having the character of a vector in the same sense. |

Physical processes do not proceed in an arbitrary way, rather they are governed by [396]
the following two axioms:

1 *Sitzungsber. d. Berliner Akad.* 1914, 1030; 1915, 778, 799, 831, 844.
2 *Ann. d. Phys.* 1912, Vol. 37, 511; Vol. 39, 1; 1913, vol. 40, 1.

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 4
Gravitation in the Twilight of Classical Physics: The Promise of Mathematics.

Axiom I (Mie's axiom of the world function[3]): *The law governing physical processes is determined through a world function* H, *that contains the following arguments:*

$$g_{\mu\nu}, \qquad g_{\mu\nu l} = \frac{\partial g_{\mu\nu}}{\partial w_l}, \qquad g_{\mu\nu lk} = \frac{\partial^2 g_{\mu\nu}}{\partial w_l \partial w_k}, \tag{1}$$

$$q_s, \qquad q_{sl} = \frac{\partial q_s}{\partial w_l} \qquad (l, k = 1, 2, 3, 4), \tag{2}$$

where the variation of the integral

$$\int H \sqrt{g} d\omega$$

$$(g = |g_{\mu\nu}|, \quad d\omega = dw_1 dw_2 dw_3 dw_4)$$

must vanish for each of the fourteen potentials $g_{\mu\nu}, q_s$.

Clearly the arguments (1) can be replaced by the arguments [1]

$$g^{\mu\nu}, \qquad g_l^{\mu\nu} = \frac{\partial g^{\mu\nu}}{\partial w_l}, \qquad g_{lk}^{\mu\nu} = \frac{\partial^2 g^{\mu\nu}}{\partial w_l \partial w_k}, \tag{3}$$

where $g^{\mu\nu}$ is the subdeterminant of the determinant g with respect to its element $g_{\mu\nu}$, divided by g.

Axiom II (axiom of general invariance[4]): *The world function* H *is invariant with respect to an arbitrary transformation of the world parameters* w_s.

Axiom II is the simplest mathematical expression of the demand that the interlinking of the potentials $g_{\mu\nu}, q_s$ is by itself entirely independent of the way one chooses to label the world's points by means of world parameters.

The guiding motive for constructing my theory is provided by the following theorem, the proof of which I shall present elsewhere. |

[397] Theorem I. If J is an invariant under arbitrary transformation of the four world parameters, containing n quantities and their derivatives, and if one forms from

$$\delta \int J g d\omega = 0$$

the n variational equations of Lagrange with respect to those n quantities, then in this invariant system of n differential equations for the n quantities there are always

3 Mie's world functions do not contain exactly these arguments; in particular the usage of the arguments (2) goes back to Born. However, what is characteristic of Mie's electrodynamics is precisely the introduction and use of such a world function in Hamilton's principle.

4 Orthogonal invariance was already postulated by Mie. In the axiom II formulated above, Einstein's fundamental basic idea of general invariance finds its simplest expression, even if Hamilton's principle plays only a subsidiary role with Einstein, and his functions H are by no means general invariants, and also do not contain the electric potentials.

four that are a consequence of the remaining $n-4$ —in this sense, that among the n differential equations and their total derivatives there are always four linear and mutually independent combinations that are satisfied identically.

Concerning the differential quotients with respect to $g^{\mu\nu}$, $g_k^{\mu\nu}$, $g_{kl}^{\mu\nu}$ occurring in (4) and subsequent formulas, let us note once for all that, due to the symmetry in μ, ν on the one hand and in k, l on the other, the differential quotients with respect to $g^{\mu\nu}$, $g_k^{\mu\nu}$ are to be multiplied by 1 resp. $\frac{1}{2}$, according as $\mu = \nu$ resp. $\mu \neq \nu$, further the differential quotients with respect to $g_{kl}^{\mu\nu}$ are to be multiplied by 1 resp. $\frac{1}{2}$ resp. $\frac{1}{4}$, according as $\mu = \nu$ and $k = l$ resp. $\mu = \nu$ and $k \neq l$ or $\mu \neq \nu$ and $k = l$ resp. $\mu \neq \nu$ and $k \neq l$.

Axiom I implies first for the ten gravitational potentials $g^{\mu\nu}$ the ten Lagrangian differential equations

$$\frac{\partial\sqrt{g}H}{\partial g^{\mu\nu}} - \sum_k \frac{\partial}{\partial w_k}\frac{\partial\sqrt{g}H}{\partial g_k^{\mu\nu}} + \sum_{k,l}\frac{\partial^2}{\partial w_k \partial w_l}\frac{\partial\sqrt{g}H}{\partial g_{kl}^{\mu\nu}} = 0, \qquad (\mu, \nu = 1, 2, 3, 4) \tag{4}$$

and secondly for the four electrodynamic potentials q_s the four Lagrangian differential equations

$$\frac{\partial\sqrt{g}H}{\partial q_h} - \sum_k \frac{\partial}{\partial w_k}\frac{\partial\sqrt{g}H}{\partial q_{hk}} = 0, \qquad (h = 1, 2, 3, 4). \tag{5}$$

We denote the left sides of the equations (4), (5) respectively by

$$[\sqrt{g}H]_{\mu\nu}, \qquad [\sqrt{g}H]_h$$

for short.

Let us call equations (4) the fundamental equations of gravitation, and equations (5) the fundamental electrodynamic equations, or generalized Maxwell equations. Due to the theorem stated above, the four equations (5) can be viewed as a consequence of equations (4), that is, because of that mathematical theorem we can directly make the claim *that in the sense as explained the electrodynamic phenomena are effects of gravitation*. I regard this | insight as the simple and very surprising solution of the problem of Riemann, who was the first to search for a theoretical connection between gravitation and light. [398]

In the following we use the easily proved fact that, if p^j $(j = 1, 2, 3, 4)$ is an *arbitrary* contravariant vector, the expression

$$p^{\mu\nu} = \sum_s (g_s^{\mu\nu}p^s - g^{\mu s}p_s^{\nu} - g^{\nu s}p_s^{\mu}), \qquad \left(p_s^j = \frac{\partial p^j}{\partial w_s}\right)$$

represents a symmetric contravariant tensor, and the expression

$$p_l = \sum_s (q_{ls}p^s + q_s p_l^s)$$

represents a covariant vector.

To proceed we establish two mathematical theorems, which express the following:

Theorem II. If J is an invariant depending on $g^{\mu\nu}$, $g_l^{\mu\nu}$, $g_{kl}^{\mu\nu}$, q_s, q_{sk}, then the following is always identically true in all arguments and for every arbitrary contravariant vector p^s: [2]

$$\sum_{\mu, \nu, l, k}\Big(\frac{\partial J}{\partial g^{\mu\nu}}\Delta g^{\mu\nu} + \frac{\partial J}{\partial g_l^{\mu\nu}}\Delta g_l^{\mu\nu} + \frac{\partial J}{\partial g_{kl}^{\mu\nu}}\Delta g_{kl}^{\mu\nu}\Big)$$
$$+\sum_{s, k}\Big(\frac{\partial J}{\partial q_s}\Delta q_s + \frac{\partial J}{\partial q_{sk}}\Delta q_{sk}\Big) = 0,$$

where

$$\Delta g^{\mu\nu} = \sum_m (g^{\mu m}p_m^\nu + g^{\nu m}p_m^\mu),$$

$$\Delta g_l^{\mu\nu} = -\sum_m g_m^{\mu\nu}p_l^m + \frac{\partial \Delta g^{\mu\nu}}{\partial w_l},$$

$$\Delta g_{lk}^{\mu\nu} = -\sum_m (g_m^{\mu\nu}p_{lk}^m + g_{lm}^{\mu\nu}p_k^m + g_{km}^{\mu\nu}p_l^m) + \frac{\partial^2 \Delta g^{\mu\nu}}{\partial w_l \partial w_k},$$

$$\Delta q_s = -\sum_m q_m p_s^m,$$

$$\Delta q_{sk} = -\sum_m q_{sm}p_k^m + \frac{\partial \Delta q_s}{\partial w_k}.$$

This theorem II can also be formulated as follows:
If J is an invariant and p^s and arbitrary vector as above, then the identity holds

$$\sum_s \frac{\partial J}{\partial w_s}p^s = PJ, \tag{6}$$

[399] | where we have put

$$P = P_g + P_q,$$

with

$$P_g = \sum_{\mu, \nu, l, k}\Big(p^{\mu\nu}\frac{\partial}{\partial g^{\mu\nu}} + p_l^{\mu\nu}\frac{\partial}{\partial g_l^{\mu\nu}} + p_{lk}^{\mu\nu}\frac{\partial}{\partial g_{lk}^{\mu\nu}}\Big)$$

$$P_q = \sum_{l, k}\Big(p_l\frac{\partial}{\partial q_l} + p_{lk}\frac{\partial}{\partial q_{lk}}\Big),$$

and used the abbreviations:

$$p_k^{\mu\nu} = \frac{\partial p^{\mu\nu}}{\partial w_k}, \qquad p_{kl}^{\mu\nu} = \frac{\partial^2 p^{\mu\nu}}{\partial w_k \partial w_l}, \qquad p_{lk} = \frac{\partial p_l}{\partial w_k}.$$

The proof of (6) follows easily; for this identity is obviously correct if p^s is a constant vector, and from this it follows in general because of its invariance.

Theorem III. If J is an invariant depending *only* on $g^{\mu\nu}$ and their derivatives, and if, as above, the variational derivatives of $\sqrt{g}J$ with respect to $g^{\mu\nu}$ are denoted by $[\sqrt{g}J]_{\mu\nu}$ then the expression—where $h^{\mu\nu}$ is understood to be any contravariant tensor—

$$\frac{1}{\sqrt{g}}\sum_{\mu,\nu}[\sqrt{g}J]_{\mu\nu}h^{\mu\nu}$$

represents an invariant; if we substitute in this sum in place of $h^{\mu\nu}$ the particular tensor $p^{\mu\nu}$ and write

$$\sum_{\mu,\nu}[\sqrt{g}J]_{\mu\nu}p^{\mu\nu} = \sum_{s,l}(i_s p^s + i_s^l p_l^s),$$

where then the expressions

$$i_s = \sum_{\mu,\nu}[\sqrt{g}J]_{\mu\nu}g_s^{\mu\nu},$$

$$i_s^l = -2\sum_{\mu}[\sqrt{g}J]_{\mu s}g^{\mu l}$$

depend only on the $g^{\mu\nu}$ and their derivatives, then we have

$$i_s = \sum_l \frac{\partial i_s^l}{\partial w_l} \tag{7}$$

in the sense that this equation is satisfied identically for all arguments, that is for the $g^{\mu\nu}$ and their derivatives.

For the proof we consider the integral

$$\int J\sqrt{g}d\omega, \qquad d\omega = dw_1 dw_2 dw_3 dw_4$$

to be taken over a finite piece of the four dimensional world. | Further, let p^s be a vector that vanishes together with its derivatives on the three dimensional surface of that piece of the world. Due to $P = P_g$ the last formula of the next page implies [400]

$$P_g(\sqrt{g}J) = \sum_s \frac{\partial \sqrt{g}Jp^s}{\partial w_s};$$

this results in

$$\int P_g(J\sqrt{g})d\omega = 0$$

and due to the way the Lagrangian derivative is formed we accordingly also have

$$\int \sum_{\mu,\nu} [\sqrt{g}J]_{\mu\nu} p^{\mu\nu} d\omega = 0.$$

Introduction of i_s, i_s^l into this identity finally shows that

$$\int\left(\sum_l \frac{\partial i_s^l}{\partial w_l} - i_s\right) p^s d\omega = 0$$

and therefore also that the assertion of our theorem is correct.

The most important aim is now the formulation of the concept of energy, and the derivation of the energy theorem solely on the basis of the two axioms I and II.

For this purpose we first form:

$$P_g(\sqrt{g}H) = \sum_{\mu,\nu,k,l}\left(\frac{\partial\sqrt{g}H}{\partial g^{\mu\nu}} p^{\mu\nu} + \frac{\partial\sqrt{g}H}{\partial g_k^{\mu\nu}} p_k^{\mu\nu} + \frac{\partial\sqrt{g}H}{\partial g_{kl}^{\mu\nu}} p_{kl}^{\mu\nu}\right).$$

Now $\dfrac{\partial H}{\partial g_{kl}^{\mu\nu}}$ is a mixed tensor of fourth rank, so if one puts

$$A_k^{\mu\nu} = p_k^{\mu\nu} + \sum_\rho\left(\begin{Bmatrix} k\rho \\ \mu \end{Bmatrix} p^{\rho\nu} + \begin{Bmatrix} k\rho \\ \nu \end{Bmatrix} p^{\rho\mu}\right),$$

$$\begin{Bmatrix} k\rho \\ \mu \end{Bmatrix} = \frac{1}{2}\sum_\sigma g^{\mu\sigma}(g_{k\sigma\rho} + g_{\rho\sigma k} - g_{k\rho\sigma}),$$

the expression

$$a^l = \sum_{\mu,\nu,k} \frac{\partial H}{\partial g_{kl}^{\mu\nu}} A_k^{\mu\nu} \tag{8}$$

becomes a contragredient vector.

Hence if we form the expression

$$P_g(\sqrt{g}H) - \sum_l \frac{\partial\sqrt{g}a^l}{\partial w_l}$$

[401] then this no longer contains the second derivatives $p_{kl}^{\mu\nu}$ and | therefore has the form

$$\sqrt{g}\sum_{\mu,\nu,k}(B_{\mu\nu}p^{\mu\nu} + B_{\mu\nu}^k p_k^{\mu\nu}),$$

where

$$B^k_{\mu\nu} = \sum_{\rho, l}\left(\frac{\partial H}{\partial g^{\mu\nu}_k} - \frac{\partial}{\partial w_l}\frac{\partial H}{\partial g^{\mu\nu}_{kl}} - \frac{\partial H}{\partial g^{\rho\nu}_{kl}}\begin{Bmatrix} l\mu \\ \rho \end{Bmatrix} - \frac{\partial H}{\partial g^{\mu\rho}_{kl}}\begin{Bmatrix} l\nu \\ \rho \end{Bmatrix}\right)$$

is again a mixed tensor.

Now we form the vector

$$b^l = \sum_{\mu, \nu} B^l_{\mu\nu} p^{\mu\nu}, \tag{9}$$

and obtain from it

$$P_g(\sqrt{g}H) - \sum_l \frac{\partial \sqrt{g}(a^l + b^l)}{\partial w_l} = \sum_{\mu, \nu}[\sqrt{g}H]_{\mu\nu} p^{\mu\nu}. \tag{10}$$

On the other hand we form

$$P_q(\sqrt{g}H) = \sum_{k, l}\left(\frac{\partial \sqrt{g}H}{\partial q_k} p_k + \frac{\partial \sqrt{g}H}{\partial q_{kl}} p_{kl}\right);$$

then $\dfrac{\partial H}{\partial q_{kl}}$ is a tensor and the expression

$$c^l = \sum_k \frac{\partial H}{\partial q_{kl}} p_k \tag{11}$$

therefore represents a contragredient vector. Correspondingly, as above, we obtain

$$P_q(\sqrt{g}H) - \sum_l \frac{\partial \sqrt{g} c^l}{d w_l} = \sum_k [\sqrt{g}H]_k p_k. \tag{12}$$

Now we note the basic equations (4) and (5), and conclude by adding (10) and (12):

$$P(\sqrt{g}H) = \sum_l \frac{\partial \sqrt{g}(a^l + b^l + c^l)}{d w_l}.$$

But we have

$$\begin{aligned} P(\sqrt{g}H) &= \sqrt{g}PH + H\sum_{\mu, \nu}\frac{\partial \sqrt{g}}{\partial g^{\mu\nu}} p^{\mu\nu} \\ &= \sqrt{g}PH + H\sum_s\left(\frac{\partial \sqrt{g}}{\partial w_s} p^s + \sqrt{g} p^s_s\right), \end{aligned}$$

and thus, due to identity (6)

$$P(\sqrt{g}H) = \sqrt{g}\sum_s \frac{\partial H}{\partial w_s}p^s + H\sum_s\left(\frac{\partial\sqrt{g}}{\partial w_s}p^s + \sqrt{g}p_s^s\right) = \sum_s \frac{\partial\sqrt{g}Hp^s}{\partial w_s}.$$

[402] | From this we finally obtain the invariant equation

$$\sum_l \frac{\partial}{\partial w_l}\sqrt{g}(Hp^l - a^l - b^l - c^l) = 0.$$

Now we note that

$$\frac{\partial H}{\partial q_{lk}} - \frac{\partial H}{\partial q_{kl}}$$

is a skew symmetric contravariant tensor; consequently

$$d^l = \frac{1}{2\sqrt{g}}\sum_{k,s}\frac{\partial}{\partial w_k}\left\{\left(\frac{\partial\sqrt{g}H}{\partial q_{lk}} - \frac{\partial\sqrt{g}H}{\partial q_{kl}}\right)p^s q_s\right\} \tag{13}$$

becomes a contravariant vector, which evidently satisfies the identity

$$\sum_l \frac{\partial\sqrt{g}d^l}{\partial w_l} = 0.$$

Let us now define

$$e^l = Hp^l - a^l - b^l - c^l - d^l \tag{14}$$

as the *energy vector, then the energy vector is a contravariant vector, which moreover depends linearly on the arbitrarily chosen vector* p^s, *and satisfies identically for that choice of this vector* p^s *the invariant energy equation*

$$\sum_l \frac{\partial\sqrt{g}e^l}{\partial w_l} = 0.$$

As far as the world function H is concerned, further axioms are needed to determine its choice in a unique way. If the gravitational field equations are to contain only second derivatives of the potentials $g^{\mu\nu}$, then H must have the form

$$H = K + L$$

where K is the invariant that derives from the Riemannian tensor (curvature of the four-dimensional manifold)

$$K = \sum_{\mu,\nu} g^{\mu\nu}K_{\mu\nu}$$

$$K_{\mu\nu} = \sum_{\kappa}\left(\frac{\partial}{\partial w_{\nu}}\begin{Bmatrix}\mu\kappa\\ \kappa\end{Bmatrix} - \frac{\partial}{\partial w_{\kappa}}\begin{Bmatrix}\mu\nu\\ \kappa\end{Bmatrix}\right) + \sum_{\kappa,\lambda}\left(\begin{Bmatrix}\mu\kappa\\ \lambda\end{Bmatrix}\begin{Bmatrix}\lambda\nu\\ \kappa\end{Bmatrix} - \begin{Bmatrix}\mu\nu\\ \lambda\end{Bmatrix}\begin{Bmatrix}\lambda\kappa\\ \kappa\end{Bmatrix}\right)$$

and where L depends only on $g^{\mu\nu}$, $g_l^{\mu\nu}$, q_s, q_{sk}. Finally we make the simplifying assumption in the following, that L does not contain the $g_l^{\mu\nu}$. |

Next we apply theorem II to the invariant L and obtain [403]

$$\sum_{\mu,\nu,m}\frac{\partial L}{\partial g^{\mu\nu}}(g^{\mu m}p_m^{\nu} + g^{\nu m}p_m^{\mu}) - \sum_{s,m}\frac{\partial L}{\partial q_s}q_m p_s^m$$
$$- \sum_{s,k,m}\frac{\partial L}{\partial q_{sk}}(q_{sm}p_k^m + q_{mk}p_s^m + q_m p_{sk}^m) = 0. \tag{15}$$

Equating to zero the coefficient of p_{sk}^m on the left produces the equation

$$\left(\frac{\partial L}{\partial q_{sk}} + \frac{\partial L}{\partial q_{ks}}\right)q_m = 0$$

or

$$\frac{\partial L}{\partial q_{sk}} + \frac{\partial L}{\partial q_{ks}} = 0, \tag{16}$$

that is, the derivatives of the electrodynamic potentials q_s occur only in the combinations

$$M_{ks} = q_{sk} - q_{ks}.$$

Thus we learn that under our assumptions the invariant L depends, besides on the potentials $g_{\mu\nu}, q_s$, only on the components of the skew symmetric invariant tensor

$$M = (M_{ks}) = \mathrm{Curl}(q_s),$$

that is, of the so-called electromagnetic six vector. *This result, which determines the character of Maxwell's equations in the first place, here derives essentially as a consequence of the general invariance, that is, on the basis of axiom II.*

If we put the coefficient of p_m^{ν} on the left of identity (15) equal to zero, we obtain, using (16)

$$2\sum_{\mu}\frac{\partial L}{\partial g^{\mu\nu}}g^{\mu m} - \frac{\partial L}{\partial q_m}q_{\nu} - \sum_{s}\frac{\partial L}{\partial M_{ms}}M_{\nu s} = 0, \qquad (\mu = 1, 2, 3, 4). \tag{17}$$

This equation admits an important transformation of the electromagnetic energy, that is the part of the energy vector that comes from L. Namely, this part results from (11), (13), (14) as follows:

$$Lp^l - \sum_k \frac{\partial L}{\partial q_{kl}} p_k - \frac{1}{2\sqrt{g}} \sum_{k,s} \frac{\partial}{\partial w_k} \left\{ \left(\frac{\partial \sqrt{g} L}{\partial q_{lk}} - \frac{\partial \sqrt{g} L}{\partial q_{kl}} \right) p^s q_s \right\}.$$

Because of (16) and by noting (5) this expression becomes |

[404]

$$\sum_{s,k} \left(L\delta_s^l - \frac{\partial L}{\partial M_{lk}} M_{sk} - \frac{\partial L}{\partial q_l} q_s \right) p^s \tag{18}$$

$$(\delta_s^l = 0, \quad l \neq s; \quad \delta_s^s = 1)$$

so because of (17) it equals

$$-\frac{2}{\sqrt{g}} \sum_{\mu,s} \frac{\partial \sqrt{g} L}{\partial g^{\mu s}} g^{\mu l} p^s. \tag{19}$$

Because of the formulas (21) to be developed below we see from this in particular that the electromagnetic energy, and therefore also the total energy vector e^l can be expressed through K alone, so that only the $g^{\mu\nu}$ and their derivatives, but not the q_s and their derivatives occur in it. If one takes the limit

$$g_{\mu\nu} = 0, \qquad (\mu \neq \nu)$$

$$g_{\mu\mu} = 1$$

in expression (18), then this limit agrees exactly with what Mie has proposed in his electrodynamics: *Mie's electromagnetic energy tensor is nothing but the generally invariant tensor that results from differentiation of the invariant L with respect to the gravitational potentials $g^{\mu\nu}$ in that limit*—a circumstance that gave me the first hint of the necessary close connection between Einstein's general relativity theory and Mie's electrodynamics, and which convinced me of the correctness of the theory here developed.

It remains to show directly how with the assumption

$$H = K + L \tag{20}$$

the generalized Maxwell equations (5) put forth above are entailed by the gravitational equations (4).

Using the notation introduced earlier for the variational derivatives with respect to the $g^{\mu\nu}$, the gravitational equations, because of (20), take the form

$$[\sqrt{g}K]_{\mu\nu} + \frac{\partial \sqrt{g} L}{\partial g^{\mu\nu}} = 0. \tag{21}$$

The first term on the left hand side becomes

$$[\sqrt{g}K]_{\mu\nu} = \sqrt{g}\left(K_{\mu\nu} - \tfrac{1}{2} K g_{\mu\nu}\right),$$

| as follows easily without calculation from the fact that $K_{\mu\nu}$, apart from $g_{\mu\nu}$, is the only tensor of second rank and K the only invariant, that can be formed using only the $g^{\mu\nu}$ and their first and second differential quotients, $g_k^{\mu\nu}$, $g_{kl}^{\mu\nu}$. [405]

The resulting differential equations of gravitation appear to me to be in agreement with the grand concept of the theory of general relativity established by Einstein in his later treatises.[5]

Further, if we denote in general the variational derivatives of $\sqrt{g}J$ with respect to the electrodynamic potential q_h as above by

$$[\sqrt{g}J]_h = \frac{\partial\sqrt{g}J}{\partial q_h} - \sum_k \frac{\partial}{\partial w_k}\frac{\partial\sqrt{g}J}{\partial q_{hk}},$$

then the basic electromagnetic equations assume the form, due to (20)

$$[\sqrt{g}L]_h = 0. \tag{22}$$

Since K is an invariant that depends only on the $g^{\mu\nu}$ and their derivatives, by theorem III the equation (7) holds identically, with

$$i_s = \sum_{\mu,\nu}[\sqrt{g}K]_{\mu\nu}g_s^{\mu\nu} \tag{23}$$

and

$$i_s^l = -2\sum_\mu[\sqrt{g}K]_{\mu s}g^{\mu l}, \qquad (\mu = 1, 2, 3, 4). \tag{24}$$

Due to (21) and (24), (19) equals $-\frac{1}{\sqrt{g}}i_\nu^m$. By differentiating with respect to w_m and summing over m we obtain because of (7)

$$\begin{aligned}
i_\nu &= \sum_m \frac{\partial}{\partial w_m}\left(-\sqrt{g}L\delta_\nu^m + \frac{\partial\sqrt{g}L}{\partial q_m}q_\nu + \sum_s\frac{\partial\sqrt{g}L}{\partial M_{sm}}M_{s\nu}\right) \\
&= -\frac{\partial\sqrt{g}L}{\partial w_\nu} + \sum_m\left\{q_\nu\frac{\partial}{\partial w_m}\left([\sqrt{g}L]_m + \sum_s\frac{\partial}{\partial w_s}\frac{\partial\sqrt{g}L}{\partial q_{ms}}\right)\right. \\
&\quad \left. + q_{\nu m}\left([\sqrt{g}L]_m + \sum_s\frac{\partial}{\partial w_s}\frac{\partial\sqrt{g}L}{\partial q_{ms}}\right)\right\} \\
&\quad + \sum_s\left([\sqrt{g}L]_s - \frac{\partial\sqrt{g}L}{\partial q_s}\right)M_{s\nu} + \sum_{s,m}\frac{\partial\sqrt{g}L}{\partial M_{sm}}\frac{\partial M_{s\nu}}{\partial w_m},
\end{aligned}$$

5 Loc. cit. *Berliner Sitzungsber.* 1915.

since of course

$$\frac{\partial\sqrt{g}L}{\partial q_m} = [\sqrt{g}L]_m + \sum_s \frac{\partial}{\partial w_s}\frac{\partial\sqrt{g}L}{\partial q_{ms}}$$

[406] | and[3]

$$-\sum_m \frac{\partial}{\partial w_m}\frac{\partial\sqrt{g}L}{\partial q_{sm}} = [\sqrt{g}L]_s - \frac{\partial\sqrt{g}L}{\partial q_s}.$$

Now we take into account that because of (16) we have

$$\sum_{m,s}\frac{\partial^2}{\partial w_m \partial w_s}\frac{\partial\sqrt{g}L}{\partial q_{ms}} = 0,$$

and then obtain by suitably collecting terms

$$\begin{aligned} i_\nu = -\frac{\partial\sqrt{g}L}{\partial w_\nu} &+ \sum_m \Big(q_\nu \frac{\partial}{\partial w_m}[\sqrt{g}L]_m + M_{m\nu}[\sqrt{g}L]_m\Big) \\ &+ \sum_m \frac{\partial\sqrt{g}L}{\partial q_m}q_{m\nu} + \sum_{s,m}\frac{\partial\sqrt{g}L}{\partial M_{sm}}\frac{\partial M_{s\nu}}{\partial w_m}. \end{aligned} \tag{25}$$

On the other hand we have

$$-\frac{\partial\sqrt{g}L}{\partial w_\nu} = -\sum_{s,m}\frac{\partial\sqrt{g}L}{\partial g^{sm}}g_\nu^{sm} - \sum_m \frac{\partial\sqrt{g}L}{\partial q_m}q_{m\nu} - \sum_{m,s}\frac{\partial\sqrt{g}L}{\partial q_{ms}}\frac{\partial q_{ms}}{\partial w_\nu}.$$

The first term on the right is nothing other than i_ν because of (21) and (23). The last term on the right proves to be equal and opposite to the last term on the right of (25); namely, we have

$$\sum_{s,m}\frac{\partial\sqrt{g}L}{\partial M_{sm}}\Big(\frac{\partial M_{s\nu}}{\partial w_m} - \frac{\partial q_{ms}}{\partial w_\nu}\Big) = 0, \tag{26}$$

since the expression

$$\frac{\partial M_{s\nu}}{\partial w_m} - \frac{\partial q_{ms}}{\partial w_\nu} = \frac{\partial^2 q_\nu}{\partial w_s \partial w_m} - \frac{\partial^2 q_s}{\partial w_\nu \partial w_m} - \frac{\partial^2 q_m}{\partial w_\nu \partial w_s}$$

is symmetric in s, m, and the first factor under the summation sign in (26) turns out to be skew symmetric in s, m.

Consequently (25) entails the equation

$$\sum_m \Big(M_{m\nu}[\sqrt{g}L]_m + q_\nu \frac{\partial}{\partial w_m}[\sqrt{g}L]_m\Big) = 0; \tag{27}$$

that is, from the gravitational equations (4) there follow indeed the four mutually independent linear combinations (27) of the basic electrodynamic equations (5) and their first derivatives. *This is the exact mathematical expression of the statement claimed in general above concerning the character of electrodynamics as a consequence of gravitation.* |

According to our assumption L should not depend on the derivatives of the $g^{\mu\nu}$; [407]
therefore L must be a function of certain four general invariants, which correspond to the special orthogonal invariants given by Mie, and of which the two simplest ones are these:

$$Q = \sum_{k,\, l,\, m,\, n} M_{mn} M_{lk} g^{mk} g^{nl}$$

and

$$q = \sum_{k,\, l} q_k q_l g^{kl}.$$

The simplest and most straightforward ansatz for L, considering the structure of K, is also that which corresponds to Mie's electrodynamics, namely

$$L = \alpha Q + f(q)$$

or, following Mie even more closely:

$$L = \alpha Q + \beta q^3,$$

where $f(q)$ denotes any function of q, and α, β are constants.

As one can see, the few simple assumptions expressed in axioms I and II suffice with appropriate interpretation to establish the theory: through it not only are our views of space, time, and motion fundamentally reshaped in the sense explained by Einstein, but I am also convinced that through the basic equations established here the most intimate, presently hidden processes in the interior of the atom will receive an explanation, and in particular that generally a reduction of all physical constants to mathematical constants must be possible—even as in the overall view thereby the possibility approaches that physics in principle becomes a science of the type of geometry: surely the highest glory of the axiomatic method, which as we have seen takes the powerful instruments of analysis, namely variational calculus and theory of invariants, into its service.

EDITORIAL NOTES

[1] The index $_l$ of $\partial\omega_l$ in the denominator of the third equation is missing in the original text.

[2] The subscript $_{sk}$ in the denominator of ∂q_{sk} is missing in the original text.

[3] The subscript $_s$ in the term $[\sqrt{g}L]_s$ is missing in the original text.

DAVID HILBERT

THE FOUNDATIONS OF PHYSICS (SECOND COMMUNICATION)

Originally published as "Die Grundlagen der Physik. (Zweite Mitteilung)" in Nachrichten von der Königlichen Gesellschaft zu Göttingen. Math.-phys. Klasse. 1917, p. 53–76. Presented in the Session of 23 December 1916.

In my first communication[1] I proposed a system of basic equations of physics. Before turning to the theory of integrating these equations it seems necessary to discuss some more general questions of a logical as well as physical nature.

First we introduce in place of the world parameters w_s $(s = 1, 2, 3, 4)$ the most general *real* spacetime coordinates x_s $(s = 1, 2, 3, 4)$ by putting

$$w_1 = x_1, \qquad w_2 = x_2, \qquad w_3 = x_3, \qquad w_4 = x_4,$$

and correspondingly in place of

$$ig_{14}, \qquad ig_{24}, \qquad ig_{34}, \qquad -g_{44},$$

we write simply

$$g_{14}, \qquad g_{24}, \qquad g_{34}, \qquad g_{44}.$$

The new $g_{\mu\nu}$ $(\mu, \nu = 1, 2, 3, 4)$ —the gravitational potentials of Einstein—shall then all be real functions of the real variables x_s $(s = 1, 2, 3, 4)$ of such a type that, in the representation of the quadratic form

$$G(X_1, X_2, X_3, X_4) = \sum_{\mu\nu} g_{\mu\nu} X_\mu X_\nu \tag{28}$$

as a sum of four squares of linear forms of the X_s, three squares always occur with positive sign, and one square with negative | sign: thus the quadratic form (28) provides our four dimensional world of the x_s with the metric of a pseudo-geometry. [54]
The determinant g of the $g_{\mu\nu}$ turns out to be negative.

1 This journal, 20 November 1915.

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 4
Gravitation in the Twilight of Classical Physics: The Promise of Mathematics.

If a curve

$$x_s = x_s(p) \qquad (s = 1, 2, 3, 4)$$

is given in this geometry, where $x_s(p)$ mean some arbitrary real functions of the parameter p, then it can be divided into pieces of curves on each of which the expression

$$G(\frac{dx_1}{dp}, \frac{dx_2}{dp}, \frac{dx_3}{dp}, \frac{dx_4}{dp})$$

does not change sign: A piece of the curve for which

$$G(\frac{dx_s}{dp}) > 0$$

shall be called a *segment* and the integral along this piece of curve

$$\lambda = \int \sqrt{G(\frac{dx_s}{dp})}dp$$

shall be the *length of the segment*; a piece of the curve for which

$$G(\frac{dx_s}{dp}) < 0$$

will be called a *time line*, and the integral

$$\tau = \int \sqrt{-G(\frac{dx_s}{dp})}dp$$

evaluated along this piece of curve shall be the *proper time of the time line*; finally a piece of curve along which

$$G(\frac{dx_s}{dp}) = 0$$

shall be called a *null line*.

To visualize these concepts of our pseudo geometry we imagine two ideal measuring devices: the *measuring thread* by means of which we are able to measure the length λ of any segment, and secondly the *light clock* with which we can determine the proper time of any time line. The thread shows zero and the light clock stops along every null line, whereas the former fails totally along a time line, and the latter along a segment. |

[55] First we show that each of the two instruments suffices to compute with its aid the values of the $g_{\mu\nu}$ as functions of x_s, as soon as a definite spacetime coordinate system x_s has been introduced. Indeed we choose any set of 10 segments, which all converge on the same world point x_s, from different directions, so that this endpoint

assumes the same parameter value p on each. At this end point we have the equation, for each of the 10 segments,

$$\left(\frac{d\lambda^{(h)}}{dp}\right)^2 = G\left(\frac{dx_s^{(h)}}{dp}\right), \qquad (h = 1, 2, \ldots, 10);$$

here the left-hand sides are known as soon as we have determined the lengths $\lambda^{(h)}$ by means of the thread. We introduce the abbreviations

$$D(u) = \begin{vmatrix} \left(\frac{dx_1^{(1)}}{dp}\right)^2, & \frac{dx_1^{(1)}}{dp}\frac{dx_2^{(1)}}{dp}, & \ldots, & \left(\frac{dx_4^{(1)}}{dp}\right)^2, & \left(\frac{d\lambda^{(1)}}{dp}\right)^2 \\ \cdots\cdots & \cdots\cdots & \cdots & \cdots\cdots & \cdots\cdots \\ \left(\frac{dx_1^{(10)}}{dp}\right)^2, & \frac{dx_1^{(10)}}{dp}\frac{dx_2^{(10)}}{dp}, & \ldots, & \left(\frac{dx_4^{(10)}}{dp}\right)^2, & \left(\frac{d\lambda^{(10)}}{dp}\right)^2 \\ X_1^2, & X_1X_2, & \ldots, & X_4^2, & u \end{vmatrix},$$

so that clearly

$$G(X_s) = -\frac{D(0)}{\frac{\partial D}{\partial u}}, \tag{29}$$

whereby also the condition on the directions of the chosen 10 segments at the point $x_s(p)$

$$\frac{\partial D}{\partial u} \neq 0$$

is seen to be necessary.

When G has been calculated according to (29), the use of this procedure for any 11th segment ending at $x_s(p)$ would yield the equation

$$\left(\frac{d\lambda^{(11)}}{dp}\right)^2 = G\left(\frac{dx_s^{(11)}}{dp}\right),$$

and this equation would then both verify the correctness of the instrument and confirm experimentally that the postulates of the theory apply to the real world.

Corresponding reasoning applies to the light clock. |

The axiomatic construction of our pseudo-geometry could be carried out without [56]
difficulty: first an axiom should be established from which it follows that length resp. proper time must be integrals whose integrand is only a function of the x_s and their first derivatives with respect to the parameter; suitable for such an axiom would be the property of development of the thread or the well-known envelope theorem for geodesic lines. Secondly an axiom is needed whereby the theorems of the pseudo-Euclidean geometry, that is the old principle of relativity, shall be valid in infinitesi-

mal regions; for this the axiom put down by W. Blaschke[2] would be particularly suitable, which states that the condition of orthogonality for any two directions—segments or time lines—shall always be a symmetric relation.

Let us briefly summarize the main facts that the Monge-Hamilton theory of differential equations teaches us for our pseudo-geometry.

With every world point x_s there is associated a cone of second order, with vertex at x_s, and determined in the running point coordinates X_s by the equation

$$G(X_1 - x_1, X_2 - x_2, X_3 - x_3, X_4 - x_4) = 0;$$

this shall be called the *null cone* belonging to the point x_s. The totality of null cones form a four dimensional field of cones, which is associated on the one hand with "Monge's" differential equation

$$G\left(\frac{dx_1}{dp}, \frac{dx_2}{dp}, \frac{dx_3}{dp}, \frac{dx_4}{dp}\right) = 0,$$

and on the other hand with "Hamilton's" partial differential equation

$$H\left(\frac{df}{dx_1}, \frac{df}{dx_2}, \frac{df}{dx_3}, \frac{df}{dx_4}\right) = 0, \tag{30}$$

where H denotes the quadratic form

$$H(U_1, U_2, U_3, U_4) = \sum_{\mu\nu} g^{\mu\nu} U_\mu U_\nu$$

reciprocal to G. The characteristics of Monge's and at the same time those of Hamilton's partial differential equation (30) are the geodesic null lines. All the geodesic null lines originating at one particular world point a_s $(s = 1, 2, 3, 4)$ generate a three
[57] dimensional point manifold, which | shall be called the *time divide* belonging to the world point a_s. This divide has a node at a_s, whose tangent cone is precisely the null cone belonging to a_s. If we transform the equation of the time divide into the form

$$x_4 = \varphi(x_1, x_2, x_3),$$

then

$$f = x_4 - \varphi(x_1, x_2, x_3)$$

is an integral of Hamilton's differential equation (30). All the time lines originating at the point a_s remain totally in the interior of that four dimensional part of the world whose boundary is the time divide of a_s.

After these preparations we turn to the problem of *causality* in the new physics.

---

2 "Räumliche Variationsprobleme mit symmetrischer Transversalitätsbedingung." *Leipziger Berichte, Math.-phys. Kl.* 68 (1916) p. 50.

Up to now all coordinate systems x_s, that result from any one by arbitrary transformation have been regarded as equally valid. This arbitrariness must be restricted when we want to realize the concept that two world points on the same time line can be related as cause and effect, and that it should then no longer be possible to transform such world points to be simultaneous. In declaring x_4 as the *true* time coordinate we adopt the following definition:

A *true* spacetime coordinate system is one for which the following four inequalities hold, in addition to $g < 0$:

$$g_{11} > 0, \qquad \begin{vmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{vmatrix} > 0, \qquad \begin{vmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ g_{31} & g_{32} & g_{33} \end{vmatrix} > 0, \qquad g_{44} < 0. \tag{31}$$

A transformation that transforms one such spacetime coordinate system into another true spacetime coordinate system shall be called a *true* spacetime coordinate transformation.

The four inequalities mean that at any world point a_s the associated null cone excludes the linear space

$$x_4 = a_4,$$

but contains in its interior the line

$$x_1 = a_1, \qquad x_2 = a_2, \qquad x_3 = a_3;$$

the latter line is therefore always a time line. |

Let any time line $x_s = x_s(p)$ be given; because [58]

$$G(\frac{dx_s}{dp}) < 0$$

it follows that in a true spacetime coordinate system we must always have

$$\frac{dx_4}{dp} \neq 0,$$

and therefore that along a time line the true time coordinate x_4 must always increase resp. decrease. Because a time line remains a time line upon every coordinate transformation, therefore two world points along one time line can never be given the same value of the time coordinate x_4 through a true spacetime transformation; that is, they cannot be transformed to be simultaneous.

On the other hand, if the points of a curve can be truly transformed to be simultaneous, then after this transformation we have for this curve

$$x_4 = \text{const.}, \qquad \text{that is} \qquad \frac{dx_4}{dp} = 0,$$

therefore

$$G(\frac{dx_s}{dp}) = \sum_{\mu\nu} g_{\mu\nu}\frac{dx_\mu}{dp}\frac{dx_\nu}{dp}, \qquad (\mu, \nu = 1, 2, 3),$$

and here the right side is positive because of the first three of our inequalities (31); the curve therefore characterizes a *segment*.

So we see that the concepts of cause and effect, which underlie the principle of causality, also do not lead to any inner contradictions whatever in the new physics, if we only take the inequalities (31) always to be part of our basic equations, that is if we confine ourselves to using *true* spacetime coordinates.

At this point let us take note of a special spacetime coordinate system that will later be useful and which I will call the *Gaussian coordinate system*, because it is the generalization of the system of geodesic polar coordinates introduced by Gauss in the theory of surfaces. In our four-dimensional world let any three-dimensional space be given so that every curve confined to that space is a segment: *a space of segments*, as
[59] I would like to call it; | let x_1, x_2, x_3 be any point coordinates of this space. We now construct at every point x_1, x_2, x_3 of this space the geodesic orthogonal to it, which will be a time line, and on this line we mark off x_4 as proper time; the point in the four-dimensional world so obtained is given coordinate values $x_1x_2x_3x_4$. In these coordinates we have, as is easily seen,

$$G(X_s) = \sum_{\mu\nu}^{1, 2, 3} g_{\mu\nu}X_\mu X_\nu - X_4^2, \tag{32}$$

that is, the Gaussian coordinate system is characterized analytically by the equations

$$g_{14} = 0, \qquad g_{24} = 0, \qquad g_{34} = 0, \qquad g_{44} = 0. \tag{33}$$

Because of the nature of the three dimensional space $x_4 = 0$ we presupposed, the quadratic form on the right-hand side of (32) in the variables X_1, X_2, X_3 is necessarily positive definite, so the first three of the inequalities (31) are satisfied, and since this also applies to the fourth, the Gaussian coordinate system always turns out to be a *true* spacetime coordinate system.

We now return to the investigation of the principle of causality in physics. As its main contents we consider the fact, valid so far in every physical theory, that from a knowledge of the physical quantities and their time derivatives in the present the future values of these quantities can always be determined: without exception the laws of physics to date have been expressed in a system of differential equations in which the number of the functions occurring in them was essentially the same as the number of independent differential equations; and thus the well-known general Cauchy theorem on the existence of integrals of partial differential equations directly offered the rationale of proof for the above fact.

Now, as I emphasized particularly in my first communication, the basic equations of physics (4) and (5) established there are by no means of the type characterized above; rather, according to Theorem I, four of them are a consequence of the rest: we regarded the four Maxwell equations (5) as a consequence of the ten gravitational equations (4), and so we have for the 14 potentials $g_{\mu\nu}$, q_s only 10 equations (4) that are essentially independent of each other. |

As soon as we maintain the demand of general invariance for the basic equations of physics the circumstance just mentioned is essential and even necessary. Because if there were further invariant equations, independent of (4), for the 14 potentials, then introduction of a Gaussian coordinate system would lead for the 10 physical quantities as per (33), [60]

$$g_{\mu\nu} \quad (\mu, \nu = 1, 2, 3), \qquad q_s \quad (s = 1, 2, 3, 4)$$

to a system of equations that would again be mutually independent, and mutually contradictory, because there are more than 10 of them.

Under such circumstances then, as occur in the new physics of general relativity, it is by no means any longer possible from knowledge of physical quantities in present and past to derive uniquely their future values. To show this intuitively on an example, let our basic equations (4) and (5) of the first communication be integrated in the special case corresponding to the presence of a single electron permanently at rest, so that the 14 potentials

$$\begin{aligned} g_{\mu\nu} &= g_{\mu\nu}(x_1, x_2, x_3) \\ q_s &= q_s(x_1, x_2, x_3) \end{aligned}$$

become definite functions of x_1, x_2, x_3, all independent of the time x_4, and in addition such that the first three components r_1, r_2, r_3 of the four-current density vanish. Then we apply the following coordinate transformation to these potentials:

$$\begin{cases} x_1 = x'_1 & \text{for } x'_4 \leq 0 \\ x_1 = x'_1 + e^{-\frac{1}{x'^2_4}} & \text{for } x'_4 > 0 \end{cases}$$
$$\begin{aligned} x_2 &= x'_2 \\ x_3 &= x'_3 \\ x_4 &= x'_4. \end{aligned}$$

For $x'_4 \leq 0$ the transformed potentials $g'_{\mu\nu}$, q'_s are the same functions of x'_1, x'_2, x'_3 as the $g_{\mu\nu}$, q_s of the original variables x_1, x_2, x_3, whereas the $g'_{\mu\nu}$, q'_s for $x'_4 > 0$ depend in an essential way also on the time coordinate x'_4; that is, the potentials $g'_{\mu\nu}$, q'_s represent an electron that is at rest until $x'_4 = 0$, but then puts its components into motion. |

Nonetheless I believe that it is only necessary to formulate more sharply the idea on which the principle of general relativity[3] is based, in order to maintain the principle of causality also in the new physics. Namely, to follow the essence of the new rel- [61]

ativity principle we must demand invariance not only for the general laws of physics, but we must accord invariance to each separate statement in physics that is to have physical meaning—in accordance with this, that in the final analysis it must be possible to establish each physical fact by thread or light clock, that is, instruments of *invariant* character. In the theory of curves and surfaces, where a statement in a chosen parametrization of the curve or surface has no geometrical meaning for the curve or surface itself, if this statement does not remain invariant under any arbitrary transformation of the parameters or cannot be brought to invariant form; so also in physics we must characterize a statement that does not remain invariant under any arbitrary transformation of the coordinate system as *physically meaningless*. For example, in the case considered above of the electron at rest, the statement that, say at the time $x_4 = 1$ this electron is at rest, has no physical meaning because this statement is not invariant.

Concerning the principle of causality, let the physical quantities and their time derivatives be known at the present in some given coordinate system: then a statement will only have physical meaning if it is invariant under all those transformations, for which the coordinates just used for the present remain unchanged; I maintain that statements of this type for the future are all uniquely determined, that is, *the principle of causality holds in this form*:

From present knowledge of the 14 physical potentials $g_{\mu\nu}$, q_s all statements about them for the future follow necessarily and uniquely provided they are physically meaningful.

To prove this proposition we use the *Gaussian* spacetime coordinate system. Introducing (33) into the basic equations (4) of the first communication yields for the 10 potentials |

[62]
$$g_{\mu\nu} \quad (\mu, \nu = 1, 2, 3), \qquad q_s \quad (s = 1, 2, 3, 4) \tag{34}$$

a system of as many partial differential equations; if we integrate these on the basis of the given initial values at $x_4 = 0$, we find uniquely the values of (34) for $x_4 > 0$. Since the Gaussian coordinate system itself is uniquely determined, therefore also all statements about those potentials (34) with respect to these coordinates are of invariant character.

The forms, in which physically meaningful, i.e. invariant, statements can be expressed mathematically are of great variety.

First. This can be done by means of an invariant coordinate system. Like the Gaussian system used above one can apply the well-known Riemannian one, as well as that spacetime coordinate system in which electricity appears at rest with unit current density. As at the end of the first communication, let $f(q)$ denote the function occurring in Hamilton's principle and depending on the invariant

3 In his original theory, now abandoned, A. Einstein (*Sitzungsberichte der Akad. zu Berlin*, 1914, p. 1067) had indeed postulated certain 4 non-invariant equations for the $g_{\mu\nu}$, in order to save the causality principle in its old form.

$$q = \sum_{kl} q_k q_l g^{kl},$$

then

$$r^s = \frac{\partial f(q)}{\partial q_s}$$

is the four-current density of electricity; it represents a contravariant vector and therefore can certainly be transformed to $(0, 0, 0, 1)$, as is easily seen. If this is done, then from the four equations

$$\frac{\partial f(q)}{\partial q_s} = 0 \quad (s = 1, 2, 3), \qquad \frac{\partial f(q)}{\partial q_4} = 1$$

the four components of the four-potential q_s can be expressed in terms of the $g_{\mu\nu}$, and every relation between the $g_{\mu\nu}$ in this or in one of the first two coordinate systems is then an invariant statement. For particular solutions of the basic equations there may be special invariant coordinate systems; for example, in the case treated below of the centrally symmetric gravitational field r, ϑ, φ, t form an invariant system of coordinates up to rotations.

Second. The statement, *according to which a coordinate system can be found* in which the 14 potentials $g_{\mu\nu}$, q_s have certain definite values in the future, or fulfill certain definite conditions, is always an invariant and therefore a physically meaningful one. The mathematically invariant expression for | such a statement is obtained by [63] eliminating the coordinates from those relations. The case considered above, of the electron at rest, provides an example: the essential and physically meaningful content of the causality principle is here expressed by the statement that the electron which is at rest for the time $x_4 \leq 0$ will, *for a suitably chosen spacetime coordinate system*, also remain at rest in all its parts for the future $x_4 > 0$.

Third. A statement is also invariant and thus has physical meaning if it is supposed to be valid in any arbitrary coordinate system. An example of this are Einstein's energy-momentum equations having divergence character. For, although Einstein's energy does not have the property of invariance, and the differential equations he put down for its components are by no means covariant as a system of equations, nevertheless the assertion contained in them, that they shall be satisfied in any coordinate system, is an invariant demand and therefore it carries physical meaning.

According to my exposition, physics is a four-dimensional pseudo-geometry, whose metric $g_{\mu\nu}$ is connected to the electromagnetic quantities, i.e. to the matter, by the basic equations (4) and (5) of my first communication. With this understanding, an old geometrical question becomes ripe for solution, namely whether and in what sense Euclidean geometry—about which we know from mathematics only that it is a logical structure free from contradictions—also possesses validity in the real world.

The old physics with the concept of absolute time took over the theorems of Euclidean geometry and without question put them at the basis of every physical theory. Gauss as well proceeded hardly differently: he constructed a hypothetical non-

Euclidean physics, by maintaining the absolute time and revoking only the parallel axiom from the propositions of Euclidean geometry; a measurement of the angles of a triangle of large dimensions showed him the invalidity of this non-Euclidean physics.

The new physics of Einstein's principle of general relativity takes a totally different position vis-à-vis geometry. It takes neither Euclid's nor any other particular geometry *a priori* as basic, in order to deduce from it the proper laws of physics, but,
[64] as I showed in my first communication, | the new physics provides at one fell swoop through one and the same Hamilton's principle the geometrical and the physical laws, namely the basic equations (4) and (5), which tell us how the metric $g_{\mu\nu}$ —at the same time the mathematical expression of the phenomenon of gravitation—is connected with the values q_s of the electrodynamic potentials.

Euclidean geometry is *an action-at-a-distance law foreign to the modern physics*: By revoking the Euclidean geometry as a general presupposition of physics, the theory of relativity maintains instead that geometry and physics have identical character and are based as *one* science on a common foundation.

The geometrical question mentioned above amounts to the investigation, whether and under what conditions the four-dimensional Euclidean pseudo-geometry

$$\begin{aligned} &g_{11} = 1, \qquad g_{22} = 1, \qquad g_{33} = 1, \qquad g_{44} = -1 \\ &g_{\mu\nu} = 0 \qquad (\mu \neq \nu) \end{aligned} \tag{35}$$

is a solution, or even the only regular solution, of the basic physical equations.

The basic equations (4) of my first communication are, due to the assumption (20) made there:

$$[\sqrt{g}K]_{\mu\nu} + \frac{\partial\sqrt{g}L}{\partial g^{\mu\nu}} = 0,$$

where

$$[\sqrt{g}K]_{\mu\nu} = \sqrt{g}\Big(K_{\mu\nu} - \frac{1}{2}K g_{\mu\nu}\Big).$$

When the values (35) are substituted, we have

$$[\sqrt{g}K]_{\mu\nu} = 0 \tag{36}$$

and for

$$q_s = 0 \qquad (s = 1, 2, 3, 4)$$

we have

$$\frac{\partial\sqrt{g}L}{\partial g^{\mu\nu}} = 0;$$

that is, when all electricity is removed, the pseudo-Euclidean geometry is possible. The question whether it is also necessary in this case, i.e. whether—or under certain additional conditions—the values (35), and those values of the $g_{\mu\nu}$ resulting from coordinate transformation of the latter, are the only regular solutions of the equations (36) is a mathematical problem not to be discussed here in general. Instead I confine myself | to presenting some thoughts concerning this problem in particular. [65]

For this we return to the original world coordinates of my first communication

$$w_1 = x_1, \qquad w_2 = x_2, \qquad w_3 = x_3, \qquad w_4 = ix_4,$$

and give the corresponding meaning to the $g_{\mu\nu}$.

In the case of the pseudo-Euclidean geometry we have

$$g_{\mu\nu} = \delta_{\mu\nu},$$

where

$$\delta_{\mu\nu} = 1, \qquad \delta_{\mu\nu} = 0 \quad (\mu \neq \nu).$$

For every metric in the neighborhood of this pseudo-Euclidean geometry the ansatz

$$g_{\mu\nu} = \delta_{\mu\nu} + \varepsilon h_{\mu\nu} + \ldots \tag{37}$$

is valid, where ε is a quantity converging to zero, and $h_{\mu\nu}$ are functions of the w_s. I make the following two assumptions about the metric (37):

I. The $h_{\mu\nu}$ shall be independent of the variable w_4.

II. The $h_{\mu\nu}$ shall show a certain regular behavior at infinity.

Now, if the metric (37) is to satisfy the differential equation (36) for all ε then it follows that the $h_{\mu\nu}$ must necessarily satisfy certain linear homogeneous partial differential equations of second order. If we substitute, following Einstein[4]

$$h_{\mu\nu} = k_{\mu\nu} - \frac{1}{2}\delta_{\mu\nu}\sum_s k_{ss}, \qquad (k_{\mu\nu} = k_{\nu\mu}) \tag{38}$$

and assume among the 10 functions $k_{\mu\nu}$ the four relations

$$\sum_s \frac{dk_{\mu s}}{dw_s} = 0, \qquad (\mu = 1, 2, 3, 4) \tag{39}$$

then these differential equations become:

$$\Box k_{\mu\nu} = 0, \tag{40}$$

where the abbreviation

4 “Näherungsweise Integration der Feldgleichungen der Gravitation.” *Berichte d. Akad. zu Berlin* 1916, p. 688.

$$\Box = \sum_s \frac{\partial^2}{\partial w_s^2}$$

has been used.

Because of the ansatz (38) the relations (39) are restrictive assumptions for the
[66] functions $h_{\mu\nu}$; however I will | show how one can always achieve, by suitable infinitesimal transformation of the variables w_1, w_2, w_3, w_4, that those restrictive assumptions are satisfied for the corresponding functions $h'_{\mu\nu}$ after the transformation.

To this end one should determine four functions $\varphi_1, \varphi_2, \varphi_3, \varphi_4$, which satisfy respectively the differential equations

$$\Box\varphi_\mu = \frac{1}{2}\frac{\partial}{\partial w_\mu}\sum_\nu h_{\nu\nu} - \sum_\nu \frac{\partial h_{\mu\nu}}{\partial w_\nu}. \tag{41}$$

By means of the infinitesimal transformation

$$w_s = w'_s + \varepsilon\varphi_s,$$

$g_{\mu\nu}$ becomes

$$g'_{\mu\nu} = g_{\mu\nu} + \varepsilon\sum_\alpha g_{\alpha\nu}\frac{\partial\varphi_\alpha}{\partial w_\mu} + \varepsilon\sum_\alpha g_{\alpha\mu}\frac{\partial\varphi_\alpha}{\partial w_\nu} + \dots$$

or because of (37) it becomes

$$g'_{\mu\nu} = \delta_{\mu\nu} + \varepsilon h'_{\mu\nu} + \dots,$$

where I have put

$$h'_{\mu\nu} = h_{\mu\nu} + \frac{\partial\varphi_\nu}{\partial w_\mu} + \frac{\partial\varphi_\mu}{\partial w_\nu}.$$

If we now choose

$$k_{\mu\nu} = h'_{\mu\nu} - \frac{1}{2}\delta_{\mu\nu}\sum_s h'_{ss},$$

then these functions satisfy Einstein's condition (39) because of (41), and we have

$$h'_{\mu\nu} = k_{\mu\nu} - \frac{1}{2}\delta_{\mu\nu}\sum_s k_{ss} \qquad (k_{\mu\nu} = k_{\nu\mu}).$$

The differential equations (40), which must be valid according to the above argument for the $k_{\mu\nu}$ we found, become due to assumption I

$$\frac{\partial^2 k_{\mu\nu}}{\partial w_1^2} + \frac{\partial^2 k_{\mu\nu}}{\partial w_2^2} + \frac{\partial^2 k_{\mu\nu}}{\partial w_3^2} = 0,$$

and, since assumption II—*mutatis mutandis*—allows the conclusion that the $k_{\mu\nu}$ approach constants at infinity, it follows that these must be constant in general, that is: *By varying the metric of the pseudo-Euclidean geometry under the assumptions I and II it is not possible to obtain a regular metric that is not likewise pseudo-Euclidean and which also corresponds to a world free of electricity.* |

The integration of the partial differential equations (36) can be performed in yet another case, first treated by Einstein[5] and by Schwarzschild.[6] In the following I present for this case a procedure that makes no assumptions about the gravitational potentials $g_{\mu\nu}$ at infinity, and which moreover offers advantages for my later investigations. The assumptions about the $g_{\mu\nu}$ are the following: [67]

1. The metric is represented in a Gaussian coordinate system, except that g_{44} is left arbitrary, i.e. we have
$$g_{14} = 0, \qquad g_{24} = 0, \qquad g_{34} = 0.$$
2. The $g_{\mu\nu}$ are independent of the time coordinate x_4.
3. The gravitation $g_{\mu\nu}$ is centrally symmetric with respect to the origin of coordinates.

According to Schwarzschild the most general metric conforming to these assumptions is represented in polar coordinates, where

$$\begin{aligned} w_1 &= r\cos\vartheta \\ w_2 &= r\sin\vartheta\cos\varphi \\ w_3 &= r\sin\vartheta\sin\varphi \\ w_4 &= l, \end{aligned}$$

by the expression

$$F(r)dr^2 + G(r)(d\vartheta^2 + \sin^2\vartheta d\varphi^2) + H(r)dl^2 \tag{42}$$

where $F(r), G(r), H(r)$ are still arbitrary functions of r. If we put

$$r^* = \sqrt{G(r)},$$

then we are equally justified in interpreting r^*, ϑ, φ as spatial polar coordinates. If we introduce r^* in (42) instead of r and then eliminate the sign $*$, the result is the expression

$$M(r)dr^2 + r^2 d\vartheta^2 + r^2\sin^2\vartheta d\varphi^2 + W(r)dl^2, \tag{43}$$

5 "Perihelbewegung des Merkur." *Situngsber. d. Akad. zu Berlin*. 1915, p. 831.
6 "Über das Gravitationsfeld eines Massenpunktes." *Sitzungsber. d. Akad. zu Berlin*. 1916, p. 189.

where $M(r), W(r)$ mean the two essential, arbitrary functions of r. The question is whether and how these can be determined in the most general way so that the differential equations (36) enjoy satisfaction. |

[68] To this end the well-known expressions $K_{\mu\nu}, K$ given in my first communication must be calculated. The first step in this is the derivation of the differential equations for geodesic lines by variation of the integral

$$\int\left(M\left(\frac{dr}{dp}\right)^2 + r^2\left(\frac{d\vartheta}{dp}\right)^2 + r^2\sin^2\vartheta\left(\frac{d\varphi}{dp}\right)^2 + W\left(\frac{dl}{dp}\right)^2\right)dp.$$

As Lagrange equations we obtain these:

$$\frac{d^2r}{dp^2} + \frac{1}{2}\frac{M'}{M}\left(\frac{dr}{dp}\right)^2 - \frac{r}{M}\left(\frac{d\vartheta}{dp}\right)^2 - \frac{r}{M}\sin^2\vartheta\left(\frac{d\varphi}{dp}\right)^2 - \frac{1}{2}\frac{W'}{M}\left(\frac{dl}{dp}\right)^2 = 0,$$

$$\frac{d^2\vartheta}{dp^2} + \frac{2}{r}\frac{dr}{dp}\frac{d\vartheta}{dp} - \sin\vartheta\cos\vartheta\left(\frac{d\varphi}{dp}\right)^2 = 0,$$

$$\frac{d^2\varphi}{dp^2} + \frac{2}{r}\frac{dr}{dp}\frac{d\varphi}{dp} + 2\cot\vartheta\frac{d\vartheta}{dp}\frac{d\varphi}{dp} = 0,$$

$$\frac{d^2l}{dp^2} + \frac{W'}{W}\frac{dr}{dp}\frac{dl}{dp} = 0;$$

here and in the following calculation the sign $'$ denotes the derivative with respect to r. By comparison with the general differential equations of geodesic lines:

$$\frac{d^2w_s}{dp^2} + \sum_{\mu\nu}\begin{Bmatrix}\mu\nu\\ s\end{Bmatrix}\frac{dw_\mu}{dp}\frac{dw_\nu}{dp} = 0,$$

we obtain for the bracket symbols $\begin{Bmatrix}\mu\nu\\ s\end{Bmatrix}$ the following values, whereby those that vanish are omitted:

$$\begin{Bmatrix}11\\ 1\end{Bmatrix} = \frac{1}{2}\frac{M'}{M}, \qquad \begin{Bmatrix}22\\ 1\end{Bmatrix} = -\frac{r}{M}, \qquad \begin{Bmatrix}33\\ 1\end{Bmatrix} = -\frac{r}{M}\sin^2\vartheta,$$

$$\begin{Bmatrix}44\\ 1\end{Bmatrix} = -\frac{1}{2}\frac{W'}{M}, \qquad \begin{Bmatrix}12\\ 2\end{Bmatrix} = \frac{1}{r}, \qquad \begin{Bmatrix}33\\ 2\end{Bmatrix} = -\sin\vartheta\cos\vartheta,$$

$$\begin{Bmatrix}13\\ 3\end{Bmatrix} = \frac{1}{r}, \qquad \begin{Bmatrix}23\\ 3\end{Bmatrix} = \cot\vartheta, \qquad \begin{Bmatrix}14\\ 4\end{Bmatrix} = \frac{1}{2}\frac{W'}{W}.$$

With these we form:

$$
\begin{aligned}
K_{11} = {} & \frac{\partial}{\partial r}\left(\begin{Bmatrix}11\\1\end{Bmatrix}+\begin{Bmatrix}12\\2\end{Bmatrix}+\begin{Bmatrix}13\\3\end{Bmatrix}+\begin{Bmatrix}14\\4\end{Bmatrix}\right)-\frac{\partial}{\partial r}\begin{Bmatrix}11\\1\end{Bmatrix} \\
& +\begin{Bmatrix}11\\1\end{Bmatrix}\begin{Bmatrix}11\\1\end{Bmatrix}+\begin{Bmatrix}12\\2\end{Bmatrix}\begin{Bmatrix}21\\2\end{Bmatrix}+\begin{Bmatrix}13\\3\end{Bmatrix}\begin{Bmatrix}31\\3\end{Bmatrix}+\begin{Bmatrix}14\\4\end{Bmatrix}\begin{Bmatrix}41\\4\end{Bmatrix} \\
& -\begin{Bmatrix}11\\1\end{Bmatrix}\left(\begin{Bmatrix}11\\1\end{Bmatrix}+\begin{Bmatrix}12\\2\end{Bmatrix}+\begin{Bmatrix}13\\3\end{Bmatrix}+\begin{Bmatrix}14\\4\end{Bmatrix}\right) \\
= {} & \frac{1}{2}\frac{W''}{W}+\frac{1}{4}\frac{W'^2}{W^2}-\frac{M'}{rM}-\frac{1}{4}\frac{M'W'}{MW}
\end{aligned}
$$

$$
\begin{aligned}
K_{22} = {} & \frac{\partial}{\partial \vartheta}\begin{Bmatrix}23\\3\end{Bmatrix}-\frac{\partial}{\partial r}\begin{Bmatrix}22\\1\end{Bmatrix} \\
& +\begin{Bmatrix}21\\2\end{Bmatrix}\begin{Bmatrix}22\\1\end{Bmatrix}+\begin{Bmatrix}22\\1\end{Bmatrix}\begin{Bmatrix}12\\2\end{Bmatrix}+\begin{Bmatrix}23\\3\end{Bmatrix}\begin{Bmatrix}32\\3\end{Bmatrix} \\
& -\begin{Bmatrix}22\\1\end{Bmatrix}\left(\begin{Bmatrix}11\\1\end{Bmatrix}+\begin{Bmatrix}12\\2\end{Bmatrix}+\begin{Bmatrix}13\\3\end{Bmatrix}+\begin{Bmatrix}14\\4\end{Bmatrix}\right) \\
= {} & -1+\frac{1}{2}\frac{rM'}{M^2}+\frac{1}{M}+\frac{1}{2}\frac{rW'}{MW}
\end{aligned}
$$

[69]

$$
\begin{aligned}
K_{33} = {} & -\frac{\partial}{\partial r}\begin{Bmatrix}33\\1\end{Bmatrix}-\frac{\partial}{\partial \vartheta}\begin{Bmatrix}33\\2\end{Bmatrix} \\
& +\begin{Bmatrix}31\\3\end{Bmatrix}\begin{Bmatrix}33\\1\end{Bmatrix}+\begin{Bmatrix}32\\3\end{Bmatrix}\begin{Bmatrix}33\\2\end{Bmatrix}+\begin{Bmatrix}33\\1\end{Bmatrix}\begin{Bmatrix}13\\3\end{Bmatrix}+\begin{Bmatrix}33\\2\end{Bmatrix}\begin{Bmatrix}23\\3\end{Bmatrix} \\
& -\begin{Bmatrix}33\\1\end{Bmatrix}\left(\begin{Bmatrix}11\\1\end{Bmatrix}+\begin{Bmatrix}12\\2\end{Bmatrix}+\begin{Bmatrix}13\\3\end{Bmatrix}+\begin{Bmatrix}14\\4\end{Bmatrix}\right)-\begin{Bmatrix}33\\2\end{Bmatrix}\begin{Bmatrix}23\\3\end{Bmatrix} \\
= {} & \sin^2\vartheta\left(-1-\frac{1}{2}\frac{rM'}{M^2}+\frac{1}{M}+\frac{1}{2}\frac{rW'}{MW}\right)
\end{aligned}
$$

$$K_{44} = -\frac{\partial}{\partial r}\begin{Bmatrix}44\\1\end{Bmatrix} + \begin{Bmatrix}41\\4\end{Bmatrix}\begin{Bmatrix}44\\1\end{Bmatrix} + \begin{Bmatrix}44\\1\end{Bmatrix}\begin{Bmatrix}41\\4\end{Bmatrix}$$

$$-\begin{Bmatrix}44\\1\end{Bmatrix}\left(\begin{Bmatrix}11\\1\end{Bmatrix} + \begin{Bmatrix}12\\2\end{Bmatrix} + \begin{Bmatrix}13\\3\end{Bmatrix} + \begin{Bmatrix}14\\4\end{Bmatrix}\right)$$

$$= \frac{1}{2}\frac{W''}{M} - \frac{1}{4}\frac{M'W'}{M^2} - \frac{1}{4}\frac{W'^2}{MW} + \frac{W'}{rM}$$

$$K = \sum_s g^{ss}K_{ss} = \frac{W''}{MW} - \frac{1}{2}\frac{W'^2}{MW^2} - 2\frac{M'}{rM^2} - \frac{1}{2}\frac{M'W'}{M^2W}$$

$$-\frac{2}{r^2} + \frac{2}{r^2M} + 2\frac{W'}{rMW}.$$

Because

$$\sqrt{g} = \sqrt{MW}r^2\sin\vartheta$$

we have

$$K\sqrt{g} = \left\{\left(\frac{r^2W'}{\sqrt{MW}}\right)' - 2\frac{rM'\sqrt{W}}{M^{3/2}} - 2\sqrt{MW} + 2\sqrt{\frac{W}{M}}\right\}\sin\vartheta,$$

and if we put

$$M = \frac{r}{r-m}, \qquad W = w^2\frac{r-m}{r},$$

where now m and w are the unknown functions of r, we finally obtain

$$K\sqrt{g} = \left\{\left(\frac{r^2W'}{\sqrt{MW}}\right)' - 2wm'\right\}\sin\vartheta,$$

[70] | so that the variation of the quadruple integral

$$\iiiint K\sqrt{g}\ dr\ d\vartheta\ d\varphi\ dl$$

is equivalent to the variation of the single integral

$$\int wm'dr$$

and leads to the Lagrange equations

$$\begin{aligned} m' &= 0 \\ w' &= 0. \end{aligned} \tag{44}$$

It is easy to convince oneself that these equations indeed imply that all $K_{\mu\nu}$ vanish; they therefore represent essentially the most general solution of equations (36) under the assumptions 1., 2., 3., we made. If we take as integrals of (44) $m = \alpha$, where α is a constant, and $w = 1$, which evidently is no essential restriction, then for $l = it$ (43) results in the desired metric in the form first found by Schwarzschild

$$G(dr, d\vartheta, d\varphi, dl) = \frac{r}{r-\alpha}dr^2 + r^2 d\vartheta^2 + r^2 \sin^2\vartheta d\varphi^2 - \frac{r-\alpha}{r}dl^2. \tag{45}$$

The *singularity* of the metric at $r = 0$ disappears only if we take $\alpha = 0$, i.e. *the metric of the pseudo-Euclidean geometry is the only regular metric that corresponds to a world without electricity under the assumptions* 1., 2., 3.

If $\alpha \neq 0$, then $r = 0$ and, for positive α also $r = \alpha$, prove to be places where the metric is not regular. Here I call a metric or gravitational field $g_{\mu\nu}$ *regular* at some place if it is possible to introduce by transformation with unique inverse a coordinate system for which the corresponding functions $g'_{\mu\nu}$ at that place are regular, that is they are continuous and arbitrarily differentiable at the place and its neighborhood, and have a determinant g' that differs from zero.

Although in my view only regular solutions of the basic physical equations represent reality directly, still it is precisely the solutions with places of non-regularity that are an important mathematical instrument for approximating characteristic regular solutions—and in this sense, following Einstein and Schwarzschild, the metric (45), not regular at $r = 0$ and $r = \alpha$, is to be viewed as the expression for | gravity of a [71]
centrally symmetric mass distribution in the neighborhood of the origin[7]. In the same sense a point mass is to be understood as the limit of a certain distribution of electricity about one point, but I refrain at this place from deriving its equations of motion from my basic physical equations. A similar situation prevails for the question about the differential equations for the propagation of light.

Following Einstein, let the *following two axioms* serve as a substitute for a derivation from the basic equations:

The motion of a point mass in a gravitational field is described by a geodesic line, which is a time line[8].

The motion of light in a gravitational field is described by a geodesic null line.

Because the world line representing the motion of a point mass shall be a time line, it is easily seen to be always possible to bring the point mass to rest by *true* spacetime transformations, i.e. there are *true* spacetime coordinate systems with respect to which the point mass remains at rest.

The differential equations of geodesic lines for the centrally symmetric gravitational field (45) arise from the variational problem

---

7 To transform the locations $r = \alpha$ to the origin, as Schwarzschild does, is not to be recommended in my opinion; Schwarzschild's transformation is moreover not the simplest that achieves this goal.

8 This last restrictive addition is to be found neither in Einstein nor in Schwarzschild.

$$\delta\int\left(\frac{r}{r-\alpha}\left(\frac{dr}{dp}\right)^2+r^2\left(\frac{d\vartheta}{dp}\right)^2+r^2\sin^2\vartheta\left(\frac{d\varphi}{dp}\right)^2-\frac{r-\alpha}{r}\left(\frac{dt}{dp}\right)^2\right)dp = 0,$$

and become, by well-known methods:

$$\frac{r}{r-\alpha}\left(\frac{dr}{dp}\right)^2+r^2\left(\frac{d\vartheta}{dp}\right)^2+r^2\sin^2\vartheta\left(\frac{d\varphi}{dp}\right)^2-\frac{r-\alpha}{r}\left(\frac{dt}{dp}\right)^2 = A, \tag{46}$$

$$\frac{d}{dp}\left(r^2\frac{d\vartheta}{dp}\right)-r^2\sin\vartheta\cos\vartheta\left(\frac{d\varphi}{dp}\right)^2 = 0, \tag{47}$$

$$r^2\sin^2\vartheta\frac{d\varphi}{dp} = B, \tag{48}$$

$$\frac{r-\alpha}{r}\frac{dt}{dp} = C, \tag{49}$$

where A, B, C denote constants of integration. |

[72] I first prove *that the orbits in the* $r\vartheta\varphi$- *space always lie in planes passing through the center of the gravitation*.

To this end we eliminate the parameter p from the differential equations (47) and (48) to obtain a differential equation for ϑ as a function of φ. We have the identity

$$\begin{aligned}\frac{d}{dp}\left(r^2\frac{d\vartheta}{dp}\right) &= \frac{d}{dp}\left(r^2\frac{d\vartheta}{d\varphi}\cdot\frac{d\varphi}{dp}\right)\\ &= \left(2r\frac{dr}{d\varphi}\frac{d\vartheta}{d\varphi}+r^2\frac{d^2\vartheta}{d\varphi^2}\right)\left(\frac{d\varphi}{dp}\right)^2+r^2\frac{d\vartheta}{d\varphi}\frac{d^2\varphi}{dp^2}.\end{aligned} \tag{50}$$

On the other hand, differentiation of (48) with respect to p gives:

$$\left(2r\frac{dr}{d\varphi}\sin^2\vartheta+2r^2\sin\vartheta\cos\vartheta\frac{d\vartheta}{d\varphi}\right)\left(\frac{d\varphi}{dp}\right)^2+r^2\sin^2\vartheta\frac{d^2\varphi}{dp^2} = 0,$$

and if we take from this the value of $\frac{d^2\varphi}{dp^2}$ and substitute on the right of (50), it becomes

$$\frac{d}{dp}\left(r^2\frac{d\vartheta}{dp}\right) = \left(\frac{d^2\vartheta}{d\varphi^2}-2\cot\vartheta\left(\frac{d\vartheta}{d\varphi}\right)^2\right)r^2\left(\frac{d\varphi}{dp}\right)^2.$$

Thus equation (47) takes the form:

$$\frac{d^2\vartheta}{d\varphi^2}-2\cot\vartheta\left(\frac{d\vartheta}{d\varphi}\right)^2 = \sin\vartheta\cos\vartheta,$$

a differential equation whose general integral is

$$\sin\vartheta\cos(\varphi + a) + b\cos\vartheta = 0,$$

where a and b denote constants of integration.

This provides the desired proof, and it is therefore sufficient for further discussion of geodesic lines to consider only the value $\vartheta = 2/\pi$. Then the variational problem simplifies as follows

$$\delta\int\left\{\frac{r}{r-\alpha}\left(\frac{dr}{dp}\right)^2 + r^2\left(\frac{d\varphi}{dp}\right)^2 - \frac{r-\alpha}{r}\left(\frac{dt}{dp}\right)^2\right\}dp = 0,$$

and the three differential equations of first order that arise from it are |

$$\frac{r}{r-\alpha}\left(\frac{dr}{dp}\right)^2 + r^2\left(\frac{d\varphi}{dp}\right)^2 - \frac{r-\alpha}{r}\left(\frac{dt}{dp}\right)^2 = A, \qquad (51)$$ [73]

$$r^2\frac{d\varphi}{dp} = B, \qquad (52)$$

$$\frac{r-\alpha}{r}\frac{dt}{dp} = C. \qquad (53)$$

The Lagrange differential equation for r

$$\frac{d}{dp}\left(\frac{2r}{r-\alpha}\frac{dr}{dp}\right) + \frac{\alpha}{(r-\alpha)^2}\left(\frac{dr}{dp}\right)^2 - 2r\left(\frac{d\varphi}{dp}\right)^2 + \frac{\alpha}{r^2}\left(\frac{dt}{dp}\right)^2 = 0 \qquad (54)$$

is necessarily related to the above equations, in fact if we denote the left sides of (51), (52), (53), (54) with [1], [2], [3], [4] respectively we have identically

$$\frac{d[1]}{dp} - 2\frac{d\varphi}{dp}\,\frac{d[2]}{dp} + 2\frac{dt}{dp}\,\frac{d[3]}{dp} = \frac{dr}{dp}[4]. \qquad (55)$$

By choosing $C = 1$, which amounts to multiplying the parameter p by a constant, and then eliminating p and t from (51), (52), (53) we obtain that differential equation for $\rho = 1/r$ as a function of φ found by Einstein and Schwarzschild, namely:

$$\left(\frac{d\rho}{d\varphi}\right)^2 = \frac{1+A}{B^2} - \frac{A\alpha}{B^2}\,\rho - \rho^2 + \alpha\rho^3. \qquad (56)$$

This equation represents the orbit of the point mass in polar coordinates; in first approximation for $\alpha = 0$ with $B = \sqrt{\alpha}b$, $A = -1 + \alpha a$ the Kepler motion follows from it, and the second approximation than leads to the most shining discovery of the present: the calculation of the advance of the perihelion of Mercury.

According to the axiom above the world line for the motion of a point mass shall be a time line; from the definition of the time line it thus follows that always $A < 0$.

We now ask in particular whether a circle, i.e. $r = \text{const.}$ can be the orbit of a motion. The identity (55) shows that in this case—because of $dr/dp = 0$—equation (54) is by no means a consequence of (51), (52), (53); the latter three equations therefore are insufficient to determine the motion; instead the necessary equations to be satisfied are (52), (53), (54). From (54) it follows that |

[74]
$$-2r\left(\frac{d\varphi}{dp}\right)^2 + \frac{\alpha}{r^2}\left(\frac{dt}{dp}\right)^2 = 0 \tag{57}$$

or that for the speed v on the circular orbit

$$v^2 = \left(r\frac{d\varphi}{dt}\right)^2 = \frac{\alpha}{2r}. \tag{58}$$

On the other hand, since $A < 0$, (51) implies the inequality

$$r^2\left(\frac{d\varphi}{dp}\right)^2 - \frac{r-\alpha}{r}\left(\frac{dt}{dp}\right)^2 < 0 \tag{59}$$

or by using (57)

$$r > \frac{3\alpha}{2}. \tag{60}$$

With (58) this implies the inequality for the speed of the mass point moving on a circle[9]

$$v < \frac{1}{\sqrt{3}}. \tag{61}$$

The inequality (60) allows the following interpretation: From (58) the angular speed of the orbiting point mass is

$$\frac{d\varphi}{dt} = \sqrt{\frac{\alpha}{2r^3}}.$$

So if we want to introduce instead of r, φ the polar coordinates of a coordinate system co-rotating about the origin, we only have to replace

$$\varphi \quad \text{by} \quad \varphi + \sqrt{\frac{\alpha}{2r^3}}t.$$

After the corresponding spacetime transformation the metric

---

9 Schwarzschild's (loc. cit.) claim that the speed of the point mass on a circular orbit approaches the limit $1/\sqrt{2}$ as the orbit radius is decreased corresponds to the inequality $r \geq \alpha$ and should not be regarded as accurate, according to the above.

$$\frac{r}{r-\alpha}dr^2 + r^2 d\varphi^2 - \frac{r-\alpha}{r}dt^2$$

becomes

$$\frac{r}{r-\alpha}dr^2 + r^2 d\varphi^2 + \sqrt{2\alpha r}\, d\varphi\, dt + \left(\frac{\alpha}{2r} - \frac{r-\alpha}{r}\right)dt^2.$$

| Here the inequality $g_{44} < 0$ is satisfied due to (60), and since the other inequalities (31) are satisfied, *the transformation under discussion of the point mass to rest is a true spacetime transformation*. [75]

On the other hand, the upper limit $1/\sqrt{3}$ found in (61) for the speed of a mass point on a circular orbit also has a simple interpretation. According to the axiom for light propagation this propagation is represented by a null geodesic. Accordingly if we put $A = 0$ in (51), instead of the inequality (59) the result for circular light propagation is the equation

$$r^2\left(\frac{d\varphi}{dp}\right)^2 - \frac{r-\alpha}{r}\left(\frac{dt}{dp}\right)^2 = 0;$$

together with (57) this implies for the radius of the light's orbit:

$$r = \frac{3\alpha}{2}$$

and for the speed of the orbiting light the value that occurs as the upper limit in (61):

$$v = \frac{1}{\sqrt{3}}.$$

In general we find for the orbit of light from (56) with $A = 0$ the differential equation

$$\left(\frac{d\rho}{d\varphi}\right)^2 = \frac{1}{B^2} - \rho^2 + \alpha\rho^3; \tag{62}$$

for $B = \frac{3\sqrt{3}}{2}\alpha$ it has the circle $r = \frac{3\alpha}{2}$ as a Poincaré "cycle"—corresponding to the circumstance that thereupon $\rho - \frac{2}{3\alpha}$ is a double factor of the right-hand side. Indeed in this case—and correspondingly for the more general equation (56)—the differential equation (62) possesses infinitely many integral curves, which approach that circle as the limit of spirals, as demanded by Poincaré's general theory of cycles.

If we consider a light ray approaching from infinity and take α small compared to the ray's distance of closest approach from the center of gravitation, then the light ray has approximately the form of a hyperbola with focus at the center.[10] |

[76] A counterpart to the motion on a circle is the motion on a straight line that passes through the center of gravitation. We obtain the differential equation for this motion if we set $\varphi = 0$ in (54) and then eliminate p from (53) and (54); the differential equation so obtained for r as a function of t is

$$\frac{d^2r}{dt^2} - \frac{3\alpha}{2r(r-\alpha)}\left(\frac{dr}{dt}\right)^2 + \frac{r(r-\alpha)}{2r^3} = 0 \tag{63}$$

with the integral following from (51)

$$\left(\frac{dr}{dt}\right)^2 = \left(\frac{r-\alpha}{r}\right)^2 + A\left(\frac{r-\alpha}{r}\right)^3. \tag{64}$$

According to (63) the acceleration is negative or positive, i.e. gravitation acts attractive or repulsive, according as the absolute value of the velocity

$$\left|\frac{dr}{dt}\right| < \frac{1}{\sqrt{3}}\frac{r-\alpha}{r}$$

or

$$> \frac{1}{\sqrt{3}}\frac{r-\alpha}{r}.$$

For light we have because of (64)

$$\left|\frac{dr}{dt}\right| = \frac{r-\alpha}{r};$$

light propagating in a straight line towards the center is always repelled, in agreement with the last inequality; its speed increases from 0 at $r = \alpha$ to 1 at $r = \infty$.

When α as well as dr/dt are small, (63) becomes approximately the Newtonian equation

$$\frac{d^2r}{dt^2} = -\frac{\alpha}{2}\frac{1}{r^2}.$$

10 A detailed discussion of the differential equations (56) and (62) will be the task of a communication by V. Fréedericksz to appear in these pages.

FROM PERIPHERAL MATHEMATICS TO
A NEW THEORY OF GRAVITATION

JOHN STACHEL

THE STORY OF NEWSTEIN
OR: IS GRAVITY JUST ANOTHER PRETTY FORCE?

1. 1. INTRODUCTION

In this paper I will argue for the following three theses:

1. The concepts of parallel displacement in Riemannian geometry and of a non-metrical affine connection were developed postmaturely (see Section 2): By the latter third of the nineteenth century, all of the mathematical prerequisites for their introduction were available, and it is a historical accident that they were not developed before the second decade of the twentieth century (see Section 3).
2. The appropriate mathematical context for implementing the equivalence principle is the theory of affine connections on the category of frame bundles, with the bundle morphisms induced by diffeomorphisms on the base manifold (see the Appendix).[1] This theory allows a mathematically precise formulation of Einstein's insight that gravitation and inertia are "essentially the same [*wesensgleich*]" as he put it (see Section 5). The absence of this context constituted a serious obstacle to the development of the general theory of relativity—indeed an insurmountable one to its development by the mathematically most direct route. Consequently, Einstein was forced to take a detour through a long and indirect route from the initial formulation of the equivalence principle in 1907 to the final formulation of the field equations in 1915 (see Section 10). The detour involved focusing attention almost exclusively on the chrono-geometrical structure of spacetime, and to this day, many discussions of the interpretation of the general theory, and of the problem of quantum gravity, still reflect the negative consequences of this detour.
3. Had the concept of an affine connection been developed in a timely manner, the affine formulation of Newtonian gravitation theory, which was actually developed only *after* the formulation of *general* relativity,[2] could have been developed *before* the formulation of *special* relativity. From the outset, such a formulation would have placed appropriate emphasis on the inertio-gravitational structure of

1 Insofar as needed for this paper, these concepts are briefly explained in the Appendix. A particularly useful reference for a more extended discussion of most of these concepts is (Crampin and Pirani 1986).

2 See (Cartan 1923; Friedrichs 1927). Excerpts from Cartan can be found in this volume.

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 4
Gravitation in the Twilight of Classical Physics: The Promise of Mathematics.

spacetime and posed the question of its relation to the chronometry and geometry of spacetime (see Sections 6 and 7). When special relativity, with its new chrono-geometry, was developed, this context for gravitation theory would have made the transition from the special to the general theory of relativity rather transparent, thereby avoiding the negative consequences of the actual transition mentioned above.

In order to vivify these rather abstract theses, I have created Isaac Albert Newstein (= Newton + Einstein), a mythical physicist who combines Newton's approach to the kinematical structure of space and time (chronometry and geometry) with Einstein's insight into the implications of the equivalence principle for (Newtonian) gravitation theory (see Section 7). He did this shortly after Hermann Weylmann (= Weyl + Grassmann), an equally mythical mathematician, formulated the concept of affine connection around 1880. Of course, Newstein had to adopt a four-dimensional treatment of space and time in order to carry out his reformulation of Newtonian gravitation theory; but, long before that, the concept of time as a fourth dimension had been introduced in analytical mechanics by d'Alembert and Lagrange.[3]

Continuing my mythical account, when in 1907 Einstein turned to the problem of extending his original (later called special) theory of relativity to include gravitation, Newstein had already shown how to describe the inertio-gravitational field by a non-flat affine connection. Einstein's problem was to combine this insight about the nature of gravitation with the new chrono-geometrical structure of spacetime that he had introduced in 1905. Once the problem is posed in this way, the step from Newstein's formulation of the gravitational field equations to the corresponding equations of Einstein's general relativity is a short one (see Section 8).

Of course, all of this is pure fable; but I believe that—in addition to their entertainment value—such scientific fables are of real value for the history and philosophy of science. First of all, they help us to combat the impression of inevitability often attached to the actual course of historical development, the idea that the "discovery" of a theory is just that: the bringing to light by the intellect of some pre-existing structure, previously hidden but predestined to emerge sooner or later and enter into the scientific corpus in just the form in which it actually did. Secondly, they help us to question the thesis that the formulation of a theory is more-or-less independent of its mode of discovery, including the peculiarities of the individual(s) who happened to "discover" it and the process of negotiation that led to its assimilation into the body of accepted knowledge by the scientific community. Such questions can lead to a more critical re-examination of the current formulation(s) of the theory. We are bound to look more critically at what actually happened, and at the accepted formulation(s) of a theory, if we can produce one or more credible scenarios showing how things might have happened quite differently.[4]

3 This is no myth. See my article on "Space-Time," in (Stachel Forthcoming).

2. POSTMATURE CONCEPTS AND THE ROLE OF ABSENCE IN HISTORY

Zuckerman and Lederberg have suggested that, just as there are premature discoveries, "there are postmature discoveries, those which are judged retrospectively to have been 'delayed'" (Zuckerman and Lederberg 1986, 629).[5] I wish to apply the concept of postmaturity to theoretical entities; but since, as noted above, the word "discovery" might suggest a Platonist attitude to mathematical and physical concepts, I shall use more epistemologically neutral phrases: "postmature development," "postmature concept," "postmature theory," etc.

As the work of Zuckerman and Lederberg suggests, in retrospect one can see that—like other forms of absence—the absence of a postmature concept can play a crucial role in the dialectical interplay that shapes the actual course of historical development. My use of word "dialectical" here is purposeful. The second chapter of Roy Bhaskar's book on dialectics (Bhaskar 1993)[6] is entitled: "Dialectic: The Logic of Absences." He equates *absence* with what he calls *real negation*, whose "primary meaning is real determinate absence or non-being (i.e., including non-existence" (Bhaskar 1993, 5). He describes real negation as:

> the central category of dialectic, whether conceived as argument, change or the augmentation of (or aspiration to) freedom, which depends upon the identification and elimination of mistakes, states of affairs and constraints, or more generally ills—argued to be absences alike (Bhaskar 1993, 393).

Elsewhere I shall argue for this viewpoint with examples drawn from the history of music as well as the history of science. But to return to the central concern of this paper, my claim is that "affine connection" is a postmature concept, the absence of which during the course of development of the general theory of relativity had a crucial negative influence on its development and subsequent interpretation. Conversely, the filling of that absence opened the way to a deeper understanding of the nature of gravitation and of its relation to other gauge field theories of physics.

3. A LITTLE HISTORY

Gauss first developed the theory of curved surfaces embedded in Euclidean three-space, including the concepts of intrinsic (or Gaussian) and extrinsic curvature. But he defined these concepts in a way that did not depend on the concept of parallelism.[7] The development of differential geometry had proceeded quite far by the time Rie-

---

4 See (Stachel 1994a) and, for other examples from the history of relativity, (Stachel 1995). For some further comments on alternative histories, see the final section, "Acknowledgements and a Critical Comment."

5 I am indebted to Gerald Holton for drawing my attention to this paper, which fills a gap in my earlier presentations of Newstein's story.

6 I regard Bhaskar's work on critical realism as the most significant attempt at a modern Marxist approach to the philosophy of science (see Stachel 2003a). For a critical introduction to Bhaskar's work, see (Collier 1994).

mann introduced the concept of a locally Euclidean manifold with curvature varying from point to point in 1854, first published in (Riemann 1868).[8] So the idea of starting with a geometrical structure defined in the infinitesimal neighborhood of a point of a manifold and proceeding from the local to the global structure was quite familiar by the last third of the nineteenth century.

Similarly, discussions of the concept of parallelism had played a central role in the development of non-Euclidean geometry in the first half of the nineteenth century.[9] Grassmann's work on affine geometry had abstracted the concepts of parallel lines, plane elements, etc., from their original three-dimensional, Euclidean contexts.[10] Few were aware of the first (1844) edition of the *Ausdehnungslehre*, or even of the second version in 1862; but after the publication of the second edition of the 1844 version in 1878, knowledge of his work began to spread among mathematicians, so that it was widely available to them by the last two decades of the century.[11] By this time, there was already a rich literature on the geometrical interpretation of the principles of mechanics for systems with n-degrees of freedom based on n-*dimensional Riemannian geometry*.[12]

In all this time no one applied Riemann's approach to intervals to the concept of parallelism. Karin Reich has drawn attention to the problem of the delay in the extension of the local approach in geometry to the concept of parallelism:

> Parallelism was and is thus a central theme for the foundations of geometry. Yet it is missing in Bernhard Riemann's Habilitation Lecture "On the Foundational Hypotheses of Geometry," indeed the word parallel does not occur here. Also in the succeeding period of rapidly occurring development of Riemannian geometry parallelism was not a theme. Perhaps this is one of the reasons why Riemannian geometry was not unconditionally accepted by pure geometers (Reich 1992, 78–79).[13]

---

7 Essentially, he defined the intrinsic curvature at a point of a surface in a way that seemed to depend on the embedding of the surface—in terms of the radius of curvature of the sphere that best fits the surface at the point in question—and then proved that the result really does not depend on the embedding. See (Gauss 1902), and for a modern discussion (Coolidge 1940, Book III, chap. III, 355–387).

8 For the history of differential geometry, see (Struik 1933; Coolidge 1940; Laptev and Rozenfel'd 1996, sec. 1: "Analytic and Differential Geometry," 3–26).

9 For the standard older historical-critical account of non-Euclidean geometry, see (Bonola 1955).

10 See (Grassman 1844; 1862; 1878), and for an English translation, (Grassmann 1995). For a survey of publications using Grassmann's approach, demonstrating that their number increased considerably after 1880, see (Crowe 1994, chap. 4); by the end of the century, interest in Grassmann's work was comparable to that in Hamilton's. Weyl was well aware of Grassmann's work. Speaking of affine geometry, he says: "For the systematic treatment of affine geometry with abstraction from the special 3-dimensional case, Grassmann's "Lineale Ausdehnungslehre" (Grassmann 1844)... is the groundbreaking work" (Weyl 1923, 325). In a recent discussion of Grassmann's role as a forerunner of category theory, Lawvere (Lawvere 1996) speaks of "the category A of affine-linear spaces and maps" as "a monument to Grassmann" (p. 255).

11 For a study of Grassmann and his influence, see (Schubring 1996).

12 See (Lützen 1995a; 1995b) for surveys of some of this work.

13 Readers of this work will realize the extent of my indebtedness to Karen Reich's work. I also gratefully acknowledge several helpful discussions with Dr. Reich.

Her retrospective critical judgement a century later is borne out by the contemporary evaluations of those who filled that gap in 1916–1917: Hessenberg, Levi-Civita, Weyl and Schouten.

Hessenberg's paper (Hessenberg 1917) was actually the first such paper, dated June 1916. It starts with a reference to relativity: "Because of the significance that the theory of quadratic differential forms has recently attained for the theory of relativity, the question of whether and how the elaborate and difficult formal apparatus of this theory can be simplified, if not bypassed, gains new significance (p. 187)." Speaking of "Christoffel's well-known transformational calculus," Hessenberg states that his aim is to "replace [it] with a geometrical argument (p. 187)." He criticizes the "formal methods of formation" of various quantities that occur because they do not bring out "the essentially *intuitive* [*anschaulich*] *meaning* of the invariants and covariants needed for the geometrical and physical applications" (p. 191). He stresses the role of Grassmann. "Access [to their geometrical significance] is opened in a way that, to me, seems surprisingly simple by means of Grassmann's ideas" (p. 192).

Levi Civita's paper (Levi Civita 1916), which is dated November 1916, also starts with a reference to Einstein's work:

> Einstein's theory of gravitation ... regards the geometrical structure of space ... as depending on the physical phenomena that take place in it ... The mathematical development of Einstein's magnificent conception ... involves as an essential element the curvature of a certain four-dimensional manifold and the related Riemann symbols [i.e., the curvature tensor] ... Working with these symbols in questions of such great general interest has led me to investigate if it is not possible to simplify somewhat the formal apparatus that is usually used to introduce them and to establish their covariant behavior. Such an improvement is indeed possible ... [This work] started with that sole objective, which little by little grew in order to make room for the geometrical interpretation [of the Riemannian curvature]. At the beginning I thought to have found it in the original work of Riemann ... ; but it is there only in embryo. ... [O]ne gets the impression that Riemann really had in mind that intrinsic and invariant characterization of the curvature, which will be made precise here. On the other hand, however, there is not a trace, either in Riemann or in Weber's commentary, of those specifications (the concept of parallel directions in an arbitrary manifold and consideration of an infinitesimal geodesic quadrilateral with two parallel sides) that we recognize to be indispensable from the geometrical point of view (pp. 173–174).

Reich comments:

> With this word "indispensible" Levi-Cività recalled Luigi Bianchi's characterization of Ricci's absolute differential calculus. Bianchi had characterized this in 1901 as "useful but not indispensable" (Reich 1992, 79–80).

Weyl (1918b) states:

> The later work of Levi-Cività [1916], Hessenberg [1917], and the author [Weyl 1918a][14] shows quite plainly that the fundamental conception on which the development of Riemann's geometry must be based if it is to be in agreement with nature, is that of the infinitesimal parallel displacement of a vector.[15]

14 For a discussion of this and the succeeding editions of Weyl's book, see the next section.

After the introduction of Riemannian parallelism by Hessenberg and Levi-Civita (and, again independently in (Schouten 1918)), it was but a brief and natural step to its generalization. Since the abstraction (in the large) of affine parallelism from parallelism in Euclidean geometry had already been made, the abstraction (in the small) of affine parallelism from parallelism in a Riemannian manifold is immediately suggested by the analogy. Indeed, Weyl took that step just a year later: In (Weyl 1918a) he defines an affinely connected manifold.[16]

The evidence thus indicates that both the Riemannian concept of parallelism and its affine generalization were introduced *postmaturely*. The absent concept of Riemannian parallelism could have been filled at any time during the last third of the nineteenth century, and followed quickly by the introduction of the concept of an affinely connected manifold, since it is a natural generalization of the Grassmannian "*lineale Ausdehnungslehre*."

Indeed, Grassmann himself might have accomplished these tasks. Towards the end of his life he learned about the work of Riemann and Helmholtz, and one of his last publications (Grassmann 1877) discusses the relation of their work to his *Ausdehnungslehre*. He discusses a method of introducing such non-linear geometries that amounts essentially to defining them as subspaces of linear spaces of higher dimensions. The path that Levi-Civita initially took to the definition of Riemannian parallelism was based on embedding a Riemannian space in a Euclidean space of sufficiently high dimension. Had Grassmann lived longer, it is conceivable that he might have introduced the concept of affine parallelism by a similar method (see the discussion in the Appendix). But he died in the same year that he wrote this paper; so I have been forced to invent Weylmann, the mathematician who introduces the concept of an affinely connected manifold around 1880, neither prematurely nor postmaturely.

4. EQUIVALENCE PRINCIPLE AND AFFINE CONNECTION

It was Albert Einstein who first realized the profound significance of the equality of inertial and gravitational mass. He soon began to speak of inertia and gravitation as "*wesensgleich*": essentially the same in nature. By an acceleration of the frame of reference, the division between inertial and gravitational "forces" can be altered, and indeed by a suitably chosen acceleration the combination of both can even be made to vanish at any point of spacetime.

Einstein's problem was to find the way to incorporate this physical insight into the mathematical structure of gravitation theory. After the development of the concept of affine connection, the way became clear: there is an inertio-gravitational field, repre-

15 Translated from (Weyl 1923, 202).

16 For references and discussion of the work of Levi-Civita, Hessenberg, Schouten and Weyl, see the indispensible (Reich 1992). For the background to Weyl's "Purely Infinitesimal Geometry," see (Scholz 1995). I am indebted to Dr. Erhard Scholz for a discussion of this work.

sented mathematically by a symmetric connection in spacetime, which incorporates this essential unity in its very nature. We can see the development of this insight by looking at the various editions of Weyl's *Raum-Zeit-Materie*. In (Weyl 1918b), Levi-Civita's concept of parallel transport, based upon the embedding of a Riemann space in a flat Euclidean space of higher dimension, is freed from this dependence by giving it an intrinsic definition. Weyl further states that the Christoffel symbols represent the gravitational field. In (Weyl 1919)—which follows the argument of Weyl (1918a)—the concept of parallel transport is freed from its dependence on the metric field by the introduction of the concept of affine connection. Weyl (1921) refers to this connection as the "guiding field" (*Führungsfeld*), incorporating the effects of both gravitation and inertia on the motion of bodies.

Soon afterwards, Cartan (1923) drew the obvious conclusion: By incorporating the equivalence of gravitation and inertia into Newton's gravitation theory, it can be formulated in terms of a Newtonian affine connection. Since then, starting with (Friedrichs 1927) and culminating—but certainly not ending—in (Ehlers 1981), a series of refinements of Cartan's approach have brought the affine version of Newton's theory to a state of considerable mathematical perfection.

However, I shall not give the most, abstract, coordinate-free characterization of the Newtonian affine connection based on the simplest set of axioms. For our purposes, it will be more useful to show how, starting from the usual form of the Newtonian theory of gravitation, the components of the connection with respect to a physically chosen basis may be defined, thus suggesting how Newstein could have proceeded—had he only existed![17]

5. NEWSTEIN'S WORLD

We shall start from the usual formulation of Newtonian gravitation theory in some inertial frame of reference (ifr, for short). Events in this frame are individuated with the help of the Newtonian absolute time t (chronometry), and three Cartesian coordinates (i.e, assuming Euclidean geometry), fixed relative to some choice of origin O and of three mutually perpendicular axes.[18] Since inertial and gravitational mass are equal, if $\boldsymbol{g}$ represents the force/unit gravitational mass, the equation of motion of a (structureless) particle will be

$$\boldsymbol{a} = \boldsymbol{g}, \tag{1}$$

where $\boldsymbol{a}$ is the acceleration of the particle with respect to the chosen ifr.

---

17 See (Stachel 1994b) for a somewhat more abstract discussion of spacetime structures in Newton-Galilean and special-relativistic spacetimes (i.e., in the absence of gravity), and in Newtonian and Einsteinian gravitational theories.

18 We assume units of time and distance fixed initially and used in all frames of reference, and shall use vector notation, so that, for example, the displacement vector from the origin $\boldsymbol{r} = (x^1, x^2, x^3)$, the velocity $\boldsymbol{v} = d\boldsymbol{r}/dt$, the acceleration $\boldsymbol{a} = d\boldsymbol{v}/dt$, etc, all with respect to the ifr, are denoted by boldface symbols.

Now consider transformations to another frame of reference, moving linearly with respect to the first:

$$\boldsymbol{r}' = \boldsymbol{r} - \boldsymbol{R}(t), \quad t' = t. \tag{2}$$

If the velocity vector $\boldsymbol{V} = d\boldsymbol{R}/dt$ is *constant*, then the transformation is to another inertial frame of reference, and the equation of motion, eq. (1), is invariant under such a transformation. That is, both $\boldsymbol{a}$ and $\boldsymbol{g}$ are invariant under such *Galilei transformations* from one inertial frame to another.

However, if $\boldsymbol{V}$ is *not* constant, then the transformation is to some linearly accelerated (rigid) frame of reference, and differentiation of eq. (2) twice with respect to the time gives

$$\boldsymbol{a}' = \boldsymbol{a} - \boldsymbol{A}(t), \quad \boldsymbol{A}(t) = d^2\boldsymbol{R}(t)/dt^2. \tag{3}$$

In Newtonian mechanics, "true" forces, such as $\boldsymbol{g}$, are assumed to be the same in all frames of reference. To compensate for the use of a non-inertial frame of reference, so-called "inertial forces" appear in the equations of motion (such forces might better be called "non-inertial"). Indeed, when we substitute eq. (3) in eq. (1), we get:

$$\boldsymbol{a}' + \boldsymbol{A} = \boldsymbol{g}, \text{or } \boldsymbol{a}' = \boldsymbol{g} - \boldsymbol{A}, \tag{4}$$

and $-\boldsymbol{A}(t)$ appears as such an "inertial force" in the equation of motion of a particle with respect to a linearly accelerating frame.

But, one may ask, if we carry out measurements in some frame of reference, and get an acceleration, let us say $\boldsymbol{a}'$, for a test particle, how do we separate it into its components, the "true force" $\boldsymbol{g}$ and the "inertial force" $-\boldsymbol{A}$? Newton would not have hesitated a moment in answering: Look for the sources of the gravitational force, and use the inverse square law to compute the total $\boldsymbol{g}$ at the point where the test particle is located. Alternatively, he might have proposed: Look at the center of mass of the "system of the world" (i.e., the solar system) and see whether you are accelerating relative to it to find $\boldsymbol{A}$.

But by the end of the nineteenth century, under the influence of Maxwell's electromagnetic theory, the field point of view towards forces was beginning to prevail; according to this viewpoint, one should look upon the gravitational field as the conveyor of all gravitational interactions between massive bodies. Accordingly, the local gravitational field at a point in space (and an instant of time) should always be ascertainable by means of local measurements in the neighborhood of that point. Now, in the case of any other force but the gravitational, there would be no obstacle to separating out the inertial from the non-gravitational effects. For electrically charged particles, for example, one would merely vary the ratio of electric charge to inertial mass: The electric force would vary with this ratio, the inertial force would not. But the ratio of gravitational charge (= gravitational mass) to inertial mass is just what *cannot* be varied—the invariance of that ratio is the primary empirical basis of the equivalence principle.

So the answer to our question is: Once we adopt the field point of view about gravitation, there is no way (locally) to distinguish inertial from gravitational effects.

We have to recognize that there is an *inertio-gravitational field*, and that how this field divides up into inertial and gravitational terms is not absolute (i.e., frame-independent), but depends on the state of motion (in particular the acceleration) of the frame of reference being used. Indeed, we see that, by choosing the value of $\boldsymbol{A}$ to coincide numerically with the value of $\boldsymbol{g}$ at some point, we can make the total inertio-gravitational field vanish at that point. Indeed, this is why we did not call it an inertio-gravitational force: Although the values of their components with respect to some frame of reference can change depending on the state of motion of that frame, non-vanishing force fields at a point, such as the electric and magnetic fields making up the electromagnetic field, cannot be made to all vanish by any change of reference frame.

Another consequence of our new, equivalence-principle viewpoint is that a basic distinction between inertial and linearly accelerated frames of reference is no longer tenable. Any rigid non-rotating frame of reference is just as good as any other.

Let us now inventory what is left after we adopt this new viewpoint:

1. the absolute time, assumed to be measurable by ideal clocks; its measurement is unaffected by the presence of an inertio-gravitational field (compatibility of chronometry with the inertio-gravitational field);
2. Euclidean geometry, which holds within each frame in the class of three-dimensional, non-rotating frames of reference; it is assumed to be measurable with ideal measuring rods; its measurement is unaffected by the presence of the inertio-gravitational field (compatibility of geometry with the inertio-gravitational field).
3. Since gravitation and inertia are no longer (absolutely) distinguished (i.e., gravity is no longer regarded as a force), the set of "force-free" inertial motions is replaced by a set of "force-free" inertio-gravitational motions. One of these is determined by specifying a velocity vector at a point of space and an instant of time. The vector is then the tangent to the "freely falling motion" through the point at this instant.
4. While the inertio-gravitational field $\boldsymbol{g}(\boldsymbol{r}, t)$ is not absolute (i.e., it depends on the frame of reference used, and only behaves like a vector with respect to transformations within a given frame of reference), its spatial derivatives $\partial_m g^n(\boldsymbol{r}, t)$ are independent of the (non-rotating) reference frame. Physically, these differential gravitational forces are usually designated as the tidal forces, since they are responsible for the tides, among other effects. The matrix of these quantities determines the relative acceleration of two freely falling test particles, i.e., the acceleration of one particle with respect to the other. The components of the tidal forces therefore may be evaluated by measurement of the components of this relative acceleration.

6. THE NEWTONIAN CONNECTION

Now we are ready to make the transition to the four-dimensional point of view, in which a point of spacetime is specified by the four coordinates (t, x^1, x^2, x^3) or $(t, \boldsymbol{r})$ for short, where x^1, x^2, x^3 are the Cartesian coordinates of the point with respect to some non-rotating frame of reference and t is the absolute time.[19] We shall refer to these as *adapted coordinates* for this frame of reference. The absolute time gives a *foliation* of spacetime, i.e., a family of non-intersecting hypersurface that fills the spacetime. In the adapted coordinate system the foliation consists of the hypersurfaces $t = \text{const}$. A vector is said to be *space-like* if it is tangent to a hypersurface of the foliation; a vector is *time-like* if it is not space-like. Any curve, the tangent vector to which is always time-like, is a *time-like* curve, with a similar definition for space-like curves. In adapted coordinates a vector is time-like if it has a non-vanishing time component, space-like if it does not.

We can use any (three-)velocity field $\boldsymbol{v}(t)$ to rig the hypersurfaces of constant time: Define a time-like four-velocity field $V(t)$, with t-component = 1 and spatial components equal to those of $\boldsymbol{v}(t)$ in adapted coordinates. Thus, $V(t)$ defines a congruence of time-like curves that fills spacetime. Indeed, we need merely select *one* such time-like curve $V(t)$ and then parallel propagate it along each hypersurface $t = \text{const}$ to get this congruence. In particular, the paths of the points $x^1, x^2, x^3 = \text{const}$, parametrized by the absolute time t, constitute such a congruence; Euclidean geometry holds for these spatial coordinates at all times. Thus we have specified the chronometry and the geometry of the initial frame of reference using the adapted coordinates.

Any $V(t)$ field provides a rigging of each hypersurface (see the discussion of rigged hypersurfaces in the Appendix). Just as a rigging was needed to go from the flat affine connection of the enveloping space to the non-flat affine connection of a hypersurface embedded in it, a rigging is needed here to relate the flat (Levi-Civita) connection on each Euclidean hypersurface to the four-dimensional non-flat connection that we want to define for spacetime as the mathematical representation of the inertio-gravitational field.

Indeed, we can define a unique symmetric, four-dimensional affine connection on the spacetime by requiring that it satisfy the following conditions:

1. The absolute time is the affine parameter for all time-like geodesic paths. A geodesic path that is time-like at any of its points is time-like at all its points.
2. There is a flat, Euclidean connection on each (three-dimensional) hypersurface of the foliation. Hence, the Euclidean distance is the affine parameter for each space-like geodesic path. A geodesic path that is space-like at any of its points is space-like at all its points.

---

19 We shall designate a time component by a sub- or superscript "t," and spatial components by sub- or superscript "$i, j, k\ldots$" or other lower-case Latin letters.

3. The three-dimensional and the four-dimensional treatments of the spatial geometry on each hypersurface are consonant with each other: The Euclidean (flat) three-dimensional affine connection on each hypersurface of some frame of reference coincides with the connection induced on that hypersurface by the four-dimensional connection when that hypersurface is rigged with *any* time-like $V(t)$ field.[20]
4. Parallel transport of any space-like vector is path-independent. By picking an orthonormal triad $\boldsymbol{e}_i$ of such vectors at some point on an initial hypersurface of the foliation, and parallel transporting the triad along any time-like curve with tangent vector $V(t)$, a frame of reference is generated: Once it is parallel transported to a point on another hypersurface of the foliation, the triad can be propagated to any other point of the hypersurface by (path-independent) parallel transport.
5. If we add any $V(t)$ to the triad field e_i, now interpreted as four-vectors, we get a four- dimensional frame of reference.[21] In any such frame of reference, any path that obeys the Newtonian gravitational equation of motion of a structureless test particle shall be a time-like geodesic of the four-dimensional connection parametrized by the absolute time. The spatial projection of its four-dimensional tangent vector onto any hypersurface of the foliation will coincide with the three-velocity of the test particle on that hypersurface.

As indicated earlier, we have not attempted to give a minimal list of assumptions, each of which is independent of the others; but rather, a physically intuitively plausible list. We now proceed to derive the components of the connection in some given non-rotating frame of reference, i.e., using coordinates adapted to the tetrad of basis vectors $V(t), e_i$ that characterize this frame of reference.

The equation of a geodesic in these coordinates is (see the Appendix):

$$d^2x^\kappa/d\lambda^2 + \Gamma^\kappa_{\rho\sigma}(dx^\rho/d\lambda)(dx^\sigma/d\lambda) = 0, \quad (\kappa, \rho, \sigma = t, x^1, x^2, x^3), \tag{5}$$

where λ is an affine parameter, i.e., the (four-dimensional) tangent vector to the curve $P(\lambda)$ is equal to $dP/d\lambda$; and the components of the connection are with respect to the chosen four-dimensional frame of reference. If we consider time-like geodesics, condition 1) requires that t be an affine parameter for all of them. The four-velocity $dx^\rho/d\lambda$ will thus have components $(1, \boldsymbol{v})$ in the adapted coordinate system, where $\boldsymbol{v}$ is the three-velocity of the particle. Considering only the t- component of eq. (5) for the moment, in adapted coordinates it takes the form:

---

20 If the requirement is fulfilled for one such field it is fulfilled for any such field, since two such fields can only differ by a *space-like* acceleration vector field. So the transition from one non-rotating frame of reference to another, which corresponds mathematically to a change of $V(t)$ field, does not affect the result.

21 Note that any such $V(t)$ field commutes with the three e_i fields, which commute with each other, so that they form a holonomic basis; so coordinates adapted to this basis will always exist.

$$d^2t/dt^2 + \Gamma^t_{tt} + 2(\Gamma^t{}_{ti})v^i + (\Gamma^t_{ij})v^iv^j = 0, \tag{5a}$$

and since the first term vanishes, the only way that eq. (5a) can hold for all values of v^i is if Γ^t_{tt}, Γ^t_{ti} and Γ^t_{ij} all vanish in the adapted coordinate system. In other words, these are the mathematical conditions that assure the compatibility of the chronometry and the inertio-gravitational field. Physically, this means that an ideal clock moving around in the inertio-gravitational field will always measure the absolute time.

Conditions 2, 3, and 4 now demand that the three space-like vectors e_i, which lie along the coordinate axes and thus have components δ^μ_i in adapted coordinates, have vanishing covariant derivates with respect to both the Euclidean (flat) three-dimensional connection on each hypersurface, and the non-flat inertio-gravitational four-dimensional connection. By a similar argument to that above, these conditions result in the vanishing of Γ^m_{ti} and Γ^m_{ni} in the adapted coordinate system. In other words, these are mathematical conditions that assure the compatibility of the geometry and the inertio-gravitational field. Physically, this means that an ideal measuring rod moving around in the inertio-gravitational field will always measure the Euclidean distance.

Condition 5 now fixes the values of the only remaining non-vanishing components of the affine connection, Γ^m_{tt}, in the adapted coordinate system. Returning to eq. (5), its spatial components in the adapted coordinate system now take the form:

$$d^2x^m/dt^2 + \Gamma^m_{tt} = 0, \tag{5b}$$

all other terms in the equation vanishing because of the previously-established vanishing of the other components of the connection. We see that we need merely set:

$$\Gamma^m_{tt} = -g^m(\boldsymbol{r}, t) \tag{6}$$

in the adapted coordinates in order to have the geodesic equation coincide with the equation of motion of a particle in the gravitational field $\boldsymbol{g}(\boldsymbol{r}, t)$.

We have now fixed all the components of the symmetric affine connection in the adapted coordinate system. We need merely apply the general transformation law for the components of the connection under a coordinate transformation $x^{\kappa'} = x^{\kappa'}(x^\kappa)$:

$$\Gamma'^{\kappa}_{\mu\nu} = \Gamma^{\kappa}_{\mu\nu}(\partial x^{\kappa'}/\partial x^{\kappa})(\partial x^{\mu}/\partial x^{\mu'})(\partial x^{\nu}/\partial x^{\nu'}) + \partial^2 x^{\kappa'}/\partial x^{\mu}\partial x^{\nu}, \tag{7}$$

to the equations for a linearly accelerated transformation (see eq. (2) of Section 5):

$$x^{m'} = x^m + R^m(t), \tag{8}$$

in order to see that the components of the connection transform correctly; i.e, that all the components but Γ'^m_{tt} continue to vanish, and the Γ^m_{tt} transform just like the components of $\boldsymbol{g}$ under such a transformation (see Section 5, eq. (4)). If we carry out a transformation to a rotating system of coordinates, the transformation of the components of the connection introduces terms that correspond to the Coriolis and centripetal "inertial forces" that must be introduced in a rotating coordinate system. To get the form of the components of the connection in an arbitrary coordinate system, one need merely apply eq. (7) to an arbitrary coordinate transformation.

What about the tidal forces, which as mentioned above are absolute? They are represented by the appropriate components of the Riemann tensor, which can be computed from the Newtonian inertio-gravitational connection. Since they are components of a tensor, they indeed possess an absolute character, in the sense that if the components do not all vanish at a point, no change in frame of reference at that point can make them all vanish. These components of the Riemann tensor enter into the equation of geodesic deviation, which describes in four-dimensional tensorial form the relative acceleration of two particles falling freely in the inertio-gravitational field; but I shall not enter into details here.

Rather, I turn to the question of the field equations for the inertio-gravitational field. The Newtonian field $\boldsymbol{g}$ obeys the field equation:

$$\nabla \cdot \boldsymbol{g} = 4\pi G\rho, \tag{9}$$

where G is the Newtonian gravitational constant, ρ is the mass density of the material sources of the gravitational field, and $\nabla \cdot \boldsymbol{g}$ is the trace of the tidal force matrix. If one works out the components of $R_{\mu\nu}$, the contracted Riemann or Ricci tensor, in the adapted coordinates, it turns out that only R_{tt} is non vanishing, and it equals $-\nabla \cdot \boldsymbol{g}$. So $R_{tt} = -4\pi G\rho$ and all other components =0 in the adapted coordinates. The only remaining problem is to write this result as a tensorial equation, independent of coordinate system; but this is easily solved by introducing a covariant vector field T_μ, such that in adapted coordinates $T_\mu = \partial_\mu t = \delta^\mu_t$. The gravitational field equations now take the tensorial form:

$$R_{\mu\nu} = 4\pi G\rho T_\mu T_\nu, \tag{10}$$

which is clearly of the same form in all coordinate systems.

In a more complete treatment,[22] one would have to go a step further: the Newtonian gravitational field $\boldsymbol{g}$ can be derived from a gravitational potential function ϕ: $\boldsymbol{g} = -\nabla\phi$, and this condition can be expressed intrinsically in terms of the properties of the corresponding Riemann tensor (the tidal force matrix introduced in Section 5, which is closely related to certain components of the Riemann tensor, becomes symmetric). Now ϕ plays an important role in taking the Newtonian limit of general relativity, but since we shall not discuss this issue, I can forego entering into further consideration of details.

The non-dynamical Newtonian chrono-geometrical structures, consisting of the absolute time and the relative spaces of the family of non-rotating frames of reference, are unmodified by the presence of gravitation. Mathematically, they are represented by a closed temporal one-form (the T_μ introduced above) and a trivector field

22 See (Stachel 2003b) for such a treatment.

whose transvection with the one-form vanishes (the e_i introduced above), from which a degenerate (rank 3) spatial "metric" may be constructed ($=\delta^{ij}e_ie_j$).

However, the compatible flat inertial structure of Newton's theory is modified. It becomes a dynamical structure, the Newtonian inertio-gravitational field, which remains compatible with the chrono-geometrical structures. Mathematically, it is represented by a symmetric affine connection (the Newtonian connection $\Gamma^{\kappa}_{\rho\sigma}$ discussed above), which can be derived from a "connection potential" (the ϕ discussed above). Its contracted Riemann tensor obeys field equations that relate it to the masses acting as its source (eq. (10) above). The compatibility of this connection with the chrono-geometrical structure means, as noted earlier, that clocks and measuring rods freely falling in the inertio-gravitational field still measure absolute temporal and spatial intervals, respectively. Mathematically, this is expressed by the vanishing of the covariant derivatives of the temporal one-form and degenerate spatial "metric" with respect to the Newtonian connection.

7. SOME MYTHICAL HISTORY: NEWSTEIN MEETS WEYLMANN

Once the concept of affine connection has been developed and the Riemann tensor geometrically interpreted in terms of parallel transport around closed curves, this version of Newton's theory—which converts gravitation from a force that pulls bodies off their (non-dynamical) inertial paths, into a (dynamical) modification of the (inertial) affine connection—is almost immediately suggested by the equality of gravitational and inertial mass. Indeed, shortly after the mythical mathematician Weylmann formulated the concept of affine parallelism, his equally mythical physicist colleague Newstein developed this reinterpretation of Newtonian gravitational theory. Brooding on the equality of gravitational and inertial mass, he became convinced of the essential unity of gravitation and inertia. Originally, he expressed this insight in the usual three-plus-one language of physics, treating space and time separately (see Section 5). He considered uniformly accelerated frames of reference in the absence of gravitation (the Newstein elevator!), and decided it was impossible to distinguish such a frame of reference from a non-accelerated frame with a constant gravitational field. This led him to consider transformations between linearly accelerated frames of reference.

He was puzzled by the strange transformation law that he had to introduce for the gravitational "force," which no longer behaves like a vector under such transformations. At some point he turned to Weylmann, who soon realized that the gravitational "force" transforms like the Γ^{m}_{tt} components of a four-dimensional affine connection, and that Poisson's law for the gravitational potential could be written as an equation linking the Ricci tensor of the connection with its material sources (see Section 6). In the now-famous Newstein-Weylmann paper, the two developed a four-dimensional geometrized formulation of Newtonian gravitation theory, which generalized Newtonian chrono-geometry to include linearly accelerated frames and a dynamized inertio-gravitational connection field, but still included the concept of absolute time.

In so far as they took any notice of this work, their contemporaries regarded it as an ingenious mathematical *tour-de-force*. But, since it had no new physical consequences, it did not much impress Newstein's positivistically-inclined physics colleagues.

Weylmann analyzed the invariance group of the new theory, which is much wider than that of the older Newtonian kinematics. The privileged role of the inertial frames of reference in Newton's theory, just beginning to be realized thanks to the work of Lange and Neumann, was lost in the new interpretation of gravitation.While rotation remained absolute (in the sense that all components of the connection representing centrifugal and Coriolis forces could be made to vanish globally by a coordinate transformation), all linearly accelerated frames of reference were now equal, and the significance of this occasioned a discussion among a few philosophers of science who concerned themselves with the foundations of mechanics. Ernst Mach added a few lines about Newstein to the latest edition of his *Mechanik*.

8. MORE MYTH: EINSTEIN CONFRONTS NEWSTEIN

Perhaps this is where Albert Einstein first read about Newstein's work. At any rate, in 1907, pursuant to his commission to write a review article on the physical consequences of his 1905 work on the relativity principle (now becoming known as the theory of relativity),[23] he turned his attention to gravitation, and (like Newstein) was struck by the equality of gravitational and inertial mass. He realized that, as a consequence, in Newtonian mechanics there is a complete equivalence between an accelerated frame of reference without a gravitational field and a non-accelerated frame of reference, in which there is a constant gravitational field. He soon generalized this to what he later called the principle of equivalence: There is no physical difference (mechanical or otherwise) between the two frames of reference.[24]

Recalling what he had read about Newstein, Einstein realized that he had rediscovered the loss of the privileged role of inertial frames once gravitation is taken into account. Like Newstein, he became convinced that inertia-cum-gravitation must be represented mathematically by an affine connection; but now this representation somehow must be made compatible with the new chronogeometry he had developed in his 1905 theory.[25] He first tried to preserve the non-dynamical nature of this chrono-geometrical structure—which Minkowski soon expressed in terms of a four-dimensional pseudo-Euclidean geometry—by developing various special-relativistic gravitational theories that incorporated the unity of gravitation and inertia by the very fact that they were based upon an affine connection. But the Riemann tensor of the inertio-gravitational connection in each of these theories was non-vanishing, while

23 For a translation of this paper, see (Stachel 1998).

24 Aside from the first sentence, this paragraph is a summary of the actual historical circumstances of Einstein's first work on gravitation, see (Einstein 1907). The fantasy begins in the next paragraph.

25 In the frame bundle language, the physically preferred subgroup of the general linear group had to be changed from the Newtonian group to the Lorentz group.

the metric-affine structure of Minkowski spacetime is flat. Physically, this meant that the inertio-gravitational and chrono-geometrical structures were not compatible: Good clocks and measuring rods, as defined by the chrono-geometrical structure, did not keep the proper time or measure the proper length when moved about in the gravitational field.

While this could be "explained away" as due to a universal distorting effect of gravitation on all measuring rods and clocks, something about such an explanation disturbed him. Since the effect was universal, the "true" Minkowski chronogeometry could be shown to have no physically observable consequences.

Finally, he realized what was bothering him: This type of explanation was all too similar to Lorentz's interpretation of the Lorentz transformations: Galilean chronogeometry is the "true" one; but the universal effect of motion through the absolute (aether) frame of reference exerts a universal effect on all physical processes that prevents any physically observable consequences of this motion. What was the way out of this new unobservability dilemma?

Suddenly the answer struck him: If he required compatibility between the inertio-gravitational and chrono-geometrical structures, the problem would disappear, just as it had in Newstein's reinterpretation of Galilean kinematics. Good measuring rods and clocks, as defined by such a chrono-geometrical structure, would measure the true proper lengths and times wherever they were placed in the inertio-gravitational field. But there was a price to pay for this compatibility: The chrono-geometrical field could no longer be flat. It would have a curvature attached to it in the Gaussian sense, the one that Riemann originally had generalized from two to an arbitrary number of dimensions. In this theory, the Riemann tensor would have two distinct (but compatible) interpretations: as the curvature of a connection, associated with parallel transport and the equation of geodesic deviation; and as the curvature of a pseudo-metric, associated with the Gaussian curvature of each of the two-dimensional sections at any point of spacetime.

And of course, since metric and connection were now compatible, this implied that the components of the connection with respect to any basis were numerically equal to the Christoffel symbols of the metric with respect to that basis. And since the connection is a dynamical field, the metric would also have to become a dynamical field. In contrast to the Newsteinian case, where the chrono-geometry remained non-dynamical, in the Einsteinian case, there are no non-dynamical spacetime structures. The bare manifold remained absolute in a certain sense;[26] but then, it had no physical characteristics other than dimensionality and local topology unless and until the iner-

26 I say this because, in actual fact, the global topology of the manifold is not given before the metric-cum-connection field, as implied in so many presentations of general relativity. One actually solves the Einstein field equations on a small patch, and then looks for the maximal extension of that patch compatible with the given metric. Certain criteria for compatibility must be given before the question of maximal extension(s) becomes meaningful, of course. For discussion of this topic, see (Stachel 1986; 1987).

tio-gravitational cum chronogeometrical field was impressed upon it. Least of all do the points of the manifold represent physical events before imposition of a metric.[27]

The new, dynamical theory of spacetime structures had a number of novel physical consequences, and Einstein soon became world-famous—but you know the rest of the story.

9. SOME REAL HISTORY: EINSTEIN WITHOUT NEWSTEIN

Unfortunately, the last section was a historical fable, and the real Einstein had to work out the general theory of relativity in the absence of the concept of affine connection—an absence which, as suggested in Section 2, played a fateful role in the actual development and subsequent history of the theory. It took Einstein without Newstein seven years to develop the general theory of relativity *after* he had adopted the equivalence principle as the key to a relativistic theory of gravitation. Rather than tell the entire story of the many genial steps and equally numerous missteps on Einstein's road from special to general relativity,[28] I shall here just highlight some of the most fateful consequences of the absence of the connection.

First of all, it is important to realize that the tensor calculus, as originally developed by Christoffel, Ricci, Levi Civita and others, was a branch of invariant theory, with only tenuous ties to geometry.[29] Einstein's introduction of the metric tensor field as the mathematical representation of both the chrono-geometry of spacetime and the potentials for the gravitational field did not carry with it most of the geometrical implications that we take for granted today. Insofar as it did carry geometrical implications, notably in fixing the geodesics of the manifold, this had to do with the interpretation of geodesics as the shortest paths (or rather longest, for time-like paths—the twin paradox) in spacetime. The interpretation of geodesics as the straightest paths in spacetime, more important for the understanding of the gravitational field—in particular, the interpretation of the Riemann tensor in terms of the equation of geodesic deviation—had to await the work of Levi Civita and Weyl on parallelism discussed in Section 3.[30] Curvature, in other words, was given the Gauss-Riemann interpretation, rather than the interpretation as the tendency of geodesics to coverge (or diverge), leading to its association with tidal forces.

---

27 For discussion of the hole argument, which bears on this point, see (Stachel 1993) and references therein.

28 See the first two volumes of this series on the development of general relativity. For earlier accounts by this author and others, see (Stachel 1995) and the references therein.

29 "The calculus developed by Gregorio Ricci in the years 1884–1887 had its roots in the theory of invariants, therefore it naturally lacked a geometrical outlook or interpretation, and was so intended by Ricci" (Reich 1992, 79). For the history of the tensor calculus, see (Reich 1994).

30 Interestingly, this interpretation was anticipated by Hertz in his geometrical version of mechanics. See (Hertz 1894) and, for a discussion of the 19th century tradition of geometrical interpretations of mechanics, (Lützen 1995a; 1995b).

It is often said that Einstein, with the help of Grossmann, found ready-to-hand the mathematical tools he needed to develop general relativity: Riemannian geometry and the tensor calculus. But this statement must be taken with a large grain of salt. It would be more correct to say that he had to make do with the tools at hand, with important negative consequences for the development of the theory, and—more importantly for us now—with negative consequences for the interpretation of the theory that continue to exert their effects to this day.[31]

To give two concrete examples of this negative influence on Einstein's work:

1. Until late in 1915, he regarded the derivatives of the metric tensor, rather than the Christoffel symbols, as the mathematical representative of the gravitational-cum-inertial field.[32] In Einstein 1915, he finally corrected this error:

   > These conservation laws [the vanishing of the covariant derivative of the stress-energy tensor] previously misled me into regarding the quantities $1/2\sum_{\mu} g^{\tau\mu}\partial g_{\mu\nu}/\partial x_{\sigma}$ as the natural expression for the components of the gravitational field, although in the light of the formulas of the absolute differential calculus it seems more obvious to introduce the Christoffel symbols instead of these quantities. This was a fateful prejudice (Einstein 1915, 782).

   The reason why this error was so fateful is that it mislead Einstein in his search for the gravitational field equations, a search that took over two years *after* he had adopted the metric tensor field as the mathematical representation of gravity.[33]

2. From 1912 onwards, Einstein expected that, in the Newtonian limit of general relativity, the spatial part of the metric field tensor would remain flat and that the g_{oo} component of the metric would reduce to the Newtonian gravitational potential. Correctly understood, in terms of a formulation of the theory taking the Newtonian limit of both the connection and the metric, these expectations are fulfilled. But one cannot properly take the Newtonian limit of general relativity without the concept of an affine connection, and the corresponding affine reformulation of Newtonian theory discussed in Section 6. Indeed, the problem of correctly taking the Newtonian limit of general relativity only began to be solved in (Friedrichs 1927), and the process was not completed in all details until (Ehlers 1981). In the absence of the affine approach, more-or-less heuristic detours through the weak-field, fast motion (i.e., special-relativistic) limit followed by a slow motion approximation basically out of step with the fast-motion approach, had to be used to "obtain" the desired Newtonian results.[34]

---

31 Perhaps the first such negative influence on work done after the final formulation of the general theory is the ultimate failure of Lorentz's attempt to give a coordinate-free geometrical interpretation of the theory. I thank Dr. Michel Janssen for pointing this out to me. For an account of Lorentz's attempt, see (Janssen 1992).

32 See (Einstein and Grossmann 1913, 7), and (Einstein 1914, 1058), for examples.

33 For details see vol. 1 of this series on the development of general relativity.

34 See (Stachel 2003b) for more details.

Einstein originally thought that he knew the form of the weak field metric in the static case. It involved a spatially flat metric tensor field, with only the g_{oo} component of the metric depending on the coordinates. He used this form of the static metric as a criterion for choosing the gravitational field equations: This form of the metric had to satisfy the field equations, which led to a disastrous result: No field equation based on the Ricci tensor had this form of the static metric as a solution, and Einstein abandoned the Ricci tensor for over two years![35] Had he known about the connection representation of the inertio-gravitational field, he would have been able to see that the spatial metric can go to a flat Newtonian limit, while the Newtonian connection remains non-flat without violating the compatibility conditions between metric and connection. As it was, using the makeshift technique described above to get the Newtonian result, he was amazed to find that the spatial metric is non-flat. Even today, almost all treatments of the Newtonian limit of general relativity are still based on this makeshift approach that employs only the metric tensor.

10. CONCLUSION

The moral of this story is that general relativity is primarily a theory of an affine connection on a four-dimensional manifold, which represents the inertio-gravitational field. The other important spacetime structure is the metric field that represents the chrono-geometry; and the peculiarity of general relativity is that the compatibility conditions between metric and connection—or in physical terms, between inertio-gravitational field and chrono-geometry—uniquely determine the connection in terms of the metric. In teaching the subject, emphasis should be put on the connection from the beginning. This can be done easily by presenting the affine version of Newtonian gravitation theory before discussing general relativity. But most textbooks still start from the metric and introduce the connection later via the Christoffel symbols in a way that does not stress the basic role of the connection.[36] Now that gauge fields have come to dominate quantum field theory, it is more important than ever to emphasize from the beginning how general relativity resembles these Yang-Mills type theories, as well as how it differs.[37]

---

35 For details, see (Stachel 1989; Norton 1984) and volume 1 of this series.

36 It is indicative of current interests that (Darling 1994), the only elementary mathematical textbook I know that introduces the connection first, does not even mention the application to gravitation theory, but concludes with a chapter on "Applications to Gauge Field Theory" (pp. 223–250).

37 The basic difference is that the affine connection lives in the frame bundle (see Section h of the Appendix), which is soldered to the spacetime manifold. The symmetries of the fibres are thus induced by spacetime diffeomorphisms. On the other hand, the Yang-Mills connections live in fibre bundles, the fibres of which have symmetry groups that are independent of the spacetime symmetries (internal symmetries). For further discussion, see (Stachel 2005).

ACKNOWLEDGEMENTS AND A CRITICAL COMMENT

I thank Dr. Jürgen Renn for a thoughtful reading of this paper, and many helpful suggestions for its improvement. I thank Dr. Erhard Scholz for his careful critique of the paper. While agreeing with its basic viewpoint, he made some critical comments on my treatment of Grassmann and the mythical Weylmann. With his kind permission I quote them:

> The (historical) *lineale Ausdehnungslehre* was so much oriented towards the investigations of linear geometric structures and their algebraic generalization that there was a deep conceptual gulf between Grassmann's approach and Riemann's differential geometry of manifolds, which could only be bridged after a tremendous amount of deep and hard work. I do not see in Grassmann's late attempt to understand the algebraic geometry of curves and surfaces in terms of his *Ausdehnungslehre* a step that might have led him even somewhat near to a generalization of parallel transport in the sense of differential geometry. In "real history" there was no natural candidate for "Weylmann."
>
> ..... So, in short, your Newstein paper is an interesting thought experiment discussing the question of what would have happened if history had gone other than it did. In doing so, and following your line of investigation, we might find more precise answers as to why there was, e.g., still a long way to go from Grassmann to a potential "Weylmann." This is contrary to your intentions, I fear, but I cannot help reading your paper that way.

Rather than going contrary to my intentions, his remarks raise a most important question that supplements my approach to alternate histories: Given that we can invent various alternatives to the actual course of events, can one attach a sort of intrinsic probability to these various alternatives? I mean probability in the sense of a qualitative ranking of the probability of the alternatives rather than attaching a numerical value to the probability of each. In a truly "postmature" case, the ranking of the actual course of events would be lower than that of at least one of the alternatives. For example, the probability of a direct mathematical route from Riemann's local metric to Levi-Civita's local metrical parallelism would rank higher than the probability of the actual route via physics through Einstein's development of general relativity. Dr. Scholz makes a strong case for ranking the probability of the actual course of events from Grassmann's affine spaces to Weyl's affine connection higher than the probability of the step from Grassmann to Weylmann in my myth. I shall not pursue this issue further here, but again thank Dr. Scholz for comments that raise it in the context of my paper.

APPENDIX: RIEMANNIAN PARALLELISM AND AFFINELY CONNECTED SPACES

I shall review the concepts of parallelism in Euclidean and affine spaces, and their generalization to non-flat Riemannian and affinely connected spaces, respectively. I shall emphasize material needed to understand the historical and mathematical discussion in the Sections 3–6 and Newstein's mythical history in Section 7. Those familiar with the mathematical concepts may refer to the Appendix as needed when reading Sections 4–7.[38]

a. affine and Euclidean spaces. The familiar concept of parallelism in Euclidean space can easily be extended from lines to vectors: two vectors at different points in that space are parallel if they are tangent to parallel lines. We say that two Euclidean vectors are equal if they are parallel and have the same length as defined by the metric of Euclidean space. But, as we shall soon see, the concepts of parallelism and equality of parallel vectors retain their significance when we abstract from the metric properties of Euclidean space to get an affine space.

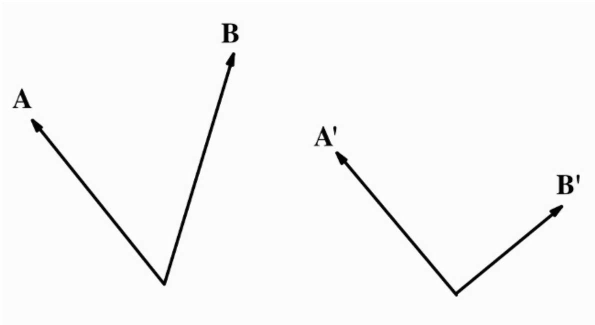


Figure 1: Any pair of non-parallel vectors $\mathbf{A}$ and $\mathbf{B}$ can be transformed into any other pair $\mathbf{A}'$ and $\mathbf{B}'$ by an (active) affine transformation.

The properties of Euclidean geometry may be defined as those that remain invariant under transformations of the Euclidean group, consisting of translations $T(3, R)$,[39] and of rotations $O(3, R)$ about any point in space.[40] A translation is a point transformation that takes the point P into the point $P + \boldsymbol{v}$, where $\boldsymbol{v}$ is any vector. A rotation is a point transformation with a fixed point P that takes the point

---

38 However, in contrast to more familiar treatments, I shall define connections in terms of frame bundles, a concept that I shall introduce informally, following (Crampin and Pirani 1986, chaps. 13–15), which may be consulted for more details.

39 I shall use the notation (n, R) to denote a group acting on a real n- dimensional space.

40 I shall give the active interpretation of all geometrical transformations: The transformations act on the points of the space in question, taking each point into another one. The idea of defining a geometry by the group of transformations that leave invariant all geometric relations goes back to (Klein 1872).

$P + \boldsymbol{r}$ into the point $P + \boldsymbol{O}r$ where $\boldsymbol{O} \in O(3, R)$ is an orthogonal transformation. The translations are clearly metric-independent; but the orthogonal transformations, being the linear transformations that preserve the distance between any pair of points, clearly do depend on the metric.

If we relax the condition that a linear transformation $\boldsymbol{L}$ preserve distances, and merely demand that it have a non-vanishing determinant), then $\boldsymbol{L} \in GL(3, R)$, the group of general linear or affine transformations. Together with the translations, they form the affine group that defines an affine geometry.[41] Parallelism of lines and vectors and the ratio of the lengths of parallel vectors (and hence the equality of two such vectors) being invariant under the affine group, are meaningful affine concepts. The (Euclidean) length of any vector is changed by an affine transformation with non-unit determinant, so it is not a meaningful affine concept.

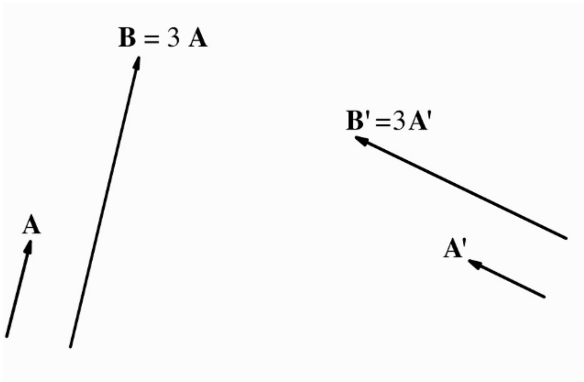


*Figure 2: Any pair of parallel vectors **A** and **B** can be transformed into any other pair of parallel vectors **A'** and **B'** with the same ratio by an (active) affine transformation.*

In order to determine the action of an affine transformation $\boldsymbol{L}$ on any vector $\boldsymbol{v}$ at some point of an n-dimensional affine space, we need merely define its action on a basis or linear frame $\boldsymbol{e}_i$ at that point, consisting of n linearly-independent vectors:

$$\boldsymbol{e}_j' = \boldsymbol{L}^i_j \boldsymbol{e}_i, \tag{11}$$

where $\boldsymbol{e}_j'$ is the new basis produced by the action of $\boldsymbol{L}$ on $\boldsymbol{e}_i$, and $\boldsymbol{L}^i_j$ is the matrix representing the action of $\boldsymbol{L}$ on some basis. (Here and throughout, we have adopted the summation convention for repeated indices, which range over the appropriate number of dimensions—here $1, ..., n$.)

If we want to restrict ourselves to Euclidean geometry and the orthogonal group, we may restrict ourselves to orthonormal bases or frames:

41 For a discussion of affine and metric spaces, with a view to the generalizations needed below, see (Crampin and Pirani 1986, chaps. 1 and 7). For these generalizations, see chaps. 9 and 11.

$$\boldsymbol{e}_i \cdot \boldsymbol{e}_j = \delta_{ij}, \tag{12}$$

where the dot symbolizes the Euclidean scalar product of two vectors, and to orthogonal changes of bases:

$$\boldsymbol{e}_j{}' = \mathcal{O}^i_j \boldsymbol{e}_i, \qquad \boldsymbol{e}_i{}' \cdot \boldsymbol{e}_j{}' = \delta_{ij}. \tag{13}$$

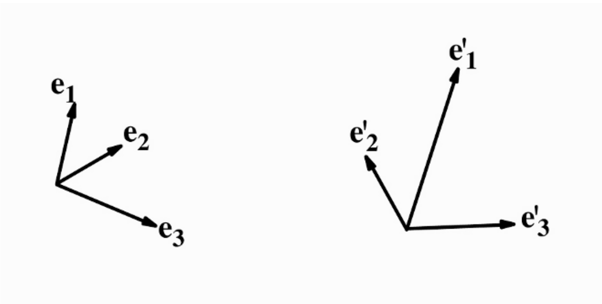


Figure 3: A (homogenous) affine transformation is defined by its action on a basis (or linear frame) $\boldsymbol{e}_A$ *of the affine space.*

Once we have chosen a basis at one point of an affine (or Euclidean) space, we can take as the basis at any other point of space the set of basis vectors equal and parallel to the original basis, thereby setting up a field of bases or *linear frames* over the entire space.

b. frame bundles. On the other hand, we can consider the set of all possible bases or linear frames at a given point of space. As is clear from eq. (11), in an affine space these frames are related to each other by the transformations of $GL(n, R)$. The set of all frames, together with the structure that the n- dimensional affine group imposes on them, is said to form a *fibre* over the point in question. Similarly, in Euclidean space, the set of all possible orthonormal frames at a point has a structure imposed on it by $O(n, R)$, the n- dimensional orthogonal group (see eq. (13)).

The set of all possible frames at every point of a space together with the space itself form a manifold that is called the *bundle of linear frames* or, more simply, the *frame bundle*. This is a special case of the more general concept of a *fibre bundle*.[42] The original space, which is affine or Euclidean in our examples but capable of generalization to any manifold, is called the *base space* of the fibre bundle; each *fibre* also need not be composed of linear frames, but may have a more general structure (below we shall consider fibres composed of tangent spaces). But there is always a *projection operation* that takes us from any fibre of the bundle to the point of the base

42 For fibre bundles in general and the frame bundle in particular, see (Crampin and Pirani 1986, chaps. 13 and 14).

space at which the fibre is located. A fibre bundle is called *trivial* if it is equivalent to the Cartesian product of a base manifold times a single fibre with a structure on it. The frame bundles we have been considering are trivial, since they are equivalent to the product of an affine (or Euclidean) space times a frame fibre with the structure imposed on it by the affine (or orthogonal) group.

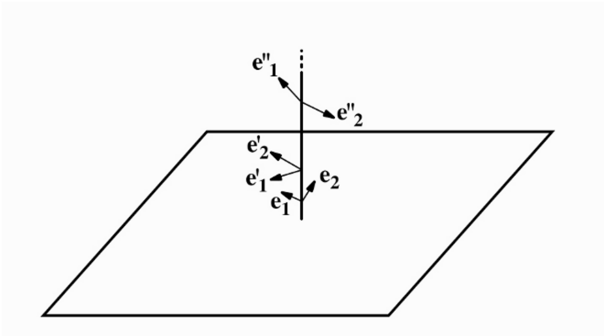


Figure 4: The set of all possible frames $\boldsymbol{e}_A$, $\boldsymbol{e}'_A$, $\boldsymbol{e}''_A$, ... *at a point of the space forms a "fibre" over the point.*

A *cross-section* of the frame bundle is a specification of a particular frame on each fibre of the bundle, i.e., at each point of the base space (see Fig. 6). (The frames must vary in a smooth way as we pass from point to point, but we shall not bother here with such mathematical details.) In an affine (or Euclidean) space, the specification of a linear (or orthonormal) frame on one fibre allows us to pick out a unique parallel cross section of the entire bundle. (The last sentence just repeats, in the language of fibre bundles, something said earlier.) A change of frame on one fibre produces a change of the entire parallel cross section that is induced by an affine (orthogonal) transformation on the original fibre.

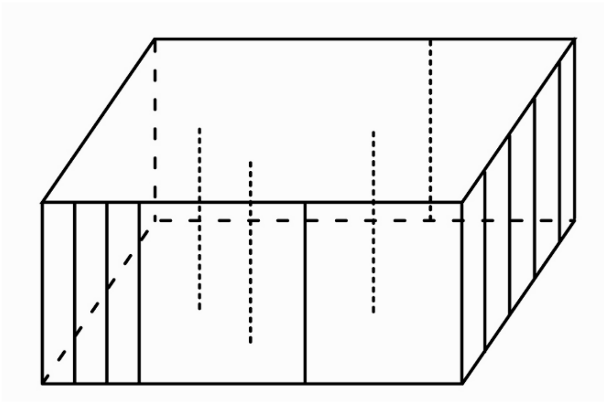


Figure 5: The fibres of a fibre bundle.

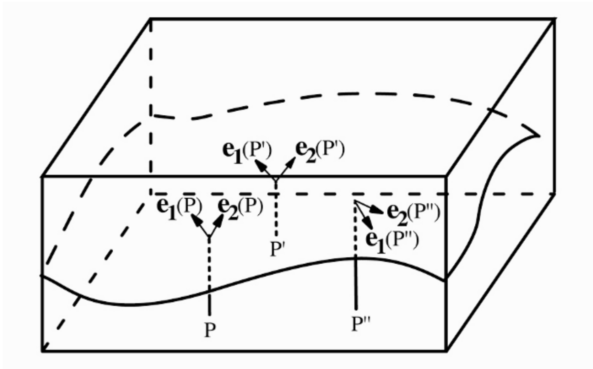


Figure 6: A "cross-section" of a frame bundle is a choice of a particular frame on each fibre of the bundle.

c. parallelism in non-flat Riemannian spaces. Now consider three-dimensional Euclidean space and some two-dimensional (generally curved) surface S in it. All vectors that are tangent to S at one of its points P form a vector space $T(P)$, called the tangent space to S at P. The collection of all such tangent spaces for all points $P \in S$ form a fibre bundle $T(S)$, called the *tangent bundle*. All vectors in $T(S)$ are intrinsically related to S,[43] and we want to define the concept of parallelism for such vectors in such a way that it will also be intrinsic to S. We cannot simply take the vector at another point P' of S that is parallel to a vector of $T(P)$ in the three-dimensional Euclidean sense: in general, that vector will not even be in $T(P')$, see Fig. 7).

We can get an idea of how to proceed by considering the case when S is a plane. The concept of parallel vectors at different points of the plane is clearly intrinsic to the plane. Consequently, the tangent spaces at each point of the plane can be identified with each other in a natural way, as can pairs of orthonormal vectors $\boldsymbol{e}_A (A = 1, 2)$ that form a basis at each point of the plane considered as a two-dimensional Euclidean space. Taken together with the unit normal vector $\boldsymbol{n}$ to the plane, the $\boldsymbol{e}_A$ form a basis for the tangent space of the three-dimensional Euclidean space.

43 These vectors can, for example, be defined as the tangent vectors to curves $C = P(s)$ lying entirely in S. We follow the usual terminology in distinguishing *curves* from *paths*, which are curves without a parametrization s.

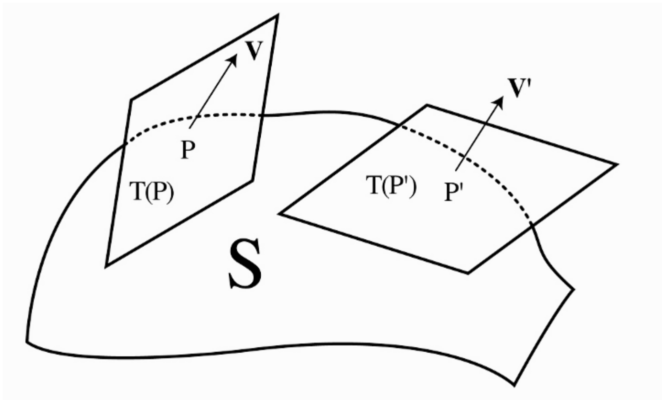


Figure 7: To define an intrinsic notion of parallelism within a surface S, we cannot use vectors that are parallel to each other in the three-dimensional sense. While V lies in the tangent plane at P, the three-dimensionally parallel vector V' does not even lie in the tangent plane at P'.

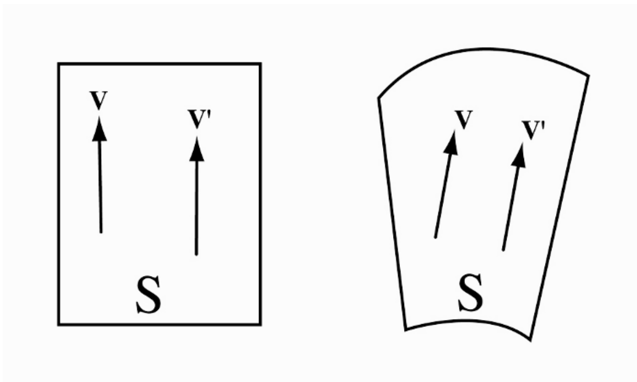


Figure 8: If V and V' are parallel vectors in the plane S, and if parallelism is intrinsic to S, then they remain parallel even when S is bent (without distortion).

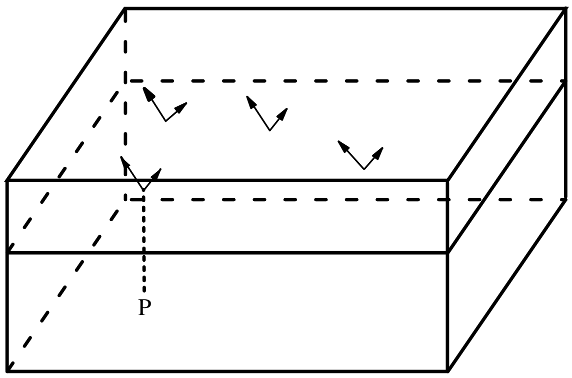


Figure 9: In an affine space, choice of a frame on one fibre picks out a unique parallel cross-section.

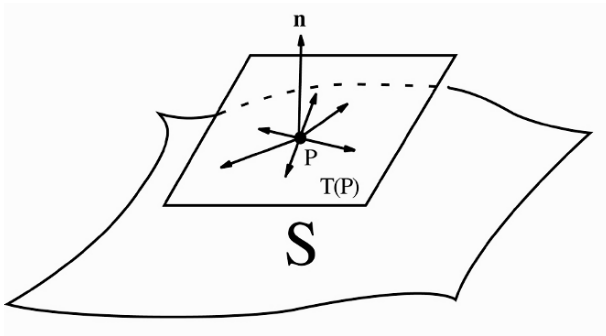


Figure 10: The tangent space $T(P)$ to a surface S at point P of the surface in Euclidean space is composed of all vectors tangent to the surface at that point. The unit normal to the tangent plane is designated by $\boldsymbol{n}$.

Now suppose we bend the plane without distorting its metric properties (i.e., the metrical relations between its points as measured *on the surface*), resulting in what is called a *developable surface*.[44] If we want the concepts of parallelism and straight line to be intrinsic to a such a surface, they must remain the same for any surface developed from the plane as they were for the plane itself. Thus, the basis vectors $\boldsymbol{e}_A$

44 Such a process of bending leaves the *intrinsic* geometry of the surface unchanged, but changes its *extrinsic* geometry. The intrinsic properties of any surface are those that remain unchanged by all such bendings; its extrinsic properties are precisely those that depend on how the surface is embedded in the enveloping Euclidean space.

at different points of the surface must still be considered parallel to each other from the intrinsic, surface viewpoint, even though they are not from the three-dimensional Euclidean point of view. Consider two neighboring points on the surface P and $P' = P + d\boldsymbol{r}$. In order to get from the tangent plane $T(P)$ at P to the tangent plane $T(P')$ at P' one must rotate the former through the angle $d\theta$ that takes $\boldsymbol{n}$ into $\boldsymbol{n}'$.[45] Thus, there must be an orthogonal transformation $\boldsymbol{\mathcal{O}}$, differing from the identity $\boldsymbol{I}$ only by an amount that depends on P, P', $\boldsymbol{n}$ and $\boldsymbol{n}'$, or equivalently on P, $\boldsymbol{n}$, $\boldsymbol{n}'$ and $d\boldsymbol{r}$:

$$\boldsymbol{\mathcal{O}} = \boldsymbol{I} + d\boldsymbol{\mathcal{O}}, \quad d\boldsymbol{\mathcal{O}} = d\boldsymbol{\mathcal{O}}(P, \boldsymbol{n}, \boldsymbol{n}', d\boldsymbol{r}) \tag{14}$$

and depends linearly on $d\boldsymbol{r}$.

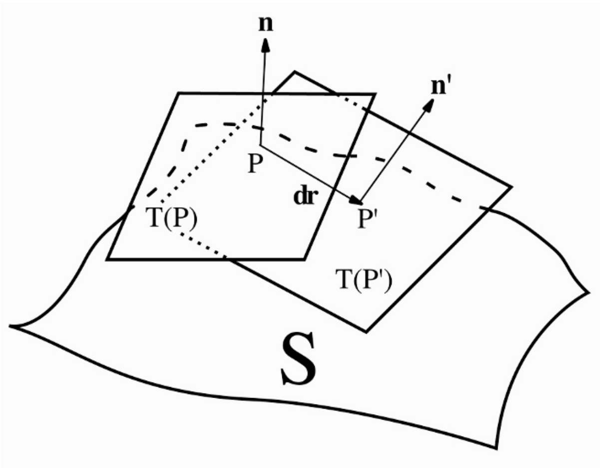


Figure 11: In order to get from the tangent plane $T(P)$ at P to a neighboring tangent plane $T(P')$ at $P' = P + d\boldsymbol{r}$, we must carry out an orthogonal transformation $\boldsymbol{\mathcal{O}} = \boldsymbol{I} + d\boldsymbol{\mathcal{O}}$ that depends on P, $d\boldsymbol{r}$, $\boldsymbol{n}$ and $\boldsymbol{n}'$.

Due to the linearity of vector spaces, the effect of this orthogonal transformation on any vector in the tangent plane to the surface can be computed once its effect on a set of basis vectors $\boldsymbol{e}_A$ in the tangent plane is known.[46] The change in each basis vector is given by:

$$\delta\boldsymbol{e}_B(P') = (d\boldsymbol{O})^A_{\ B}\,\boldsymbol{e}_A, \tag{15}$$

45 The concept of parallelism in the Euclidean space allows us to draw the vector at P that is equal and parallel to $\boldsymbol{n}'$ at P', and so define the angle $d\theta$ between $\boldsymbol{n}$ and $\boldsymbol{n}'$.

46 Note that we need the normals $\boldsymbol{n}$ and $\boldsymbol{n}'$ to define the orthogonal transformation between parallel vectors lying in the tangent planes at P and P'; but since we are only interested in the change in vectors lying in the surface we may omit $\boldsymbol{n}$ from explicit mention in eq. (5), since it is determined by the $\boldsymbol{e}_A$ and the orthonormality conditions.

where $(d\mathbf{0})^A_B\ (P, d\boldsymbol{r})$ are the elements of a matrix that determines the effect of the infinitesimal rotation on the orthonormal basis vectors.

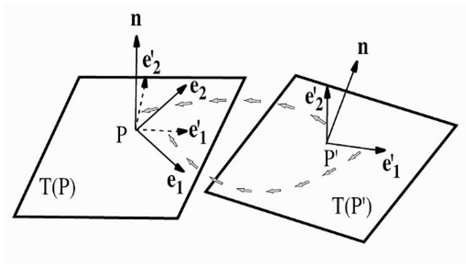


Figure 12: The effect of $d\boldsymbol{O}$ on any vector in $T(P)$ is determined by its effect on a set of basis vectors $\boldsymbol{e}_A$ of the space.

It is this connection between parallel vectors in neighboring tangent planes, given by eqs. (14) and (15), that we shall preserve for all surfaces, in particular for those that are not intrinsically plane. Since it was introduced by Levi-Civita (see Section 4), it is often called the Levi-Civita connection. If two points P and Q are not neighboring, we must choose some path C on the surface connecting P and Q, and break it up into small straight line segments PP', $P'P''$, ..., Q. If we move from P to P' along straight-line segment PP', we must rotate $T(P)$ at P through some small angle $\Delta\theta$ about the normal $\boldsymbol{n}$ at P in order to get the tangent plane $T(P')$ at P'. For the next segment $P'P''$, we have to rotate the tangent plane $T(P')$ through an angle $\Delta\theta'$ about the normal $\boldsymbol{n}'$ at P' in order to get the tangent plane $T(P'')$ at P''. We keep doing this until we reach the endpoint Q. Now we increase the number of intermediate points indefinitely, and take the limit of this process so that the broken straight line segments approach the curve. This defines the vector in $T(Q)$ that is parallel to one in $T(P)$ *with respect to the path* C.

Note that we must add the last qualification because, unless S is a developable surface, the resulting parallelism in general will be path dependent. We can see this by looking at a small parallelogram with sides PP', $P'Q$ and PP'', $P''Q$. Since $T(P')$ and $T(P'')$ are not in general parallel to each other, the correspondence between vectors in the tangent planes $T(P)$ and $T(Q)$ that is set up by going via $T(P')$ is not in general the same as the one we get by going via $T(P'')$.

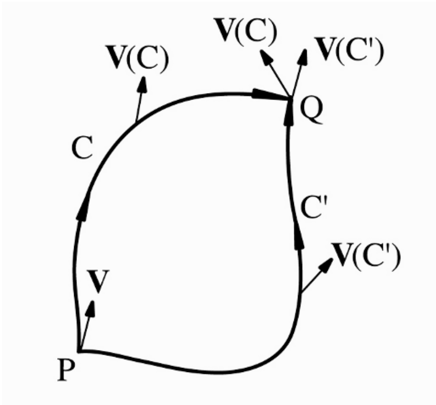


Figure 13: In general the vector $V'(C)$ at Q that is parallel to V at P depends on the path taken between P and Q.

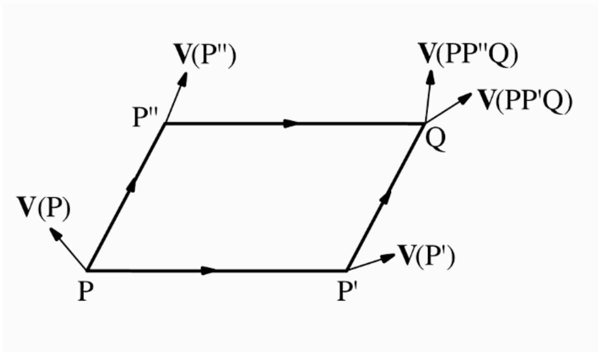


Figure 14: We can see this by looking at the parallel transport of a vector $\boldsymbol{v}$ along the sides PP', $P'Q$ and PP'', $P''Q$ of a small parallelogram.

d. the Riemann tensor. By carrying out the analysis of this parallelogram quantitatively, we can define the Riemann tensor of the surface.[47] Take a vector $\boldsymbol{v}$ in $T(P)$, and let the corresponding (i.e. intrinsically parallel) vector in $T(P')$ be $\boldsymbol{v} + \delta\boldsymbol{v}$. Then $\boldsymbol{v} + \delta\boldsymbol{v}$ results from $\boldsymbol{v}$ by a rotation operation that acts on $\boldsymbol{v}$; we shall symbolize it by the operator $\boldsymbol{I} + d\boldsymbol{O}$ (see eq. (14)), so that:

47 In the case of a two-dimensional surface, it reduces to a scalar R; i.e., all non-vanishing components of the Riemann tensor reduce to $\pm R$. But we prefer to keep the tensorial designation in view of the impending generalization to higher dimensions.

$$\delta \boldsymbol{v} = d\boldsymbol{\mathcal{O}}\boldsymbol{v}.$$

Here, $d\boldsymbol{\mathcal{O}}$ represents a first order infinitesimal rotation operator that depends linearly on $d\boldsymbol{r}$. Similarly, if $d\boldsymbol{r}'$ represents the displacement PP'', then the change $\delta\boldsymbol{v}'$ in $\boldsymbol{v}$ when we go from $T(P)$ to $T(P'')$ is given by:

$$\delta \boldsymbol{v}' = d\boldsymbol{\mathcal{O}}'\boldsymbol{v}.$$

Then the change in $\boldsymbol{v}$ at $T(Q)$ when we go via $d\boldsymbol{r}$ first, then $d\boldsymbol{r}'$ (i.e., via $PP'Q$) is given by:

$$\delta \boldsymbol{v}_1 = (\boldsymbol{\mathcal{I}} + d\boldsymbol{\mathcal{O}}')(\boldsymbol{\mathcal{I}} + d\boldsymbol{\mathcal{O}})\boldsymbol{v} - \boldsymbol{v} = (d\boldsymbol{\mathcal{O}}' + d\boldsymbol{\mathcal{O}} + d\boldsymbol{\mathcal{O}}'d\boldsymbol{\mathcal{O}}v)\boldsymbol{v};$$

while, if we proceed in the reverse order (i.e., via $PP''Q$), the change is given by:

$$\delta \boldsymbol{v}_2 = (d\boldsymbol{\mathcal{O}} + d\boldsymbol{\mathcal{O}}' + d\boldsymbol{\mathcal{O}}d\boldsymbol{\mathcal{O}}')\boldsymbol{v}.$$

Since $\delta\boldsymbol{v}_1$ and $\delta\boldsymbol{v}_2$ are vectors at the same point, their difference is a (second order infinitesimal) vector $\delta^2\boldsymbol{v}$. It indicates by how much the two vectors in $T(Q)$ that are parallel to $\boldsymbol{v}$ in $T(P)$, depending on which of the two paths is taken, differ from each other:

$$\delta^2 \boldsymbol{v} = (d\boldsymbol{\mathcal{O}}d\boldsymbol{\mathcal{O}}' - d\boldsymbol{\mathcal{O}}'d\boldsymbol{\mathcal{O}})\boldsymbol{v}.$$

Note the operator in parentheses is the same for all vectors in $T(P)$ since they are all rotated by the same amount. And since $d\boldsymbol{\mathcal{O}}$ and $d\boldsymbol{\mathcal{O}}'$ are linear in $d\boldsymbol{r}$, $d\boldsymbol{r}'$, respectively, this operator is proportional to $(d\boldsymbol{r}d\boldsymbol{r}' - d\boldsymbol{r}'d\boldsymbol{r})$. Such an antisymmetric tensorial product of two vectors is abbreviated as $d\boldsymbol{r} \wedge d\boldsymbol{r}'$ and called a simple bivector; it represents the (signed) area of the infinitesimal parallelogram with sides $d\boldsymbol{r}, d\boldsymbol{r}'$. This second order infinitesimal term is also same for all vectors taken from P to Q along the sides of the parallelogram.[48] So there must be a finite tensorial operator $\boldsymbol{R}$, such that, when it operates on an area bivector $d\boldsymbol{r} \wedge d\boldsymbol{r}'$ and a vector $\boldsymbol{v}$, it produces the change in $\boldsymbol{v}$ when it is parallel transported around the area $d\boldsymbol{r} \wedge d\boldsymbol{r}'$. Note that, to the second differential order we are considering, it makes no difference whether we parallel transport a vector from P to Q in two different ways, and compare the results in $T(Q)$, or take it around the parallelogram and compare the result with the original vector in $T(P)$. Further, the result is independent of the shape of the infinitesimal plane figure we carry it around so long as this has the same area as, and lies in the plane defined by, $d\boldsymbol{r} \wedge d\boldsymbol{r}'$. The tensorial operator $\boldsymbol{R}$, which operates on a bivector and a vector to produce another vector, is called the *Riemann tensor*; when it operates on an infinitesimal area element, it measures how much Riemannian parallelism

48 One should actually distinguish between $d\boldsymbol{r}$ at P and $d\boldsymbol{r}$ at P'', which is the result of parallel transporting $d\boldsymbol{r}$ at P along $d\boldsymbol{r}'$. But to the order we are considering, the difference may be neglected. The more serious problem of whether the parallelogram resulting from these displacements actually "closes" will be discussed later.

on that surface element differs from flat, path-independent, parallelism, for which the Riemann tensor would vanish.

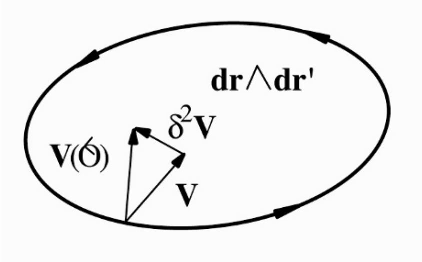


Figure 15: The operator $\boldsymbol{R}$, *operating on the area* $d\boldsymbol{r} \wedge d\boldsymbol{r}'$ *and the vector* $\boldsymbol{v}$, *produces the change* $\delta^2\boldsymbol{v}$ *in* $\boldsymbol{v}$ *when it is parallel transported around that area.*

e. non-flat affine spaces. Our discussions of parallelism on a surface and of the Riemann tensor made essential use of the metric of the enveloping Euclidean space. First of all, this metric induced a notion of distance on the surface; but this is intrinsic to the surface, and can be defined without using the fact that the surface is embedded in a Euclidean space. More serious is the fact that we used the normals to the surface at each point in order to develop the relation between tangent spaces at neighboring points in terms of an orthogonal transformation (rotation through some angle). The notions of orthogonality and angle are intrinsically metrical.

Suppose we abstract from these metric concepts and consider an affine space, as discussed above. Using only affine concepts, can we still define concepts of parallelism and straight line on a surface in an affine space? The answer is yes, but we must introduce a substitute for the unit normal field given naturally in a Euclidean space.

First of all, the concept of surface is independent of a metric, as are those of the tangent space at each point of a surface, and (hence) of the tangent bundle. But now we have no natural way of relating the tangent spaces at different point of the surface by means of a general linear transformation. At each point of the surface, a basis in its tangent space must be supplemented by a vector that does not lie in the tangent space; i.e., a vector that takes the place of the normal vector to the surface in a Euclidean space. Together with the chosen basis in the tangent space, this vector constitutes a basis for the enveloping affine space. This vector field is said to *rig* the surface, and the process is called *rigging*. Once the surface is rigged, one can carry out in an affine space a procedure to relate neighboring tangent spaces that is entirely analogous to the procedure used in the Euclidean case. The only difference is that, instead of the infinitesimal orthogonal transformation $d\mathbf{O}$ that carries the orthonormal basis at P into the orthonormal basis at P', one considers the infinitesimal general linear transformation $d\mathbf{L}$ that takes a basis for the enveloping affine space at P into the corresponding basis at P'. Due to the linearity of vector spaces, carrying out the transformation $d\mathbf{L}$ on any vector in $T(P)$ yields the corresponding parallel vector in $T(P')$. Such a connection between tangent spaces, which generalizes to surfaces in an affine space the Levi-Civita connection for surfaces in a Euclidean space, is called *a general linear connection*. Once the connection is defined, everything proceeds in a way that is entirely analogous to that for Euclidean spaces (see the previous subsection), up to and including the definition of the Riemann tensor operator.

Instead of eq. (15), giving the effect of an infinitesimal orthogonal transformation (rotation) matrix on an orthonormal basis, we can now specify the effect on the vectors in a tangent plane of an infinitesimal general linear transformation, by specifying the infinitesimal general linear transformation matrix $(dL)^A_B$ that gives the effect of this transformation on an arbitrary basis:

$$\delta \boldsymbol{e}_B(P') = (dL)^A_B \boldsymbol{e}_A. \tag{16}$$

f. covariant differentiation, geodesics. Once we have the concept of parallelism along a path, we can define a derivative operation for a vector field on a surface. The essence of the usual derivative operation for a vector field in Euclidean space consists in comparing the value of the vector field $\boldsymbol{v}$ at some point with its values at some neighboring points. But we can only compare vectors in the same tangent space: what we actually do to compare vectors at two points P and Q is to compare $v(Q)$ with the vector at Q that is parallel to $\boldsymbol{v}(P)$. We shall proceed in the same way on a surface and compare values at two neighboring points P and $P + d\boldsymbol{r}$:

$$\boldsymbol{v}(P + d\boldsymbol{r}) - [\boldsymbol{v}(P) + \delta\boldsymbol{v}] = d\boldsymbol{r} \cdot \partial\boldsymbol{v} - \delta\boldsymbol{v},$$

since

$$\boldsymbol{v}(P + d\boldsymbol{r}) = \boldsymbol{v}(P) + d\boldsymbol{r} \cdot \partial\boldsymbol{v},$$

where ∂ represents the ordinary-derivative gradient operation; operating on a scalar field $\phi(\boldsymbol{r})$, it gives the gradient vector field $\partial\phi$; but operating on a vector (or tensor) field it does not produce another vector (or tensor) field. It must be supplemented by the second term $\delta\boldsymbol{v}$ for a vector (and similar terms for higher-order tensors). Since $\delta\boldsymbol{v} = d\boldsymbol{\mathcal{O}}\boldsymbol{v}$, we can write the invariant combination as

$$\boldsymbol{v}(P + d\boldsymbol{r}) - [\boldsymbol{v}(P) + \delta\boldsymbol{v}] = d\boldsymbol{r} \cdot \partial\boldsymbol{v} - d\boldsymbol{\mathcal{O}}\boldsymbol{v}.$$

Since $d\boldsymbol{\mathcal{O}}\boldsymbol{v}$ is also linear in $d\boldsymbol{r}$, we can abbreviate the right-hand side as:

$$d\boldsymbol{r} \cdot \partial\boldsymbol{v} - d\boldsymbol{\mathcal{O}}\boldsymbol{v} = d\boldsymbol{r} \cdot \nabla\boldsymbol{v}.$$

The expression $d\boldsymbol{r} \cdot \nabla$ represents an invariant directional derivative in the $d\boldsymbol{r}$ direction. Since the result is linear in $d\boldsymbol{r}$, there must be a tensorial operator $\boldsymbol{\nabla}$ called the *covariant derivative* operator, that operates on a vector to produce a mixed tensor $\boldsymbol{\nabla v}$ with one covariant and one contravariant (i.e.,vectorial) place.

On a surface, we may generalize the concept of a straight line in an affine space to that of a geodesic by requiring that the parallel transport of its tangent vector along a geodesic remain the tangent vector. If $\boldsymbol{t}$ represents the tangent vector to the curve $C(\lambda)$, $\boldsymbol{t} = dC/d\lambda$, this means that a geodesic must satisfy the equation:

$$\boldsymbol{t} \cdot \nabla \boldsymbol{t} = 0.$$

g. generalizations, intrinsic characterizations. Nothing in the discussion above depends essentially on the number of dimensions being three, and it can be immediately generalized to n-dimensional metric and affine spaces, defined by the translation groups $T(n)$ and $O(n, R)$ and $T(n)$ and $GL(n, R)$ respectively; and to their m-dimensional sub-spaces. If m is less than $n-1$, then there are n-m normals, and n-m rigging vectors must be defined; but otherwise the discussion proceeds quite analogously. Since any m-dimensional Riemannian or affinely-connected space can be embedded in an n-dimensional Euclidean or affine space of sufficiently high dimension (locally, if not globally), such embedding arguments can handle the generic case.

Of course, once the basic geometrical concepts have been grasped, an intrinsic method of characterizing curved spaces, independently of any embedding in flat spaces of higher dimension, is preferable. It is clear from the previous discussion how to proceed. One must specify a connection between vectors in $T(P)$ and $T(P')$ that defines when a vector in one is parallel to a vector in the other. In contrast to the order in the previous embedding considerations, I shall first give the definition for a general affine linear connection, and then indicate how to specialize it to a Riemannian or Levi-Civita connection.

As indicated earlier (see discussion around eqs. (15) and (16) above), in order to connect arbitrary vectors in the two tangent spaces, it suffices to indicate how sets of basis vectors in the two tangent spaces are connected. Let $\boldsymbol{e}_i(P)$ be a set of basis vectors in $T(P)(i = 1, 2, \ldots, n)$. The changes in these basis vectors when we move to $T(P + d\boldsymbol{r})$ will be given by (generalizing eq. (6) above):

$$\delta \boldsymbol{e}_i(P') = (dL)_i^j\, \boldsymbol{e}_j, \qquad (i, j = 1, 2, \ldots, n).$$

Our connection is linear in $d\boldsymbol{r}$, so it suffices to know the change in $\boldsymbol{e}_i$ for a small change in each of the basis directions, $d\boldsymbol{r} = \in e_k$, where $\in$ is an infinitesimal of first order.

$$\delta \boldsymbol{e}_i(P') = \in L_i^j(P, \boldsymbol{e}_k)\boldsymbol{e}_j.$$

On the other hand $\delta \boldsymbol{e}_i$ itself must be a linear combination of the basis vectors, so we may decompose it into the infinitesimal changes in each of these directions:

$$(\delta \boldsymbol{e}_i)_j = \in \Gamma_{ij}^k \boldsymbol{e}_k.$$

Thus, specification of the set of quantities $\Gamma_{ij}^k(P)$ at all points of the manifold fixes the affine connection intrinsically.[49] We call the Γ_{ij}^k the *components of the connection* with respect to the basis $\boldsymbol{e}_i$.[50]

If we now want to construct the parallelogram as described above in the definition of the Riemann tensor, we must make sure that it “closes,” that is, that we reach the same point if we parallel transport $d\boldsymbol{r}$ along $d\boldsymbol{r}'$ as we do if we parallel transport $d\boldsymbol{r}'$

along $d\boldsymbol{r}$. It is relatively simple to show that this will be the case if Γ^k_{ij} is symmetric in its two lower indices; we shall consider only such symmetric affine connections.[51]

The Riemann tensor operator $\boldsymbol{R}$ can now be defined in terms of its effect on the basis vectors. If we transport $\boldsymbol{e}_k$ around an area defined by $\boldsymbol{e}_i \wedge \boldsymbol{e}_j$, then its change in the $\boldsymbol{e}_l$ direction is given by R^k_{ijl}. These are the components of the Riemann tensor with respect to the basis $\boldsymbol{e}_i$, which can easily be related to the derivatives of the Γ^j_{ki}, but we omit the details. For future reference, we note that R^k_{ijl} is antisymmetric in its last pair of indices, and that if we contract its upper index with either of the last two indices, say the second, we get (plus or minus) the Ricci tensor R_{il}.

The covariant derivative operator will have components:

$$\nabla_i = \boldsymbol{e}_i \cdot \nabla;$$

the components of the covariant derivative of a vector $\boldsymbol{\nabla v}$, for example, are:

$$\nabla_i v^j = \partial_i v^j + \Gamma^j_{ki} v^k.$$

The components of the geodesic equation in an adapted coordinate system are:

$$d^2 x^j / d\lambda^2 + \Gamma^j_{ki}(dx^k / d\lambda)(dx^i / d\lambda) = 0.$$

The components of the Riemann tensor with respect to a basis can be similarly calculated.

Turning to Riemannian spaces, it is natural to demand that parallel transport along any path preserve the length of all vectors. If we impose this condition on a symmetric affine connection, we are led uniquely to the Levi-Civita connection discussed above; but again we omit the details.

For future reference, we also note that, just as in the case of a surface in a linear (flat) affine space discussed above, a connection is induced on a hypersurface in a non-flat affinely connected space if that hypersurface is rigged with an arbitrary vector field.[52]

---

49 Note that these quantities transform as scalars under a coordinate transformation, but as tensors under a change of basis. If we use the natural basis associated with a coordinate system (see the following note) and carry out a simultaneous coordinate transformation and change of natural basis, they transform under a more complicated, non-tensorial transformation law (see Section 6, eq. (7)).

50 Note that a basis need not be holonomic, i.e., coordinate forming. It will be if and only if the Lie bracket of any pair of basis vectors vanishes. We shall only need holonomic bases, for which an associated coordinate system x^i exists, such that in this coordinate system e^j_i, the coordinate components of $\boldsymbol{e}_i$, are equal to δ^j_i, the Kronecker delta. Conversely, a basis is associated with any coordinate system by the same relations.

51 If the parallelogram does not close, the antisymmetric part of Γ^j_{ki} defines the so-called *torsion tensor*.

52 It is customary, when discussing spaces of more than three dimensions, to refer to subspaces of one less dimension than that of the space as *hypersurfaces*. Thus, when the discussion is generalized to more than three dimensions, "surfaces" become "hypersurfaces."

h. frame bundles and connections. We introduced the concept of affine connection in the currently-habitual way, in terms of its local action on vector or frame fields in some manifold. But a connection is more naturally introduced globally in terms of the frame bundle over that base manifold (see Section b). A curve C in the base manifold together with a frame field defined along the curve corresponds to a curve $\mathcal{C}$ in the frame bundle; and conversely $\mathcal{C}$ projects down to C in the base manifold, together with a frame field along the curve. Now a connection provides a rule for defining such curves in the frame bundle: given a curve C in the base manifold together with an initial frame at some point on the curve, parallel transport of the initial frame along the curve thus defines a unique curve $\mathcal{C}$ in the frame bundle. The only thing we have to worry about is what happens if we change the initial frame by the action of some element $\mathcal{L}^i_j$ of the general linear group (see eq. (11)).The curve in the frame bundle is then transformed into another curve that differs from the first only by the same action of $\mathcal{L}^i_j$ on the frame at each point of the curve in the base manifold.

We can use this idea to define a connection globally as a collection of curves in the frame bundle, each passing only once through any fibre of the bundle, that satisfy the following condition: if two such curves $\mathcal{C}$ and $\mathcal{C}'$ project into the same curve C in the base manifold, and hence have all of their fibres in common, then on each fibre the frames on the two curves are related by the global application of the same $\mathcal{L}^i_j$.

If we want to restrict the structure group of the frame fibres to some subgroup of $GL(n, r)$, then we must assure that the connection introduced is compatible with the structure of this subgroup. For example, if we required compatibility with any of the orthogonal or pseudo-orthogonal subgroups, the Levi-Civita connection would result.[53]

REFERENCES

Bhaskar, Roy. 1993. *Dialectic: The Pulse of Freedom*. London/New York: Verso.

Bonola, Roberto. 1955. *Non-Euclidean Geometry: A Critical and Historical Study of its Development*. Dover reprint: New York: Dover Publications. (Open Court 1912)

Cartan, Élie. 1923. "Sur les variétés à connection affine et la théorie de la relativité généralisée." *Ecole Normale Supérieure* (Paris). *Annales* 40: 325–412. English translation in (Cartan 1986).

———. 1986. *On Manifolds with an Affine Connection and the Theory of Relativity.* Translated by Anne Magnon and Abhay Ashtekar. Naples: Bibliopolis. (Chapter 1 printed in this volume.)

Collier, Andrew. 1994. *Critical Realism: An Introduction to Roy Bhaskar's Philosophy.* London/New York: Verso.

Coolidge, Julian Lowell. 1940. *A History of Geometrical Methods*. Oxford: Clarendon Press.

Crowe, Michael J. 1994. *A History of Vector Analysis: The Evolution of the Idea of a Vectorial System*. New York: Dover Publications (reprint ed.).

Crampin, Michael, and Felix A.E. Pirani. 1986. *Applicable Differential Geometry.* Cambridge/London/New York: Cambridge University Press.

Darling, R. W. 1994. *Differential Forms and Connections*. Cambridge/New York/Melbourne: Cambridge University Press.

Ehlers, Jürgen. 1981. "Über den Newtonschen Grenzwert der Einsteinschen Gravitationstheorie." In J. Nitsch et al. (eds.), *Grundlagenprobleme der modernen Physik*. Mannheim: Bibliographisches Institut, 65–84.

53 See (Crampin and Pirani 1986, chap. 15) for details.

Einstein, Albert. 1907. "Über das Relativitätsprinzip und die aus demselben gezogenen Folgerungen." *Jahrbuch der Radioaktivität und Elektronik*. 4: 411–462.

———. 1914. "Die formale Grundlage der allgemeinen Relativitätstheorie." *Königlich Preussische Akademie der Wissenschaften* (Berlin). *Mathematisch-physikalische Klasse*. *Sitzungsberichte*: 1030–1085.

———. 1915. "Zur allgemeinen Relativitätstheorie." *Königlich Preussische Akademie der Wissenschaften* (Berlin). *Mathematisch-physikalische Klasse*. *Sitzungsberichte*: 778–786.

Einstein, Albert and Marcel Grossmann. 1913. *Entwurf einer verallgemeinerten Relativitätstheorie und einer Theorie der Gravitation. I. Physikalischer Teil von Albert Einstein. II. Mathematischer Teil von Marcel Grossmann*. Leipzig: Teubner.

Eisenstaedt, Jean and A. J. Kox (eds.). 1992. *Studies in the History of General Relativity* (*Einstein Studies*, vol 3). Boston/Basel/Berlin: Birkhäuser.

Friedrichs, Kurt O. 1927. "Eine invariante Formulierung des Newtonschen Gravitationsgesetzes und des Grenzüberganges vom Einsteinschen zum Newtonschen Gesetz." *Mathematische Annalen* 98: 566–575.

Gauss, Carl Friedrich. 1902. *General Investigations of Curved Surfaces*. Princeton: Princeton University Press. English translation by A. Hiltebeitel and J. Morehead of *Disquisitiones generales circa superficies curvas* (Göttingen: Dietrich, 1828).

Grassmann, Hermann. 1844. *Die lineale Ausdehnungslehre, ein neuer Zweig der Mathematik*. Leipzig: Otto Wigand.

———. 1862. *Die Ausdehnungslehre. Vollständig und in strenger Form bearbeitet*. Berlin: Enslin.

———. 1877. "Ueber die Beziehung der nicht-Euklidischen Geometrie zur Ausdehnungslehre."Appendix I to (Grassmann 1878). English translation in (Grassmann 1995, 279–280).

———. 1878. *Die Ausdehnungslehre von 1844 oder Die lineale Ausdehnungslehre*. (2nd. Edition.) Leipzig: Otto Wigand.

———. 1995. *A New Branch of Mathematics: The Ausdehnungslehre of 1844 and Other Works*. Translated by Lloyd C. Kannenberg. Chicago and LaSalle: Open Court. (Appendix printed in this volume.)

Hertz, Heinrich. 1894. *Die Prinzipien der Mechanik. In neuem Zusammenhange dargestellt*, Philipp Lenard (ed.). *Gesammelte Werke*, Vol. 3. Leipzig: Barth.

Hessenberg, Gerhard. 1917. "Vektorielle Begründung der Differentialgeometrie." *Mathematische Annalen* 78: 187–217.

Howard, Don and John Stachel (eds.). 1989. *Einstein and the History of General Relativity* (*Einstein Studies*, Vol. 1). Boston/Basel/Berlin: Birkhäuser.

Janssen, Michel. 1992. "H. A. Lorentz's Attempt to Give a Coordinate-Free Formulation of the General Theory of Relativity." In (Eisenstaedt and Kox 1992, 344–363).

Jost, Jürgen. 1991. *Riemannian Geometry and Geometric Analysis*, 2nd. ed. Berlin/Heidelberg/New York: Springer.

Klein, Felix. 1872. *Vergleichende Betrachtungen über neuere geometrische Forschungen*. Erlangen: A. Dühechert. Revised version in *Mathematische Annalen* 43: 63–100 (1893).

Kolmogorov, Andrei N. and Adolf P. Yushkevich. 1996. *Mathematics of the 19th Century*. Basel/Boston/Berlin: Birkhäuser.

Laptev, B. L. and B. A. Rozenfel'd. 1996. "Chapter 1. Geometry." In (Kolmogorov and Yushkevich 1996, 1–118).

Lawvere, F. William. 1996. "Grassmann's Dialectics and Category Theory." In (Schubring 1996, 255–264).

Levi-Civita, Tullio. 1916. "Nozione de Parallelismo in una Varieta Qualunque e Conseguente Specificazione Geometrica della Curvatura Riemanniana." *Rendiconti del Circolo Matematico di Palermo* 42: 17–205. (English translation of excerpts given in this volume.)

Lützen, Jesper. 1995a. "Interactions Between Mechanics and Differential Geometry in the 19th Century." *Archive for History of Exact Sciences* 49: 1–72.

———. 1995b. "Renouncing Forces: Geometrizing Mechanics. Hertz's Principles of Mechanics." *Københavns Universitet Matematisk Institut, Preprint Series* 1995, No. 22.

Norton, John. 1984. "How Einstein Found his Field Equations: 1912–1915." *Historical Studies in the Physical Sciences* 14: 253–316.

Reich, Karen. 1992. "Levi -Civitasche Parallelverschiebung, affiner Zusammenhang, Übertragungsprinzip: 1916/17–1922/23." *Archive for History of Exact Sciences* 44: 78–105.

———. 1994. *Die Entwicklung des Tensorkalküls. Vom absoluten Differentialkalkuel zur Relativitätstheorie*. Basel: Birkhäuser.

Riemann, Bernhard. 1868. "Über die Hypothesen, welche die Geometrie zu Grunde liegen." In Richard Dedekind and H. Weber (eds.), *Gesammelte mathematische Werke*, 2nd ed. 1892, (reprinted in New York: Dover Publications, 1953), 272–287.

Scholz, Erhard. 1995. "Hermann Weyl's Purely Infinitesimal Geometry." In *Proceedings of the International Congress of Mathematicians, Zurich, Switzerland 1994*. Basel: Birkhäuser.

Schouten, Jan A. 1918. "Die direkte Analysis zur neueren Relativitätstheorie." *Verhandelingen der Koninklijke Akademie van Wetenschappen te Amsterdam*, XII, no. 6.

Schubring, Gert (ed.). 1996. *Hermann Guenther Grassmann (1809–1877): Visionary Mathematician, Scientist and Neohumanist Scholar*. Dordrecht/Boston/London: Kluwer Academic.

Stachel, John. 1986. "What a Physicist Can Learn from the Discovery of General Relativity." In Remo Ruffini (ed.), *Proceedings of the Fourth Marcel Grossmann Meeting on General Relativity.* Amsterdam: Elsevier, 1857–1862.

———. 1987. "How Einstein Discovered General Relativity: A Historical Tale With Some Contemporary Morals." In Malcolm A. H. MacCallum (ed.), *General Relativity and Gravitation: Proceedings of the 11th International Conference on General Relativity and Gravitation*. Cambridge: Cambridge University Press, 200–208.

———. 1989. "Einstein's Search for General Covariance, 1912–1915." In (Howard and Stachel 1989, 63–100).

———. 1993. "The Meaning of General Covariance: The Hole Story." In John Earman et al. (eds.), *Philosophical Problems of the Internal and External World, Essays on the Philosophy of Adolf Grünbaum*. (Konstanz: Universitätsverlag, Pittsburgh University Press, 129–160.

———. 1994a. "Scientific Discoveries as Historical Artifacts." In Kostas Gavroglu (ed.), *Current Trends in the Historiography of Science*. Dordrecht: Reidel, 139–148.

———. 1994b. "Changes in the Concepts of Space and Time Brought About by Relativity." In Carol C. Gould and Robert S. Cohen (eds.), *Artefacts, Representations and Social Practice*. Dordrecht: Kluwer, 141–162.

———. 1995. "History of Relativity." In Laurie M. Brown, Abraham Pais and Brian Pippard (eds.), *Twentieth Century Physics*, Vol. I. Bristol and Phila.: Institute of Physics Pub., New York: American Institute of Physics Press.

——— (ed.). 1998. *Einstein's Miraculous Year: Five Papers That Changed the Face of Physics*. Princeton: Princeton University Press.

———. 2003a. "Critical Realism: Bhaskar and Wartofsky." In Carol C. Gould (ed.), *Constructivism and Practice: Towards a Historical Epistemology.* Lanham, Md.: Rowman and Littlefield, 137–150.

———. 2003b. "Einstein's Intuition and the Post-Newtonian Approximation." Talk at the Mexico City Conference in Honor of Jerzy Plebanski, May 26, 2003. To appear in the Proceedings of the Conference.

———. 2005. "Fibered Manifolds, Geometric Objects, Structured Sets, G-Spaces and All That: The Hole Story from Space-Time to Elementary Particles." To appear in (Stachel forthcoming).

———. Forthcoming. *Going Critical: Selected Essays*. Dordrecht: Kluwer.

Struik, Dirk. 1933. "Outline of the History of Differential Geometry." *Isis* 19: 92–120, *Isis* 20: 161–191.

Weyl, Hermann. 1918a. "Reine Infinitesimalgeometrie." *Mathematische Zeitschrift* 2: 384–411. (English translation of excerpt given in this volume.)

———. 1918b. *Raum-Zeit-Materie. Vorlesungen über die allgemeine Relativitätstheorie*. Berlin: Springer.

———. 1919. *Raum-Zeit-Materie. Vorlesungen über die allgemeine Relativitätstheorie*. 3rd ed.

———. 1921. *Raum-Zeit-Materie. Vorlesungen über die allgemeine Relativitätstheorie*. 4th ed.

———. 1923. *Raum-Zeit-Materie. Vorlesungen über die allgemeine Relativitätstheorie*. 5th ed.

Zuckerman, Harriet and Joshua Lederberg. 1986. "Forty years of genetic recombination in bacteria: postmature scientific discovery?" *Nature* 324: 629–631.

HERMANN GRASSMANN

ON THE RELATION OF NON-EUCLIDEAN GEOMETRY TO EXTENSION THEORY

Originally published as Appendix 1 (1877) to "A New Branch of Mathematics: The 'Ausdehnungslehre' of 1844 and Other Works" (Chicago: Open Court, 1995), pp. 279–280.

(Cf. §§15–23)[1]

To the detriment of science, the entire presentation in §§15–23 still remains almost totally unnoticed. Neither Riemann in his *Habilitationsschrift*[1] of 1854, first published in 1867, nor Helmholtz,[2] in his paper "Über die Tatsachen, welche der Geometrie zur Grunde liegen" (1868), nor even in his excellent lecture "Über den Ursprung und die Bedeutung der geometrischen Axiome" (1876) mention it, even though the foundations of geometry come into view much more simply than in those later publications.

In extension theory the straight line is quite special and, in contrast to Euclid, is the foundation for geometric definitions. In §16 the plane is defined as a collection of parallels that intersect a straight line, and space as a collection of parallels that intersect a plane; geometry can proceed no further, but the abstract science is not so limited. Since all points of a straight line may be numerically derived from two of its points, the straight line appears as a simple elementary domain of second order, and correspondingly the plane as a simple elementary domain of third, and ultimately space as one of fourth order.[3]

Thus for example the points of a plane are numerically derivable from three noncollinear points, e.g. by numbers x_1, x_2, x_3. Upon establishing a homogeneous equa-

1 Here is meant his *Habilitationsrede* "Über die Hypothesen, welche der Geometrie zu Grunde liegen," *Ges. Werke* 1st ed. p. 54ff, 2nd ed. p. 272ff.

2 The article appears in *Gott. Nachr.*, 1878, pp. 193–221, cf. also *Ges. wiss. Abh.*, vol. II., pp. 618–639. The lecture is found in his *Vortragen und Reden*, vol. II., p. 1 ff; Braunschweig: 1884.

3 To forestall confusion, I observe that the *displacements* in a plane form an elementary *extensive* domain of second order, those in space an elementary *extensive* domain of third order, and in general the displacements in a simple *elementary* domain of $(n+1)$ -th order an elementary *extensive* domain of n-th order.

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 4
Gravitation in the Twilight of Classical Physics: The Promise of Mathematics.

tion between these three numbers, the collection of points satisfying this equation is reduced to a domain of second order. If this homogeneous equation is of first degree, then the domain so defined is elementary, that is a straight line; if however that equation is of higher degree it forms a curved line for which only some of the longimetric laws for the straight line are valid.

Turning to space, each of its points is numerically derivable from four points forming a tetrahedron, by four numbers $x_1, \ldots, x_4$. If, among these magnitudes, there exist two mutually independent homogeneous equations, neither of them of first degree, we then obtain doubly curved lines for which again only a part of the longimetric laws are true.

Now if we proceed another step beyond space, as a domain of fourth order, to a domain of fifth order (which does not exist geometrically), then one has five basis numbers $x_1, \ldots, x_5$, and if a homogeneous equation of first degree holds between them, then one returns to the simple elementary domain of fourth order, that is to Euclidean space. On the other hand, upon imposing on them a homogeneous equation of higher degree one also produces elementary domains of fourth order, but ones to which the Euclidean axioms no longer apply, and thus as it were to non-Euclidean spaces;[4] furthermore, one can proceed to an elementary domain of sixth order, and between the six determining numbers assume two higher homogeneous equations to obtain once more new elementary domains of fourth order,[5] and can in this way form an infinite sequence of non-Euclidean spaces, the equations of which immediately illuminate the extent to which the Euclidean axioms apply.

Thus extension theory also provides a fully adequate and completely general basis for these and similar considerations.

EDITORIAL NOTE

[1] The reference §§15–23 is to the body of Grassmann's text, which has not been reproduced in this book.

---

4 Thus for example resulting in Helmholtz's spherical space if one assumes a certain homogeneous equation of second degree between the five basis numbers mentioned above (or more generally a curved space upon adoption of an equation of arbitrary degree).

5 One could perhaps call such a space doubly curved, in contrast to the (simple) curved space just mentioned.

TULLIO LEVI-CIVITA

NOTION OF PARALLELISM ON A GENERAL MANIFOLD AND CONSEQUENT GEOMETRICAL SPECIFICATION OF THE RIEMANNIAN CURVATURE (EXCERPTS)

MEMORIA DI T. LEVI-CIVITA (PADOVA)

Originally published as "Nozione di parallelismo in una varietà qualunque e conseguente specificazione geometrica della curvatura riemanniana" in Circolo Matematico di Palermo. Rendiconti, Vol. 42, 1916, pp. 173–204. Received 24 December 1916. Author's date: Padova, November 1916. The Introduction, §15, and the Critical Note are translated here.[1]

INTRODUCTION

Einstein's theory of relativity (now corroborated by the explanation of the famous secular inequality, revealed by observations on Mercury's perihelion, which was not predicted by Newton's law) considers the geometrical structure of space as very tenuously, but nonetheless intimately, dependent on the physical phenomena taking place in it, differently from classical theories, which assume the whole physical space as given *a priori*. The mathematical development of Einstein's grandiose conception (which finds in Ricci's absolute differential calculus its natural algorithmic instrument) utilizes as an essential element the curvature of a certain four-dimensional manifold and the Riemann symbols relative to it. Meeting these symbols—or, better said, continuously using them—in questions of such a general interest, led me to investigate whether it would be possible to somewhat reduce the formal apparatus commonly used in order to introduce them and to establish their covariant behaviour.[1]

Some progress in this direction is actually possible, and essentially forms the content of sections 15 and 16 of the present paper, which, initially conceived with this

1 Cf. e.g. L. Bianchi, *Lezioni di geometria differenziale*, Vol. I (Pisa, Spoerri, 1902), pp. 69–72.

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 4
Gravitation in the Twilight of Classical Physics: The Promise of Mathematics.

only purpose, gradually expanded to make some room for the geometric interpretation too.

At first, I thought I would undoubtedly find it [that interpretation] in Riemann's original works "*Über die Hypothesen welche [der] Geometrie zu Grunde liegen*" and "*Commentatio Mathematica…*,"[2] but only an embryo of it can be found there. Indeed, on the one hand, looking closely at the quoted sources, one gets the impression that Riemann had actually in mind the characterization of the intrinsic and
[174] invariant curvature that shall be specified here (sections 17–18). | On the other hand, neither in Riemann nor in Weber's explicative comment,[3] is to be found a trace of those specifications (notion of parallel directions on a general manifold and consideration of a geodetic infinitesimal quadrangle with two parallel sides), that we shall recognize as indispensable from the geometrical point of view. Moreover, one cannot—or at least, I was not able to—justify the formal step in terms of which, according to Riemann, from the premises, which are impeccable, one should obtain an equivalently impeccable final expression of the curvature.

I will present to the reader this doubt of mine, providing him with the elements required to form an opinion in a final critical note.

The first and more extended part of the paper (sections 1–14) is devoted to an introduction and an illustration of the notion of parallelism in a V_n with any metric.

One begins with the infinitesimal field, trying to characterize the parallelism of two directions $(\alpha), (\alpha')$ through two very close points P and P'. To this purpose, one should remember that any manifold V_n can be looked at as embedded in an Euclidean space S_N of a sufficiently high number N of dimensions, and notice, first of all, that, for any direction (f) of S_N through P, ordinary parallelism would require, in such a space,

$$\mathrm{angle}(\widehat{f)(\alpha}) = \mathrm{angle}(\widehat{f)(\alpha'})$$

for any (f). Now, parallelism in V_n is defined limiting oneself to require that the condition be satisfied *for all the* (f) *belonging to* V_n (namely to the set of directions of S_N tangent to V_n in P).

In order to justify this definition, it should be noted that, while for an Euclidean V_n, it reproduces, as is necessary, the elementary behaviour, it has in any case an intrinsic character, since it ultimately turns out to depend only on the metric of V_n, and not on the auxiliary ambient space S_N as well. Indeed, the analytic version of our definition of parallelism is realized as follows: Once V_n is given general coordinates x_i $(i = 1, 2, \ldots, n)$, , let dx_i be the increments corresponding to the displacement from P to P'; $\xi^{(i)}$ the parameters of a generic direction (α) through P; $\xi^{(i)} + d\xi^{(i)}$ those belonging to an infinitely close direction (α') through P'. The condition of parallelism is expressed by the n equations

2 B. Riemann, *Gesammelte mathematische Werke* (Leipzig, Teubner, 1876), pp. 261–263, 381–382.

3 loc. cit.[2] pp. 384–389.

$$d\xi^{(i)} + \sum_{j,l=1}^{n} \begin{Bmatrix} jl \\ i \end{Bmatrix} dx_j\, \xi^{(l)} = 0 \qquad (i = 1, 2, \ldots, n), \tag{A}$$

where the $\begin{Bmatrix} jl \\ i \end{Bmatrix}$ denote well-known Christoffel symbols.

Once the law by means of which one goes from one point to an infinitely close point is acquired, one is provided with all the means required in order to perform the transport of parallel directions along any curve C. If $x_i = x_i(s)$ are its parametric equations, | one only needs to consider in eqs. (A) the x_i and subordinately the [175]

$$\begin{Bmatrix} jl \\ i \end{Bmatrix},$$

as assigned functions, the $\xi^{(i)}$ as functions to be determined of the parameter s, and one has the ordinary linear system

$$\frac{d\xi^{(i)}}{ds} + \sum_{j,l=1}^{n} \begin{Bmatrix} jl \\ i \end{Bmatrix} \frac{dx_j}{ds}\, \xi^{(l)} = 0 \qquad (i = 1, 2, \ldots, n)$$

reducible to a typical form (said "with hunched determinant"), which already appeared in other researches and was the object of a systematic investigation by Mr. Eiesland,[4] Laura,[5] Darboux,[6] Vessiot.[7]

Here is some geometrical consequence.

1. The direction through a generic point P parallel to a direction (α) through any other point P_0 depends in general on the path followed from P_0 to P. Independence from the path is an exclusive property of Euclidean manifolds.

2. Along a given geodesic, directions of the tangents are parallel, a result that generalizes an obvious feature of the straight line in a Euclidean space (the one that Euclid himself sets as a primordial intuitive notion of a straight line at the beginning of *Elements*).

3. The parallel transport along any path of two concurrent directions preserves their angle. By this we obviously mean that the angle formed by two generic direc-

---

4 J. Eiesland, "On the Integration of a System of Differential Equations in Kinematics" *American Journal of Mathematics*, vol. XXVIII (1906), pp. 17–42.

5 E. Laura, "Sulla integrazione di un sistema di quattro equazioni differenziali lineari a determinante gobbo per mezzo di due equazioni di Riggati" *Atti della Accademia delle Scienze di Torino*, vol. XLII, 1906–1907, pp. 1089–1108; vol. XLII, 1907–1908, pp. 358–378.

6 G. Darboux, "Sur certains systèmes d'equations linéaires" *Comptes rendu hebdomadaires des séances de l'Académie des Sciences*, t. CXLVIII (1er semestre 1909), pp. 332–335, and "Sur les systèmes d' équations différentielles homogènes" (*Ibid*., pp. 673–679 and pp. 745–754.

7 E. Vessiot, "Sur l'intégration des systèmes linéaires à déterminant gauche" *Comptes rendus hebdomadaires des séances de l'Académie des Sciences*, t. CXLVIII (1er semestre 1909), pp. 332–335.

tions through the same point is also the angle formed by their parallels through another point. Taking into account the mentioned property of the geodesics, one derives as a corollary that, along a geodesic, parallel directions are always equally inclined with respect to the geodesic itself. If in particular one deals with a V_2, this condition is also sufficient; hence, for ordinary surfaces, parallelism along a geodesic is equivalent to isogonality.

I am not specifying how the content is arranged in the various sections. A look at the summary at the end of the paper will supply the necessary information.

[…]

[195]

§15.
2° ORDER DIFFERENTIALS - INVARIANT DETERMINATIONS - RICCI'S LEMMA.

In a given investigation, let the independent variables, for instance $x_1, x_2, \ldots, x_n$, be fixed. As is known from the calculus, it is always legitimate to consider the second order differentials $d^2x_1, d^2x_2, \ldots, d^2x_n$ as vanishing. Such a convention, however, does not have an invariant character with respect to changes of variables. Indeed, if the x_i are replaced by n independent combinations thereof $\chi_i(x_1, x_2, \ldots, x_n)$, the second differentials

$$d^2\chi_i = \sum_{j,l=1}^{n} \frac{\partial^2 \chi_i}{\partial x_j \, \partial x_l} dx_j \, dx_l$$

(computed on the basis of the hypothesis $d^2x_i = 0$) turn out to be in general different from zero.

If to the variables a quadratic differential form is associated, referring for instance to the metric of a V_n (in the notations of the preceding sections), the way to an invariant characterization is facilitated. It is sufficient to assume the d^2x_i (not vanishing, but) defined as follows:

$$d^2x_i + \sum_{j,l=1}^{n} \begin{Bmatrix} jl \\ i \end{Bmatrix} dx_j \, dx_l = 0 \qquad (i = 1, 2, \ldots, n).$$

From the geodesic equations (sec. 7), multiplied by ds^2, it appears that such d^2x_i are those belonging to the variables along the geodesic through the generic point $(x_1, x_2, \ldots, x_n)$ in the similarly generic direction $(dx_1, dx_2, \ldots, dx_n)$. This geometric interpretation guarantees *a priori* that the above convention has the desired invariant character, making unnecessary a material check, which, on the other hand, could be done straightforwardly.

Similarly for the superposition of two independent systems of increments dx_i and δx_i, one might set $d\delta x_i = \delta d x_i = 0$, but, while the invertibility of the increments d

and δ has, as is easily checked, an invariant character, the same does not hold as far as setting $d\delta x_i = 0$ is concerned. We shall replace them by:

$$d\delta x_i + \sum_{j,l=1}^{n} \left\{ \begin{matrix} jl \\ i \end{matrix} \right\} dx_i \delta x_j = 0, \tag{31}$$

which imply

$$d\delta x_i = \delta dx_i, \tag{31'}$$

and contain, as a particular case, for $d = \delta$, the previous expressions for the $d^2 x_i$.

The invariance of eqs. (31) with respect to changes of variables can be derived from the geometric interpretation as well. One only needs to observe that, writing $\delta x_i = \varepsilon \xi^{(i)}$ (with ε an infinitesimal constant), eqs. (31) become identical with eqs. (I_a), so that they express how the δx_i must be altered, as a consequence of the displacement $(dx_1, dx_2, \ldots, dx_n)$, | in order that they define directions parallel to one [196]
another. This invariant property, besides verifying it directly, could be controlled with an elegant formal device sketched by Riemann[8] and made explicit by Weber.[9]

From eqs. (31), taking into account

$$d a_{ik} = \sum_{j=1}^{n} \frac{\partial a_{ik}}{\partial x_j} dx_j = \frac{1}{2} \sum_{j=1}^{n} (a_{ij,k} + a_{jk,i}) dx_j = \frac{1}{2} \sum_{j,l=1}^{n} \left[a_{lk} \left\{ \begin{matrix} ij \\ l \end{matrix} \right\} + a_{li} \left\{ \begin{matrix} jk \\ l \end{matrix} \right\} \right] dx_j,$$

it follows identically[2]

$$d \sum_{i,k=1}^{n} a_{ik}\, \delta x_h\, \delta x_k = 0, \tag{32}$$

as well as

$$d \sum_{i,k=1}^{n} a_{ik}\, dx_h\, \delta x_k = 0.$$

These relations are equivalent to the well-known result of the absolute differential calculus that the covariantly derived system of the coefficients a_{ik} of the fundamental form vanish identically (Ricci's lemma).

[...]

1. CRITICAL NOTE [201]

We have already pointed out in sec. 15, the expressions (31)

8 loc. cit.[2] p. 381.
9 loc. cit.[2] p. 388.

$$d\delta x_i + \sum_{j,l=1}^{n} \begin{Bmatrix} jl \\ i \end{Bmatrix} dx_j \, \delta x_l = 0$$

[202] | of the second order differentials do not differ from those which are arrived at by making Riemann's comprehensive definition explicit.

From the same section it is also deduced that, with these expressions of the second differentials, one has identically (Ricci's lemma)

$$\delta\, ds^2 = d\,\delta s^2 = d\Phi = \delta\Phi = 0, \tag{48}$$

where

$$\Phi = \sum_{i,k=1}^{n} a_{ik}\, dx_i \delta x_k,$$

and ds^2, δs^2 stand, of course, respectively for

$$\sum_{i,k=1}^{n} a_{ik}\, dx_i\, dx_k \qquad \text{and} \qquad \sum_{i,k=1}^{n} a_{ik}\, \delta x_i\, \delta x_k.$$

In this context the meaning to be attributed to the trinomial considered by Riemann:

$$R = \delta^2 ds^2 - 2\, d\delta\Phi + d^2\, \delta s^2$$

seems unambiguous, and such meaning, by virtue of (48), implies necessarily $R = 0$.

Riemann states[10] instead that: "Haec expressio (that is R) invenietur $= J$" (J having the value (45)). Weber, in his elucidations, dwells on the way the second differentials are introduced,[11] but, after deriving their explicit expression, simply says:[12] "woraus man leicht den Ausdruck erhält $R = J$."

Probably, there is just some blemish in Riemann's explicit expression for R that obscures the concept. I flatter myself that I have substantially reconstructed such a concept, but I was not able to adjust the symbol. If this can be achieved, it will be the case to pay full tribute, on this point too, to Riemann's genius.

---

10 loc. cit.[2] p. 381

11 Adding, with no further justification, the supplementary conditions

$$d^2\delta s^2 = \delta^2 ds^2 = -2d\delta\Phi.$$

By virtue of (48) (and provided the formulae are read as they are actually written) everything vanishes.

12 loc. cit.[2] p. 388

I shall end with an observation about the calculation of the curvature with reference to particular variables, which is indicated by Riemann[13] and developed by Weber.[14] Here is, to begin with, what the matter is about.

Let us choose coordinates $x_1, x_2, \ldots, x_n$ such that, *at a given point* P, all the symbols

$$\begin{Bmatrix} jl \\ i \end{Bmatrix}$$

vanish (which is always possible, as was pointed out by Weber). Let us consider two independent sets of differentials $dx_i, \delta x_i$, considering | all the second differentials [203]
$d^2x_i, d\delta x_i, \delta dx_i, \delta^2 x_i$ as vanishing. Let P' and Q denote the points of coordinates $x_i + dx_i, x_i + \delta x_i$, and a'_{hk} the coefficients of the squared line element in P'. Set, in particular,

$$(\delta s^2)_{P'} = \sum_{h,k=1}^{n} a'_{hk}\, \delta x_h\, \delta x_k,$$

let us apply to the a' the Taylor expansion with respect to the increments d up to the second order. In such approximation one has

$$(\delta s^2)_{P'} = \delta s^2 + \frac{1}{2} \sum_{h,k,j,l=1}^{n} \frac{\partial^2 a_{hk}}{\partial x_j\, \partial x_l} dx_j\, dx_l\, \delta x_h\, \delta x_k\,,$$

δs^2 and the second derivatives referring, of course, to P. As shown by Weber, due to the way the variables were fixed, special relations hold between the values of the second derivatives of the a_{hk} in P. Taking them into account, one finds, with some manipulation,

$$(\delta s^2)_{P'} = \delta s^2 + \frac{1}{3} \sum_{h,k,j,l=1}^{n} \left[\frac{\partial^2 a_{hk}}{\partial x_j \partial x_l} + \frac{\partial^2 a_{jl}}{\partial x_h \partial x_k} - \frac{\partial a_{hj}}{\partial x_k \partial x_l} - \frac{\partial a_{kl}}{\partial x_h \partial x_j}\right] dx_j\, dx_l \delta x_h \delta x_k.$$

The sum can be looked at as the expression which, as the basis of formula (45), is taken on by $-I$, when variables x specified as above are adopted.

Therefore, taking into account (47), we derive

$$\frac{(\delta s^2)_{P'} - \delta s^2}{(ds\,\delta s \sin\psi)^2} = -\frac{1}{3} K, \tag{49}$$

which Riemann, in the quoted passage, states in words (multiplying both sides by 4, in order to show up the area of the triangle $PP'Q$ in the denominator).

13 loc. cit.[2] p. 261
14 loc. cit.[2] pp. 384–387

I come, at last, to my point:

If Q^* denotes the extremum of the line element $(\delta s^2)_{P'}$ (corresponding to the increments δx_i), eq. (49) can be written

$$\frac{\overline{P'Q^*}^2 - \overline{PQ}^2}{(ds\delta s \sin\psi)^2} = -\frac{1}{3}K; \tag{49'}$$

whereas eq. (46) (with an overall change of sign) reads

$$\frac{\overline{P'Q'}^2 - \overline{PQ}^2}{(ds\delta s \sin\psi)^2} = -K. \tag{46'}$$

As can be seen, the right-hand sides are in the ratio 1 to 3. The lack of coincidence is manifestly due to the fact that the point Q' (fourth vertex of the parallelogrammoid), which is reached through the invariant procedure, is well distinct from Riemann's point Q^*, analytically defined with reference to particular variables.

To localize the discrepancy about the formulae, it helps to work out our procedure too (as is of course allowed given its invariant character) in Riemann's special vari-
[204] ables. | Eqs. (31) give then, in so far as they refer to the point P,

$$d^2x_i = d\delta x_i = \delta dx_i = \delta^2 x_i = 0;$$

but it does not follow that the higher differentials, such as $\delta d^2 x_i, d^2\delta x_i$, etc. must vanish at the same point as well. Riemann's calculation on the contrary is based on the hypothesis that all differentials of an order higher than the first must vanish: a legitimate hypothesis too, but not one endowed with an invariant character (with respect to changes of variables). Therefore, it should not come as a surprise that the results are different: one should rather notice the fortuitous analogy between formulae $(49')$ and $(46')$, whose right-hand sides differ only by a numerical factor.

EDITORIAL NOTES

[1] This text has been translated by Silvio Bergia.

[2] In the original text, the index $_h$ is missing from the expression δx_h in eq. (32).

HERMANN WEYL

PURELY INFINITESIMAL GEOMETRY (EXCERPT)

Originally published as "Reine Infinitesimalgeometrie" in Mathematische Zeitschrift 2, 1918, pp. 384–411. Excerpt covers pp. 384–401.

1. INTRODUCTION: CONCERNING THE RELATION BETWEEN GEOMETRY AND PHYSICS

The real world, into which we have been placed by virtue of our consciousness, *is not there* simply and all at once, but *is happening*; it passes, annihilated and newly born at each instant, a continuous one-dimensional succession of states in *time*. The arena of this temporal happening is a three-dimensional Euclidean *space*. Its properties are investigated by *geometry*, the task of *physics* by comparison is to conceptually comprehend the real that exists in space and to fathom the laws persisting in its fleeting appearances. Therefore, physics is a science which has geometry as its foundation; the concepts however, through which it represents reality—matter, electricity, force, energy, electromagnetic field, gravitational field, etc.—belong to an entirely different sphere than the geometrical.

This old view concerning the relation between the form and the content of reality, between geometry and physics, has been overturned by Einstein's theory of relativity.[1] The *special theory of relativity* led to the insight that space and time are fused into an indissoluble whole which shall here be called the *world*; the world, according to this theory, is a four-dimensional Euclidean manifold—Euclidean with the modification that the underlying quadratic form of the world metric is not positive definite but is of inertial index 1. The *general theory of relativity*, in accordance with the spirit of modern physics of local action [*Nahewirkungsphysik*], admits that as valid only in the infinitely small, hence for the world metric it makes use of the more general concept of a metric [*Maßbestimmung*] based on a quadratic *differential* form, developed by Riemann in his habilitation lecture. | But what is new in principle in this [385]
is the insight that the metric is not a property of the world in itself, rather, spacetime as the form of appearances is a completely formless four-dimensional continuum in

1 I refer to the presentation in my book *Raum, Zeit, Materie*, Springer 1918 (in the sequel cited as RZM), and the literature cited there.

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 4
Gravitation in the Twilight of Classical Physics: The Promise of Mathematics.

the sense of analysis situs. The metric, however, expresses something real that exists in the world, which produces physical effects on matter by means of centrifugal and gravitational forces, and whose state is in turn determined according to natural laws by the distribution and composition of matter. By removing from Riemannian geometry, which claims to be a purely "local geometry," [*Nahe-Geometrie*] an inconsequence still currently adhering to it, ejecting one last element of non-local geometry [*ferngeometrisches Element*] which it had carried along from its Euclidean past, I arrived at a world metric from which not only arises gravitation, but also the electromagnetic effects, and therefore, as one may assume with good reason, accounts for all physical processes.[2] According to this theory, *everything real that exists in the world is a manifestation of the world metric*; the physical concepts are none other than the geometric ones. The only difference that exists between geometry and physics is that geometry fathoms in general what lies in the nature of the metric concepts,[3] whereas physics has to determine the law by which the real world is distinguished among all the four-dimensional metric spaces possible according to geometry and pursue its consequences.[4]

In this note, I want to develop that *purely infinitesimal geometry* which, according to my conviction, contains the physical world as a special case. The construction of the local geometry proceeds adequately in three steps. On the first step stands the *continuum* in the sense of analysis situs, without any metric—physically speaking, *the empty world*; on the second the *affinely connected continuum*—I so call a manifold in which the concept of infinitesimal parallel displacement of vectors is meaningful; in | physics, the affine connection appears as *the gravitational field*—; finally [386]
on the third, the *metric* continuum—physically: *the "aether,"* whose states are manifested in the phenomena of matter and electricity.

2. SITUS-MANIFOLD (EMPTY WORLD)

As a consequence of the difficulty in grasping the intuitive character of the continuous connection by means of a purely logical construction, a completely satisfactory analysis of the concept of an *n -dimensional manifold* is not possible today.[5] The following is sufficient for us: An n-dimensional manifold refers to n coordinates $x_1\ x_2\ \dots\ x_n$, of which each possesses at each point of the manifold a particular numerical value: different sets of values of the coordinates correspond to different

---

2 A first communication about this appeared under the title "Gravitation und Elektrizität" in *Sitzungsber. d. K. Preuß. Akad. d. Wissenschaften* 1918, p. 465.

3 Naturally, traditional geometry leaves the path of this, its principal task, and immediately takes on the less specific one by not making space itself anymore the object of its investigation, but the structures possible in space, special classes and their properties they are endowed with on the basis of the space-metric.

4 I am bold enough to believe that the totality of physical phenomena can be derived from a single universal world law of greatest mathematical simplicity.

5 See also H. Weyl, *Das Kontinuum* (Leipzig 1918), specifically pp. 77 ff.

points; if $\bar{x}_1\ \bar{x}_2\ \ldots\ \bar{x}_n$ is a second system of coordinates then there exist between the x- and the $\bar{x}$-coordinates of the same arbitrary point regular relations

$$x_i = f_i(\bar{x}_1\ \bar{x}_2\ \ldots\ \bar{x}_n) \qquad (i = 1, 2, \ldots, n),$$

where f_i denote purely logically-arithmetically constructible functions; of these we presuppose not only that they are continuous, but also that they posses continuous derivatives

$$\alpha_{ik} = \frac{\partial f_i}{\partial \bar{x}_k},$$

whose determinant does not vanish. The last condition is necessary and sufficient for the affine geometry to be valid in the infinitely small, namely that there exist invertible linear relationships between the coordinate differentials in the two systems:

$$dx_i = \sum_k \alpha_{ik} d\bar{x}_k. \tag{1}$$

We assume the existence and continuity of higher order differentials where required during the course of the investigation. In any case, the concept of the continuous and continuously differentiable point-function, if necessary also the $2, 3, \ldots$ times continuously differentiable, has therefore an invariant meaning independent of the coordinate system. The coordinates themselves are such functions. An n-dimensional manifold for which we regard no properties other than those lying within the concept of an n-dimensional manifold, we call—in physical terminology—an (n-dimensional) *empty world*. |

The relative coordinates dx_i of a point $P' = (x_i + dx_i)$ infinitely close to the point $P = (x_i)$ are the components of a *line element* in P, or an *infinitesimal displacement* $\overrightarrow{PP'}$ of P. In going to a different coordinate system the formulae (1) apply for these components, the α_{ik} denoting the corresponding derivatives at the point P. More generally, on the basis of a definite coordinate system in the neighborhood of P, any n numbers ξ^i $(i = 1, 2, \ldots, n)$ given in a definite order, characterize at the point P a *vector* (or a *displacement*) at P. The components ξ^i respectively $\bar{\xi}^i$ of the same vector in any two coordinate systems, the "unbarred" one and the "barred" one, are related by the same linear transformation equations (1): [387]

$$\xi^i = \sum_k \alpha_{ik} \bar{\xi}^k.$$

Vectors at P can be added and multiplied by numbers; thus they form a "linear" or "affine" totality [*Gesamtheit*]. With each coordinate system are associated n "unit vectors" e_i at P, namely those vectors which in the coordinate system in question have the components

$$\begin{array}{c|c} \mathrm{e}_1 & 1, 0, 0, \ldots, 0 \\ \mathrm{e}_2 & 0, 1, 0, \ldots, 0 \\ .. & \ldots \ \ldots \ \ldots \ \ldots \ \ldots \\ \mathrm{e}_n & 0, 0, 0, \ldots, 1 \end{array}$$

Any two (linearly independent) line elements at P with the components dx_i and δx_i respectively span a (two-dimensional) area element at P with the components

$$dx_i \delta x_k - dx_k \delta x_i = \Delta x_{ik},$$

each three (independent) line elements dx_i, δx_i, $\mathfrak{d} x_i$ at P, a (three-dimensional) volume element with the components

$$\begin{vmatrix} dx_i & dx_k & dx_l \\ \delta x_i & \delta x_k & \delta x_l \\ \mathfrak{d} x_i & \mathfrak{d} x_k & \mathfrak{d} x_l \end{vmatrix} = \Delta x_{ikl} ;$$

etc. A linear form depending on an arbitrary line- or area- or volume- or ... element at P is called a *linear tensor* of order $1, 2, 3 \ldots$ respectively. By using a particular coordinate system, the coefficients a of this linear form

$$\sum_i a_i dx_i, \quad \text{resp.} \quad \frac{1}{2!}\sum_{ik} a_{ik} \Delta x_{ik}, \quad \frac{1}{3!}\sum_{ikl} a_{ikl} \Delta x_{ikl} \ , \ldots$$

[388] | can be uniquely normalized through the alternation requirement; e.g., for the case just written down this implies that the triple of indices (ikl), which arise through an even permutation of itself corresponds to the same coefficient a_{ikl}, whereas under odd permutations the coefficient changes into its negative, that is

$$a_{ikl} = a_{kli} = a_{lik} = -a_{kil} = -a_{lki} = -a_{ilk}.$$

The coefficients normalized in this manner are called the *components* of the tensor in question. From a scalar field f one obtains through differentiation a linear tensor field of order 1 with the components

$$f_i = \frac{\partial f}{\partial x_i};$$

from a linear tensor field f_i of order 1, one of 2nd order:

$$f_{ik} = \frac{\partial f_i}{\partial x_k} - \frac{\partial f_k}{\partial x_i};$$

from one of order 2, a linear tensor field of order 3:

$$f_{ikl} = \frac{\partial f_{kl}}{\partial x_i} + \frac{\partial f_{li}}{\partial x_k} + \frac{\partial f_{ik}}{\partial x_l};$$

etc. These operations are independent of the coordinate system used.[6]

A linear tensor of the 1st order at P we will call a *force* acting there. Assuming a definite coordinate system, such a force is thus characterized by n numbers ξ_i, which transform contragrediently to the components of the displacement under a change to another coordinate system:

$$\xi_i = \sum_k \alpha_{ki}\xi_k.$$

If η^i are the components of an arbitrary displacement at P, then

$$\sum_i \xi_i \eta^i$$

is an invariant. By a *tensor* at P, one generally understands a linear form of one or more arbitrary displacements and forces at P. For example, if we are dealing with a linear form of three arbitrary displacements ξ, η, ζ and two arbitrary forces ρ, σ:

$$\sum a_{ikl}^{pq} \xi^i \eta^k \zeta^l \rho_p \sigma_q,$$

then we speak of a tensor of order 5, with the components a being covariant with respect to the indices ikl and contravariant with respect to the indices pq. A displacement is itself a contravariant | tensor of 1st order, the force a covariant one. [389] The fundamental operations of tensor algebra are:[7]

1. Addition of tensors and multiplication by a number;
2. Multiplication of tensors;
3. Contraction.

Accordingly, tensor algebra can already be constructed in the empty world—it does not presuppose any metric [*Maßbestimmung*]—of tensor analysis, however, only that of "linear" tensors.

A "*motion*" in our manifold is given, if to each value s of a real parameter is assigned a point in a continuous manner; by using the coordinate system x_i, the motion is expressed by the formulae $x_i = x_i(s)$, in which the x_i on the right are to be understood as function symbols. If we presuppose continuous differentiability, then we obtain, independently of the coordinate system, for each point $P = (s)$ of the motion a vector at P with the components:

6 RZM, §13.

7 RZM, §6.

$$u^i = \frac{dx_i}{ds},$$

the *velocity*. Two motions, arising from one another through continuous monotonic transformation of the parameter s describe the same *curve*.

3. AFFINELY CONNECTED MANIFOLD (WORLD WITH GRAVITATIONAL FIELD)

3.1 The Concept of the Affine Connection

If P' is infinitely close to the fixed point P, then P' is *affinely connected with* P, if for each vector at P it is determined into which vector at P' it will transform under *parallel displacement* from P to P'. The parallel displacement of all vectors at P from there to P' must evidently satisfy the following requirement.

A. *The transfer of the totality of vectors from P to the infinitely close point P' by means of parallel displacement produces an affine transformation of the vectors at P to the vectors at P'.*

If we use a coordinate system in which P has the coordinates x_i, P' the coordinates $x_i + dx_i$, an arbitrary vector at P the components ξ^i, and the vector at P', that results from it through parallel displacement to P', the components $\xi^i + d\xi^i$, then
[390] $d\xi^i$ must therefore depend linearly on the ξ^i: |

$$d\xi^i = -\sum_r d\gamma^i{}_r \xi^r.$$

$d\gamma^i{}_r$ are infinitesimal quantities which depend only on the point P and the displacement $\overrightarrow{PP'}$ with the components dx_i, but not on the vector ξ subject to parallel displacement. From now on, we consider affinely connected manifolds; in such a manifold, each point P is affinely connected to all its infinitely close points. A second requirement is still to be imposed on the concept of parallel displacement, that of *commutativity.*

B. *If P_1, P_2 are two points infinitely close to P and if the infinitesimal vector $\overrightarrow{PP_1}$ becomes $\overrightarrow{P_2P_{21}}$ under parallel displacement from P to P_2, and $\overrightarrow{PP_2}$ becomes $\overrightarrow{P_1P_{12}}$ under parallel displacement to P_1, then the points P_{12} and P_{21} coincide.* (An infinitely small parallelogram results.)

If we denote the components of $\overrightarrow{PP_1}$ by dx_i, and those of $\overrightarrow{PP_2}$ by δx_i, then the requirement in question obviously implies that

$$d\delta x_i = -\sum_r d\gamma^i{}_r \cdot \delta x_r \tag{2}$$

is a symmetric function of the two line elements d and δ. Consequently, $d\gamma^{i}{}_{r}$ must be a linear form of the differentials dx_i,

$$d\gamma^{i}{}_{r} = -\sum_{s} \Gamma^{i}{}_{rs}\, dx_s,$$

and the coefficients Γ, the "*components of the affine connection*," which depend only on the location of P, must satisfy the symmetry condition

$$\Gamma^{i}{}_{sr} = \Gamma^{i}{}_{rs}.$$

Because of the way in which the infinitesimal quantities are dealt with in the formulation of the requirement B, it could be objected that the latter lacks a precise meaning. Therefore, we want to determine explicitly through a rigorous proof that the symmetry of (2) is a condition independent of the coordinate system. For this purpose, we make use of a (twice differentiable) scalar field f. From the formula for the total differential

$$df = \sum_{i} \frac{\partial f}{\partial x_i} dx_i$$

we infer, that if ξ^{i} are the components of an arbitrary vector at P, |

$$df = \sum_{i} \frac{\partial f}{\partial x_i} \xi^{i}$$ [391]

is an invariant independent of the coordinate system. We form its variation under a second infinitesimal displacement δ, in which the vector ξ shall be displaced parallel to itself from P to P_2, and obtain

$$\delta df = \sum_{ik} \frac{\partial^2 f}{\partial x_i \partial x_k} \xi^{i} \delta x_k - \sum_{ir} \frac{\partial f}{\partial x_i} \cdot d\gamma^{i}{}_{r} \xi^{r}.$$

If we replace in this expression ξ^{i} again by dx_i and subtract from this equation the one obtained by interchanging d and δ, then the invariant

$$\Delta f = (\delta d - d\delta) f = \sum_{i} \left\{ \frac{\partial f}{\partial x_i} \sum_{r} (d\gamma^{i}{}_{r} \delta x_r - \delta\gamma^{i}{}_{r} dx_r) \right\}.$$

results. The relations

$$\sum_{r} (d\gamma^{i}{}_{r} \delta x_r - \delta\gamma^{i}{}_{r} dx_r) = 0$$

contain the necessary and sufficient condition that for any scalar field f the equation $\Delta f = 0$ is satisfied.

In physical terms, an affinely connected continuum is to be described as a world in which a *gravitational field* exists. The quantities $\Gamma^i{}_{rs}$ are the components of the gravitational field. The formulae, according to which these components transform in changing from one coordinate system to another, we need not state here. Under linear transformations the $\Gamma^i{}_{rs}$ behave with respect to r and s like the covariant components of a tensor and with respect to i like the contravariant components, but lose this character under non-linear transformations. However, the changes $\delta\Gamma^i{}_{rs}$, which are experienced by the quantities Γ, if one arbitrarily varies the affine connection of the manifold, form the components of a generally-invariant tensor of the given character.

What is to be understood by *parallel displacement of a force* at P from there to the infinitely close point P' results from the requirement that the invariant product of this force and an arbitrary vector at P is preserved under parallel displacement. If ξ_i are the components of the force, η^i those of the displacement, then[8]

$$d(\xi_i\eta^i) = (d\xi_i \cdot \eta^i) + \xi_r d\eta^r = (d\xi_i - d\gamma^r{}_i\xi_r)\eta^i = 0$$

yields the formula

$$d\xi_i = \sum_r d\gamma^r{}_i\xi_r.$$

[392] | At each point P, one can introduce a coordinate system x_i of a kind—I call it *geodesic* at P—such that in it, the components of the affine connection $\Gamma^i{}_{rs}$ vanish at the point P. If x_i are initially arbitrary coordinates that vanish at P, and $\Gamma^i{}_{rs}$ designate the components of the affine connection at the point P in this coordinate system, then one obtains a geodesic coordinate system $\bar{x}_i$ via the transformation

$$x_i = \bar{x}_i - \frac{1}{2}\sum_{rs}\Gamma^i{}_{rs}\bar{x}_r\bar{x}_s. \tag{3}$$

Namely, if we consider the $\bar{x}_i$ as independent variables and their differentials $d\bar{x}_i$ as constants, then one has in the sense of Cauchy at $P(\bar{x}_i = 0)$:

$$dx_i = d\bar{x}_i, \qquad d^2x_i = -\Gamma^i{}_{rs}d\bar{x}_r d\bar{x}_s,$$

therefore,

$$d^2x_i + \Gamma^i{}_{rs}d\bar{x}_r d\bar{x}_s = 0.$$

Because of their invariant nature, the last equations in the coordinate system $\bar{x}_i$ become:

$$d^2\bar{x}_i + \bar{\Gamma}^i{}_{rs}d\bar{x}_r d\bar{x}_s = 0.$$

8 In the following we will use Einstein's convention that summation is always to be carried out over indices which occur twice in a formula without our finding it necessary to always place a summation sign in front of it.

For arbitrary constant $d\bar{x}_i$ these are, however, satisfied only if all the $\bar{\Gamma}^i{}_{rs}$ vanish. *Therefore, through an appropriate choice of the coordinate system, the gravitational field can always be made to vanish at a single point*. Through the requirement of "geodesy" at P the coordinates in the neighborhood of P are determined up to linear transformation excluding terms of third order; i.e., if x_i, $\bar{x}_i$ are two coordinate systems geodesic at P, and if the x_i as well as the $\bar{x}_i$ vanish at P, then by neglecting terms in $\bar{x}_i$ of order 3 and higher, linear transformation equations $x_i = \sum_k \alpha_{ik}\bar{x}_k$ with constant coefficients α_{ik} apply.

3.2 Tensor Analysis, Straight Line

Only in an affinely connected space can *tensor analysis* be fully established. If for example f_i^k are the components of a 2nd order tensor field, covariant in i and contravariant in k, then with the aid of an arbitrary displacement ξ and a force η at the point P, we form the invariant

$$f_i^k \xi^i \eta_k$$

and its change under an infinitely small displacement d of the point P, in which ξ and η are displaced parallel with respect to themselves. We have

$$d(f_i^k \xi^i \eta_k) = \frac{\partial f_i^k}{\partial x_l}\xi^i \eta_k dx_l - f_r^k \eta_k d\gamma^r{}_i \xi^i + f_i^r \xi^i d\gamma^k{}_r \eta_k,$$

| and therefore [393]

$$f_{il}^k = \frac{\partial f_i^k}{\partial x_l} - \Gamma_{il}^r f_r^k + \Gamma_{rl}^k f_i^r$$

are the components of 3rd order tensor field, covariant in il and contravariant in k, which arises from the given 2nd order tensor field in a coordinate independent manner.

In the affinely connected space, the concept of *straight or geodesic line* gains a definite meaning. The straight line arises as the trajectory of the initial point of the vector which is displaced in its own direction keeping it parallel to itself; it can therefore be described as that curve the direction of which remains unchanged. If u^i are the components of that vector, then during the course of the motion the equations

$$du^i + \Gamma^i{}_{\alpha\beta} u^\alpha dx_\beta = 0,$$
$$dx_1 : dx_2 : \ldots : dx_n = u^1 : u^2 : \ldots : u^n$$

should always hold. The parameter s used in describing the curve can thus be normalized in such a way that

$$\frac{dx_i}{ds} = u^i$$

identically along s, and the differential equations of the straight line are then

$$w^i \equiv \frac{d^2 x_i}{ds^2} + \Gamma^i_{\alpha\beta}\frac{dx_\alpha}{ds}\frac{dx_\beta}{ds} = 0.$$

For each arbitrary motion $x_i = x_i(s)$, the left hand sides of these equations are the components of a vector invariantly linked to the motion at the point s, the acceleration. Actually, if ξ_i is an arbitrary force at that point, which during the transition to the point $s + ds$ is displaced parallel to itself, then

$$\frac{d(u^i \xi_i)}{ds} = w^i \xi_i.$$

A motion, whose acceleration vanishes identically is called a *translation*. A straight line—this is another way of grasping our above explanation—is to be understood as the trajectory of a translation.

3.3 Curvature

If P and Q are two points connected by a curve, and a vector is given at the first point, then one can displace this vector parallel to itself along the curve from P to Q. The resulting *vector transfer* is however in general *not integrable*; i.e. the vector |
[394] which one ends up with at Q depends on the path along which the transport takes place. Only in the special case of integrability does it make sense to speak of the *same* vector at two different points P and Q; these are understood to be vectors which arise from one another under parallel transport. In this case, the manifold is called *Euclidean*. In such a manifold, special "linear" coordinate systems can be introduced which are distinguished by the fact that equal vectors at different points have equal components. Any two such linear coordinate systems are related by linear transformation equations. In a linear coordinate system the components of the gravitational field vanish identically.

On the infinitely small parallelogram constructed above (§3, I., B.), we attach at the point P an arbitrary vector with components ξ^i and in the first case displace it parallel to itself to P_1, and from there to P_{12}, and in the second case first to P_2, and from there to P_{21}. Since P_{12} and P_{21} coincide, we can form the difference of these two vectors at this point and through this obviously obtain there a vector with the components

$$\Delta\xi^i = \delta d\xi^i - d\delta\xi^i.$$

From

$$d\xi^i = -d\gamma^i_k\,\xi^k = -\Gamma^i{}_{kl}dx_l\xi^k,$$

it follows that

$$\delta d\xi^i = -\frac{\partial \Gamma^i{}_{kl}}{\partial x_m} dx_l \delta x_m \xi^k - \Gamma^i{}_{kl} \delta dx_l \cdot \xi^k + d\gamma^i{}_r \delta\gamma^r{}_k \xi^k,$$

and because of the symmetry of δdx_l:

$$\Delta\xi^i = \left\{ \left(\frac{\partial \Gamma^i{}_{km}}{\partial x_l} - \frac{\partial \Gamma^i{}_{kl}}{\partial x_m} \right) dx_l \delta x_m + (d\gamma^i{}_r \delta\gamma^r{}_k - d\gamma^r{}_k \delta\gamma^i{}_r) \right\} \xi^k.$$

Therefore, we obtain

$$\Delta\xi^i = \Delta R^i{}_k \xi^k,$$

where the $\Delta R^i{}_k$ are linear forms of the two displacements d and δ, or rather of the area element spanned by them, independent of the vector ξ and with the components

$$\Delta x_{lm} = dx_l \delta x_m - dx_m \delta x_l,$$

$$\Delta R^i{}_k = R^i{}_{klm} dx_l \delta x_m = \frac{1}{2} R^i{}_{klm} \Delta x_{lm} \qquad (R^i{}_{kml} = -R^i{}_{klm}), \tag{4}$$

$$R^i{}_{klm} = \left(\frac{\partial \Gamma^i{}_{km}}{\partial x_l} - \frac{\partial \Gamma^i{}_{kl}}{\partial x_m} \right) + (\Gamma^i{}_{lr} \Gamma^r{}_{km} - \Gamma^i{}_{mr} \Gamma^r{}_{kl}). \tag{5}$$

If η_i are the components of an arbitrary force at P, then $\eta_i \Delta\xi^i$ is | an invariant; consequently, $R^i{}_{klm}$ are the components of a 4th order tensor at P, covariant in klm and contravariant in i, the *curvature*. That the curvature vanishes identically is the necessary and sufficient condition for the manifold to be Euclidean. In addition to the condition of "skew" symmetry given beside (4), the curvature components satisfy the condition of "cyclic" symmetry: [395]

$$R^i{}_{klm} + R^i{}_{lmk} + R^i{}_{mkl} = 0.$$

By its nature, the curvature at a point P is a linear map or transformation $\Delta \boldsymbol{P}$, which assigns to each vector ξ there another vector $\Delta\xi$; this transformation itself depends linearly on an element of area at P:

$$\Delta \boldsymbol{P} = \boldsymbol{P}_{ik} dx_i \delta x_k = \frac{1}{2} \boldsymbol{P}_{ik} \Delta x_{ik} \qquad (\boldsymbol{P}_{ki} = -\boldsymbol{P}_{ik}).$$

Accordingly, the curvature is best described as a "linear transformation-tensor of 2nd order."

In order to counter objections to the proof of the invariance of the curvature tensor, which could be raised against the above considerations involving infinitesimals, one uses a force field f_i, and forms the change $d(f_i \xi^i)$ of the invariant product $f_i \xi^i$ in such a way that under the infinitely small displacement d the vector ξ is displaced parallel to itself. Replacing in the expression obtained the infinitesimal displacement

dx with an arbitrary vector ρ at P, one obtains an invariant bilinear form of two arbitrary vectors ξ and ρ at P. From this one forms the change which corresponds to a second infinitely small displacement δ, by parallely taking along the vectors ξ and ρ, and replacing thereafter the second displacement by a vector σ at P. One obtains the form

$$\delta d(f_i \xi^i) = \delta d f_i \cdot \xi^i + d f_i \delta \xi^i + \delta f_i d\xi^i + f_i \delta d\xi^i.$$

Through the interchange of d and δ and subsequent subtraction, this yields, because of the symmetry of $\delta d f_i$, the invariant

$$\Delta(f_i \xi^i) = f_i \Delta \xi^i,$$

and thus the desired proof has been completed.

4. METRIC MANIFOLD (THE AETHER)

4.1 The Concept of The Metric Manifold

A manifold *carries at the point P a metric*, if the line elements at P can be compared with respect to their lengths. For this purpose, we assume the validity of the
[396] Pythagorean-Euclidean | laws in the infinitely small. Hence, to any two vectors ξ, η at P shall correspond a number $\xi \cdot \eta$, the *scalar product*, which is a symmetric bilinear form with respect to the two vectors. This bilinear form is certainly not absolute, but is only determined up to an arbitrary non-zero factor of proportionality. Hence, it is actually not the form $\xi \cdot \eta$, that is given but only the equation $\xi \cdot \eta = 0$; two vectors which satisfy this equation are called *perpendicular* to one another. We presuppose that this equation is non-degenerate, i.e. that the only vector at P, to which all vectors at P can be perpendicular is the 0 vector. We do not however presuppose that the associated quadratic form $\xi \cdot \xi$ is positive definite. If it has the index of inertia q, and if $n - q = p$, then we say in brief, the manifold at the point considered is $(p + q)$ -dimensional. As a result of the arbitrary factor of proportionality, the two numbers p, q are only determined up to their order. We now assume that our manifold carries a metric [*Maßbestimmung*] at each point P. For the purpose of analytic representation, we consider (1) a definite coordinate system, and (2) the factor of proportionality appearing in the scalar product and which can be arbitrarily chosen at each point as fixed; with this, a "*frame of reference*"[9] for the analytic representation is obtained. If the vector ξ at the point P with the coordinates x_i has the components ξ^i, and η the components η^i, then one has

$$(\xi \cdot \eta) = \sum_{ik} g_{ik} \xi^i \eta^k \qquad (g_{ki} = g_{ik}),$$

9 I thus differentiate between "coordinate system" and "frame of reference."

where the coefficients g_{ik} are functions of the x_i. The g_{ik} should not only be continuous, but also be twice continuously differentiable. Since they are continuous and their determinant g by assumption does not vanish anywhere, the quadratic form $(\xi \cdot \xi)$ has the same index of inertia q at all points; therefore, we can describe the manifold in its entirety as $(p+q)$-dimensional. If we retain the coordinate system, but make a different choice for the undetermined factor of proportionality, then instead of the g_{ik} we obtain for the coefficients of the scalar product the quantities

$$g'_{ik} = \lambda \cdot g_{ik},$$

where λ is a nowhere vanishing continuous (and twice continuously differentiable) function of position.

According to the previous assumption, the manifold is only equipped with an *angle-measurement*; the geometry which is solely based on this, would be described as "*conformal geometry*"; it has, | as is well known, in the realm of two-dimensional [397] manifolds ("Riemannian surfaces") experienced extensive development, because of its importance for complex function theory. If we make no further assumptions, then the individual points of the manifold remain completely isolated from one another with respect to metrical properties. The manifold becomes endowed with a metric connection from point to point, only when a *principle exists for the transfer of the unit of length from a point* P *to an infinitely close one*. Instead, Riemann made the much farther reaching assumption, that line elements can be compared not only at the same location, but that they can be compared as to their lengths at two finitely distant locations. *But the possibility of such a "non-local geometric" comparison definitely cannot be admitted in a purely infinitesimal geometry.* Riemann's assumption has also entered the Einsteinian world geometry of gravitation. Here, this inconsequence shall be removed.

Let P be a fixed point and P_* an infinitely close point obtained from P through the displacement with the components dx_i. We assume a definite frame of reference. In relation to the unit of length thus defined at P (as well as at all other points in the space), the square of the length of an arbitrary vector ξ at P is given by

$$\sum_{ik} g_{ik}\xi^i\xi^k.$$

Now, *if we transfer the unit of length chosen at* P *to* P_*, which we presuppose as possible, the square of the length of an arbitrary vector ξ_* at P_* is given by

$$(1+d\varphi)\sum_{ik}(g_{ik}+dg_{ik})\xi_*^i\xi_*^k,$$

where $1+d\varphi$ is a factor of proportionality deviating infinitesimally from 1; $d\varphi$ must be a homogeneous function of degree 1 of the differentials dx_i. Namely, if we transplant the unit of length chosen at P from point to point along a curve leading from P to a finitely distant point Q, then on the basis of the unit of length so obtained at Q we obtain for the square of the length of an arbitrary vector at Q the expression $g_{ik}\xi^i\xi^k$, multiplied by the factor of proportionality which results from the product of the infinitely many individual factors of the form $1 + d\varphi$, which arise each time that we move from one point on the curve to the next.

$$\prod(1+d\varphi) = \prod e^{d\varphi} = e^{\Sigma d\varphi} = e^{\int_P^Q d\varphi}.$$

[398] | In order that the integral appearing in the exponent makes sense, $d\varphi$ must be a function of the differentials of the kind asserted.

If one replaces g_{ik} by $g'_{ik} = \lambda g_{ik}$, then in place of $d\varphi$ a different quantity $d\varphi'$ will appear. If λ denotes the value of this factor at the point P, one must have

$$(1+d\varphi')(g'_{ik}+dg_{ik}) = \lambda(1+d\varphi)(g_{ik}+dg_{ik}),$$

and this yields

$$d\varphi' = d\varphi - \frac{d\lambda}{\lambda}. \tag{6}$$

Of the initially possible assumptions about $d\varphi$, that it is a linear differential form, or the root of a quadratic one, or the cubic root of a cubic one etc., only the first, as we can now see from (6), has an invariant meaning. We have thus arrived at the following result.

The metric of a manifold is based on a quadratic and on a linear differential form

$$ds^2 = g_{ik}dx_i dx_k \qquad \textit{and} \qquad d\varphi = \varphi_i dx_i. \tag{7}$$

However, conversely these forms are not absolutely determined by the metric, but each pair of forms ds'^2 and $d\varphi'$, which arise from (7) according to the equations

$$ds'^2 = \lambda \cdot ds^2, \qquad d\varphi' = d\varphi - \frac{d\lambda}{\lambda} \tag{8}$$

is equivalent to the first pair in the sense that both express the same metric. In this λ is an arbitrary, nowhere vanishing continuous (more precisely: twice continuously differentiable) *function of position*. Into all quantities or relations which represent metric relations analytically, the functions g_{ik}, φ_i must thus enter in such a way that invariance holds (1) with respect to an arbitrary coordinate transformation ("coordinate-invariant"), and (2) with respect to the replacement of (7) by (8) ("measure-invariance").

$$\frac{d\lambda}{\lambda} = d\lg\lambda$$

is a total differential. Hence, whereas in the quadratic form ds^2, a factor of proportionality remains arbitrary at each location, the indeterminacy of $d\varphi$ consists of an additive total differential.

A metric manifold we describe physically as a world filled with *aether*. The particular metric existing in the manifold represents a particular state of the world filling aether. This state is thus to be described relative to a frame of reference through the specification (arithmetic construction) of the functions g_{ik}, φ_i. |

From (6) it follows that the linear tensor of 2nd order with the components [399]

$$F_{ik} = \frac{\partial\varphi_i}{\partial x_k} - \frac{\partial\varphi_k}{\partial x_i}$$

is uniquely determined by the metric of the manifold; I call it the *metric vortex*. It is the same, I believe, as what in physics one calls the *electromagnetic field*. It satisfies the "first system of Maxwell's equation"

$$\frac{\partial F_{kl}}{\partial x_i} + \frac{\partial F_{li}}{\partial x_k} + \frac{\partial F_{ik}}{\partial x_l} = 0.$$

Its vanishing is the necessary and sufficient condition for the transfer of length to be integrable, i.e., for those conditions which Riemann placed at the foundations of metric geometry to prevail. We understand from this how Einstein through his world geometry, which mathematically follows Riemann, could only account for gravitation but not for the electromagnetic phenomena.

4.2 Affine Connection of a Metric Manifold

In a metric space, in place of the requirement A imposed on the concept of parallel displacement in §3, I., we have the more specific one

A: that the parallel displacement of all vectors at a point P to an infinitely close point P', must not only be an affine but also a congruent transfer of the totality of these vectors.*

Using the previous notation, this requirement yields the equation

$$(1 + d\varphi)(g_{ik} + dg_{ik})(\xi^i + d\xi^i)(\xi^k + d\xi^k) = g_{ik}\xi^i\xi^k. \tag{9}$$

For all quantities a^i, which carry an upper index (i), we define the "lowering" of the index through the equations

$$a_i = \sum_k g_{ik}a^k.$$

(and the reverse process of raising an index through the inverse equations). Using this symbolism, for (9) we can write

$$(g_{ik}\xi^i\xi^k)d\varphi + \xi^i\xi^k dg_{ik} + 2\xi_i d\xi^i = 0.$$

The last term is

$$= -2\xi_i\xi^k d\gamma^i{}_k = -2\xi^i\xi^k d\gamma_{ik} = -\xi^i\xi^k(d\gamma_{ik} + d\gamma_{ki});$$

[400] | and therefore

$$d\gamma_{ik} + d\gamma_{ki} = dg_{ik} + g_{ik}d\varphi. \tag{10}$$

This equation can certainly be satisfied only if $d\varphi$ is a linear differential form; an assumption to which we were already driven above as the only reasonable one. From (10) or

$$\Gamma_{i,kr} + \Gamma_{k,ir} = \frac{\partial g_{ik}}{\partial x_r} + g_{ik}\varphi_r \tag{10*}$$

follows, as a consequence of the symmetry property $\Gamma_{r,ik} = \Gamma_{r,ki}$:

$$\Gamma_{r,ik} = \frac{1}{2}\left(\frac{\partial g_{ir}}{\partial x_k} + \frac{\partial g_{kr}}{\partial x_i} - \frac{\partial g_{ik}}{\partial x_r}\right) + \frac{1}{2}(g_{ir}\varphi_k + g_{kr}\varphi_i - g_{ik}\varphi_r); \quad (\Gamma_{r,ik} = g_{rs}\Gamma^s{}_{ik}). \tag{11}$$

It turns out that on a metric manifold the concept of the infinitesimal parallel displacement of a vector is uniquely determined through the requirements put forward.[10] I consider this as the *fundamental fact of infinitesimal geometry*, that with the metric also the affine connection of a manifold is given, that *the principle of transfer of length inherently carries with it that of transfer of direction*, or expressed physically, *that the state of the aether determines the gravitational field*.

If the quadratic form $g_{ik}dx_idx_k$ is indefinite, then among the geodesic lines, the null lines are distinguished as those along which the form vanishes. They depend only on the ratios of the g_{ik}, but not at all on the φ_i, they are thus structures of conformal geometry.[11]

We had imposed certain axiomatic requirements on the concept of parallel transport and shown that they can be satisfied on a metric manifold in one and only one way. However, it is also possible to define that concept explicitly in a simple manner. If P is a point in our metric manifold, then we call a frame of reference *geodesic* in P, if upon its use the φ_i vanish at P and the g_{ik} assume stationary values:

$$\varphi_i = 0, \qquad \frac{\partial g_{ik}}{\partial x_r} = 0.$$

[401] | D. *For each point P there exist geodesic frames of reference. If ξ is a given vector at P, and P' is an infinitely close point to P, then we understand by the vector which arises from x through parallel transport to P' that vector at P', which has the same components as ξ in the geodesic coordinate system belonging to P. This definition is independent of the choice of the geodesic frame of reference.*

---

10 See also Hessenberg, "Vektorielle Begründung der Differentialgeometrie," *Math. Ann.* vol. 78 (1917), p. 187–217, especially p. 208.

11 With this comment, I would like to correct a mistake on page 183 of my book *Raum, Zeit, Materie*.

It is not difficult to demonstrate the assertion contained in this explanation independently of the train of thought followed here through direct calculation, and to show by the same means that the process of parallel transport so defined is, in an arbitrary coordinate system, described by the equation

$$d\xi^r = -\Gamma^r{}_{ik}\xi^i dx_k \tag{12}$$

with the coefficients Γ to be taken from (11).[12] But here, where the invariant meaning of equation (12) is already established, we conclude more simply as follows. According to (11), the $\Gamma^r{}_{ik}$ vanish in a geodesic frame of reference and the equations (12) reduce to $d\xi^r = 0$. Hence, the concept of parallel transfer that we derived from the axiomatic requirements agrees with the one defined in D. Only the existence of a geodesic frame of reference is left to be shown. For this purpose, we choose a coordinate system x_i, geodesic at P, having the point P as its origin $(x_i = 0)$. If the unit of length at P and in its vicinity is for the time being chosen arbitrarily, and if furthermore the φ_i denote the value of these quantities at P, then one only needs to complete the transition from (7) to (8) with

$$\lambda = e^{\Sigma_i \varphi_i x_i},$$

in order to obtain that, besides the $\Gamma^i{}_{rs}$, the φ_i also vanish at P. From this then follows—see (10*)—the geodesic nature of the frame of reference so obtained. The coordinates of a frame of reference geodesic at P are in the immediate vicinity of P determined up to terms of 3rd order, leaving aside linear transformation, and the unit of length up to terms of 2nd order, leaving aside the addition of a constant factor.

12 In this one could follow the approach I have taken in RZM, §14.

ELIE CARTAN

THE DYNAMICS OF CONTINUOUS MEDIA AND THE NOTION OF AN AFFINE CONNECTION ON SPACE-TIME

Originally published as chapter 1 of "Sur les variétés à connexion affine et la théorie de la relativité généralisée" in Annales Scientifiques de L'Ecole Normale Supérieure (1923): 325–412. Translation, by Anne Magnon and Abhay Ashtekar, taken from "On Manifolds With An Affine Connection And The Theory Of General Relativity," (Napoli: Bibliopolis, 1986), p. 31–55.

PRINCIPLE OF INERTIA AND NEWTONIAN GRAVITY

1. Newton's foundation of classical mechanics rests on the concepts of absolute time and absolute space. Thus, analytically, any event can be labelled in time and space provided a choice is made of an origin, a unit of time, and a frame of spatial coordinates. For example, the frame might originate at the center of mass of the solar system and its axes might point towards fixed stars. Of course, any other frame which is invariantly related to this one would be also admissible. As is well known, the laws of mechanics remain unaffected if the frame of spatial coordinates is made to undergo a rectilinear, uniform translation with respect to Newton's absolute space, keeping the absolute time undisturbed. One is thus led to the notion of *Galilean frames of reference*.

The principle of inertia may be stated as follows: *in absence of interactions with other bodies, the velocity of a point mass remains constant in direction and magnitude in any Galilean frame*. The fact that the validity of this principle in one Galilean frame implies its validity in any other Galilean frame follows immediately from the transformation laws governing these frames. Let us label a point in space by arbitrary Cartesian coordinates,[1] not necessarily orthogonal. Then the transformation laws are as follows:

1 In the passage from orthogonal to general coordinates, the laws of classical mechanics retain their form and the formulas of *theoretical mechanics* remain unchanged.

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 4
Gravitation in the Twilight of Classical Physics: The Promise of Mathematics.

$$\left.\begin{aligned} x' &= a_1x + b_1y + c_1z + g_1t + h_1 \\ y' &= a_2x + b_2y + c_2z + g_2t + h_2 \\ z' &= a_3x + b_3y + c_3z + g_3t + h_3 \\ t' &= t + h, \end{aligned}\right\}$$

where the coefficients are constants. One can now give an alternate formulation of the
[32] principle of inertia. Two reference frames will be | said to be *equivalent* provided they are equivalent in the usual geometric sense and motionless relative to each other. Thus, the coordinate transformation between equivalent frames is given by:

$$\left.\begin{aligned} x' &= x + h_1, \\ y' &= y + h_2, \\ z' &= z + h_3, \\ t' &= t + h. \end{aligned}\right\}$$

Now, consider a moving point mass and attach to it, at each instant of time, a Galilean frame which has that point as its origin.[2] Then the principle of inertia can be stated as follows: *if a system of equivalent Galilean frames is attached to a moving point mass as above, then, at any instant of time, the velocity of the point mass in the Galilean frame corresponding to that instant is constant if the point mass is not subject to interaction with other bodies*.

2. Clearly, the structure of mechanics is based on two notions:

(i)The notion of a Galilean frame (which enables one to define the velocity of a moving point mass);

(ii)The notion of equivalent Galilean frames (which enables one to state the principle of inertia).

It is important to note the advantage of the second formulation of the principle of inertia: in essence, *it uses the notion of equivalent Galilean frames only for those frames whose origins are infinitesimally close*. All generalizations of classical or relativistic mechanics retain the notion of Galilean frames; it is the notion of equivalent frames that has undergone modifications.

Let us continue to use the framework of classical mechanics with the notion of absolute time (as measured by a unit which is fixed once and for all). We shall see that a modification of the notion of equivalent frames will enable us to extend the principle of inertia so that it incorporates not only isolated point masses but also *point*

2 That is, the point is the origin of the coordinate axes, and the instant, at which one examines it, is taken to be the origin of time.

masses placed in a gravitational field. Let us fix a Galilean frame and denote by T_0 the corresponding spatial triad of coordinates. Next, let us introduce a field of forces analogous to a gravitational field, i.e., an acceleration field (X, Y, Z). Then, if the velocity of a point mass w.r.t. the fixed Galilean frame is given by

$$u, \quad v, \quad w$$

at time t, at time $t + dt$ the velocity will be

$$u + Xdt, \quad v + Ydt, \quad w + Zdt.$$

| At time t, let us attach to the point mass a triad T which is equivalent to T_0 in the [33]
usual geometrical sense. Similarly, at time $t + dt$, let us attach a triad T'. These triads will define Galilean frames only if one specifies that they are in a rectilinear, uniform motion w.r.t. T_0. When this is done, the resulting Galilean frames will have origins

$$x, \quad y, \quad z, \quad t \text{ and } x + dx, \quad y + dy, \quad z + dz, \quad t + dt$$

respectively. Denote by

$$a, \quad b, \quad c, \text{ and } a', \quad b', \quad c'$$

the translation velocities of T and T' w.r.t. T_0. Then the velocity of the point mass in the Galilean frame attached to it at time t has components

$$u - a, \quad v - b, \quad w - c,$$

and in the Galilean frame attached at time $t + dt$,

$$u + Xdt - a', \quad v + Ydt - b', \quad w + Zdt - c'.$$

Thus the components will not have changed if

$$a' - a = Xdt, \quad b' - b = Ydt, \text{ and } \quad c' - c = Zdt.$$

Consequently, *the motion of an arbitrary point mass which is placed in the field of forces described above will satisfy the principle of inertia provided two Galilean frames with infinitesimally close origins,*

$$x, \quad y, \quad z, \quad t;$$

$$x + dx, \quad y + dy, \quad z + dz, \quad t + dt,$$

are considered as equivalent if their triads T and T' are equivalent in the usual geometrical sense, and if T is in a rectilinear, uniform translational motion w.r.t. T' with velocity (Xdt, Ydt, Zdt). Note that, again, we have used mutual relations only between the infinitesimally close frames of reference.

3. One can express the same ideas in a way which is perhaps more intuitive, and which has the advantage of being closer to the viewpoint adopted by Einstein than the [34] point of departure for his theory of | gravitation. Consider a point particle moving in the field of forces discussed above and attach to it a spatial triad T originating at the point and carried by it in a *translational* motion. At each instant of time, introduce a Galilean reference frame consisting of a triad $\bar{T}$ which coincides with T at the instant considered and which is in a *rectilinear*, *uniform* translational motion, with the velocity which the particle has at that precise instant.[3] Obviously, the velocity of the particle w.r.t. these frames is zero. Thus, the motion of the point mass agrees with the principle of inertia (constancy of velocity) if the successive Galilean reference frames defined by the triads $\bar{T}$ are considered as equivalent, *step by step*. Clearly, the constant velocity of the triad $\bar{T}'$, corresponding to time $t + dt$, w.r.t. the triad $\bar{T}$, corresponding to time t, is given by

$$Xdt, \quad Ydt, \quad Zdt.$$

Consider for example the uniform field due to earth's gravity and assume for a moment that one can neglect the motion of the earth w.r.t. the absolute space. Choose the z-axis along the vertical upwards direction. Two triads, T and T', associated with instants of time t and $t + dt$, will be said to define two equivalent Galilean frames of reference if T' has a constant vertical velocity gdt w.r.t. T. Thus, one can attach a Galilean frame to each event (x, y, z, t) of space-time in such a way that *all the resulting frames are equivalent*; given the frame associated with a particular event (x_0, y_0, z_0, t_0), all others will be completely determined. In a general case, however, such a situation does not occur: Whether two frames are equivalent or not will depend on the space-time paths connecting their origins, since the notion of equivalence itself has been introduced via a step by step procedure. We shall return to this fundamental issue later on.

4. Let us say that the conditions determining the equivalence of two Galilean frames with infinitesimal close origins define the *geometrical*[4] properties of the space-time. Thus, the gravitational phenomena are shifted from the domain of physics to that of geometry[5] and the components X, Y and Z of the gravitational field capture the basic geometrical structure of space-time.[6] The relations

$$\frac{\partial Z}{\partial y} - \frac{\partial Y}{\partial z} = 0, \qquad \frac{\partial X}{\partial z} - \frac{\partial Z}{\partial x} = 0, \qquad \frac{\partial Y}{\partial x} - \frac{\partial X}{\partial y} = 0 \tag{1}$$

---

3 Actually, for the purposes for which the triad $\bar{T}$ has been used, one could have replaced it by T itself; thus, as far as the velocity of a point at instant t is concerned, T plays the role of a Galilean triad.

4 Actually, these properties are geometrical as well as kinematical.

5 This is essentially another way of stating the equality of inertial and gravitational masses, or, the fact that the gravitational field is kinematical (a field of accelerations) rather than dynamical (a field of forces).

which hold in *orthogonal* coordinates express the properties of this | structure. [35]
Finally, the fundamental Poisson equation,

$$\frac{\partial X}{\partial x}+\frac{\partial Y}{\partial y}+\frac{\partial Z}{\partial z} = -4\pi\rho, \tag{2}$$

which, together with the above relations, yields a complete formulation of the laws of Newtonian gravitation,[7] shows that *the matter density of a continuous medium is the physical manifestation of a local geometrical property of space-time*. Thus, we recover some features of Einstein's theory of gravity within the framework of classical mechanics itself. The only essential difference is the lack of relation between gravitational and electromagnetic phenomena. But we have already recovered the structure which intertwines space-time, geometry and matter.

5. All these considerations call for further remarks. Does the reduction of gravitation to geometry occur only for a specific definition of the equivalence of two infinitesimally close Galilean frames? We shall examine this question in detail later. For the time being, let me just say that the answer is in the negative. Let us consider two Galilean frames with infinitesimally close origins which are equivalent in the sense of section 3 above. Thus, the corresponding triads T and T', with origins M and M' are parallel, T' undergoing a uniform rectilinear translation w.r.t. (T). Let us now replace T' by T'', a triad which is fixed w.r.t. T', has M' as its origin, and is obtained from (T) by a helicoidal displacement along the axis MM', the sense and the magnitude of the displacement being fixed once and for all. Consider a point mass which if freely falling in a gravitational field such that it finds itself at M at time t and at M' at time $t + dt$. Since the velocity of this point mass is almost colinear with MM' at time $t + dt$, it will have the same components w.r.t. the frame S' defined by T' as those w.r.t. the frame S'' defined by T''. Thus, if the motion of the point mass obeys the principle of inertia when S and S' are equivalent, *it will continue to obey this principle with the modified definition of equivalence*. This example leads us to the following conclusion: *As far as the dynamics of a point mass is concerned, there exists an infinite number of definitions of equivalence of Galilean frames whose origins are infinitesimally close.*

6. One might expect that these conclusions would have to be modified for the dynamics of material *systems* since the dynamics of a point mass neglects the important issue of rotation. Consider a small spherical ball undergoing an absolute, uniform

---

6 In fact, as we shall see later, this structure requires the functions X, Y, Z only to be defined up to arbitrary additive constants. This is because mechanical experiments performed inside a system which is embedded in a *uniform* gravitational field cannot detect this field. In particular, if one assumes that the gravitational field due to distant stars is *uniform over the solar system,* the laws of celestial Mechanics governing the motion of sun and its planets remain unchanged.

7 In addition, we must assume that the functions X, Y, Z vanish at infinity.

rotation. The axis of rotation, along which its angular momentum points, should be |
[36] considered as remaining equivalent to itself. Thus, our first convention by which spatial directions remain parallel to themselves in the usual sense, appears to be the only one that is permissible. However, it is simply too early to draw such a conclusion. In fact, we shall see later on that this conclusion is premature and *the high degree of indeterminacy in the notion of equivalence persists in its entirety when one deals with the laws of dynamics of material systems*.[8] However, to investigate this issue in a fruitful way, it is important to note that the new viewpoint which we have now adopted requires that the laws of mechanics should be formulated only *locally*. In other words, we must go back to mechanics of continuous media. Indeed, we do not have the notion of equivalence of two frames unless their origins are infinitesimally close.

In order to facilitate the transition from Newtonian to relativistic mechanics, we shall now formulate the equations of mechanics of continuous media using a 4-dimensional manifold as the model for space-time.

FOUR DIMENSIONAL SPACE-TIME AND CLASSICAL DYNAMICS OF CONTINUOUS MEDIA

7. Let us adopt the viewpoint of classical mechanics. Space-time or the universe will be represented by an *affine manifold*. By this, we mean the following. Let us call a *space-time vector* a set consisting of two events (each located in time and space) one of which is the origin of the vector and, the other, the extremity. In a Galilean frame the components of a space-time vector are the four numbers

$$t'-t, \quad x'-x, \quad y'-y, \quad z'-z,$$

obtained by subtracting the coordinates of the origin from those of the extremity. If the components of two space-time vectors are identical in one Galilean frame, they are identical in all Galilean frames. Thus, here we have a property of space-time vectors which is independent of the Galilean frame used to represent these vectors mathematically. Vectors which have this property will be said to be *equivalent*. It is clear that if two space-time vectors are equivalent to a third, they are equivalent to each other. It is the existence of this notion of equivalence among space-time vectors that we express when we say that space-time has an affine structure.

Of the four numbers,

$$t'-t, \quad x'-x, \quad y'-y, \quad z'-z$$

[37] | which mathematically represent a space-time vector, the first will be referred to as the *time component* and the remaining three will be called *space components*. Note that *the time component is independent of the choice of reference frame*. The situation is different for the spatial vector whose components in the coordinate triad T defin-

8 Except for one possible restriction; see section no. 16.

ing the given Galilean frame are $x' - x, y' - y$ and $z' - z$; the spatial vector depends not only on the given space-time vector and the triad T but also on the velocity of this triad w.r.t. the absolute space.

Let us now consider a point mass moving w.r.t. a Galilean frame. In this frame, the components of the space-time vector joining the position of this point mass at time t to that at time $t + dt$ are:

$$dt, \quad dx, \quad dy, \quad dz.$$

The vector itself does not depend on the Galilean frame. The same remarks hold for the space-time vector

$$1, \ \frac{dx}{dt}, \ \frac{dy}{dt}, \ \frac{dz}{dt}$$

which is obtained by dividing the first space-time vector by dt. Finally, if we denote the mass of the particle by m, the space-time vector

$$m, \ m\frac{dx}{dt}, \ m\frac{dy}{dt}, \ m\frac{dz}{dt}$$

is itself again independent of one's choice of the frame of reference. This is the *energy-momentum vector*. While its time component, the mass, is independent of the choice of the frame of reference, its space-component, the momentum, does depend on this choice.

We can now state the fundamental principle of particle dynamics: *The time-derivative of the energy-momentum space-time vector is equal to the spatial force vector.* This statement contains both the principle of conservation of mass and the law relating force and acceleration.

8. Let us now consider a continuous medium equipped with a given Galilean frame. Fix a 3-dimensional volume of space-time. As I have shown elsewhere,[9] the total mass contained in this volume is given by the integral

$$\iiint \rho dxdydz - \rho udydzdt - \rho vdzdxdt - \rho wdxdydt$$

where ρ denotes the density and u, v, w denote the components of | the velocity of [38]
each element of matter. Let us first assume that the matter is *free of pressure as well as stress*. Then the x component of momentum of the same volume is given by the integral

$$\iiint \rho udxdydz - \rho u^2 dydzdt - \rho uvdzdxdt - \rho uwdxdydt.$$

The y and the z components can be expressed similarly. Let us denote by

9 E. Cartan, *Leçons sur les Invariants intégraux,* Paris, Hermann, 1922, p. 35-37. [Translator's note: The integral is just $\iiint p^a_{\varepsilon abcd} dx^b \wedge dx^c \wedge dx^d$, where $p^a \equiv (\rho, \rho u, \rho v, \rho w)$ is the 4-momentum density.]

$$\Pi,\quad \Pi_x,\quad \Pi_y,\quad \Pi_z$$

the integrands of these integrals. These are the four components of the energy-momentum vector of an element of matter in the medium. Finally, let us denote by $X, Y, Z,$ the components of the force per unit volume.

To obtain the equations of mechanics of continuous media, we proceed as follows. Consider a 4-dimensional domain of space-time and decompose it into *world tubes* formed by elements of the matter under consideration, taken between time t_1 and t_2. Thus the boundary of this domain consists of elements of matter at the extremities t_1 and t_2 of the time interval. Now, the geometrical difference between the energy-momentum vectors of a matter element evaluated at time t_1, when the element enters the domain, and at time t_2, when it leaves, is given by a spatial vector with components

$$\int_{t_1}^{t_2}(Xdxdydz)dt, \qquad \int_{t_1}^{t_2}(Ydxdydz)dt, \qquad \int_{t_1}^{t_2}(Zdxdydz)dt.$$

In other words, *the integral of the "energy-momentum 4-vector" over the boundary of the domain is equal to the four dimensional integral of the "force 3-vector" over the domain itself.* This statement finds its expression in the following formulas:

$$\left.\begin{aligned} \Pi' &= 0 \\ \Pi'_x &= Xdtdxdydz \\ \Pi'_y &= Ydtdxdydz \\ \Pi'_z &= Zdtdxdydz, \end{aligned}\right\} \tag{5}$$

where

$$\Pi' = \left[\frac{\partial\rho}{\partial t} + \frac{\partial(\rho u)}{\partial x} + \frac{\partial(\rho v)}{\partial y} + \frac{\partial(\rho w)}{\partial z}\right]dtdxdydz$$

$$\Pi'_x = \left[\frac{\partial(\rho u)}{\partial t} + \frac{\partial(\rho u^2)}{\partial x} + \frac{\partial(\rho uv)}{\partial y} + \frac{\partial(\rho uw)}{\partial z}\right]dtdxdydz$$

$$\Pi'_y = \left[\frac{\partial(\rho v)}{\partial t} + \frac{\partial(\rho uv)}{\partial x} + \frac{\partial(\rho v^2)}{\partial y} + \frac{\partial(\rho vw)}{\partial z}\right]dtdxdydz$$

$$\Pi'_z = \left[\frac{\partial(\rho w)}{\partial t} + \frac{\partial(\rho wu)}{\partial x} + \frac{\partial(\rho wv)}{\partial y} + \frac{\partial(\rho w^2)}{\partial z}\right]dtdxdydz.$$

[39] | After some calculations and simplifications these equations yield the familiar ones:

$$\frac{\partial\rho}{\partial t} + \frac{\partial(\rho u)}{\partial x} + \frac{\partial(\rho v)}{\partial y} + \frac{\partial(\rho w)}{\partial z} = 0$$

$$\rho\left(\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} + w\frac{\partial u}{\partial z}\right) = X$$

$$\rho\left(\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + w\frac{\partial v}{\partial z}\right) = Y$$

$$\rho\left(\frac{\partial w}{\partial t} + u\frac{\partial w}{\partial x} + v\frac{\partial w}{\partial y} + w\frac{\partial w}{\partial z}\right) = Z.$$

Let us call *exterior derivative*[10] the operation which enables one to convert an integral over a closed p-dimensional manifold to the integral over the $(p+1)$-dimensional manifold enclosed by the p-manifold. Then the fundamental principle of mechanics of continuous media can be stated as follows: *The exterior derivative of the energy-momentum field is equal to the product of dt with the force field.*

9. In the above discussion, we had assumed the absence of pressure as well as stress. However, the general case can be reduced to the one discussed above by defining the (generalized) momentum of an element of matter to be the vector whose components are obtained by adding the following quantities to the components introduced previously:

$$\begin{aligned}
&-p_{xx}dydzdt - p_{xy}dzdxdt - p_{xz}dxdydt\\
&-p_{yx}dydzdt - p_{yy}dzdxdt - p_{yz}dxdydt\\
&-p_{zx}dydzdt - p_{zy}dzdxdt - p_{zz}dxdydt.
\end{aligned}$$

In the kinetic theory of gases, one can in effect consider pressure to be the flux of momentum resulting from irregularities in the molecular velocities. On the other hand, the quantities u, v, w introduced previously represent only an *average* velocity.

The usual equations of mechanics of continuous media can be | now recovered by [40]
expanding equations (5):

$$\frac{\partial\rho}{\partial t} + \frac{\partial(\rho u)}{\partial x} + \frac{\partial(\rho v)}{\partial y} + \frac{\partial(\rho w)}{\partial z} = 0,$$

$$\rho\left(\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} + w\frac{\partial u}{\partial z}\right) + \frac{\partial p_{xx}}{\partial x} + \frac{\partial p_{xy}}{\partial y} + \frac{\partial p_{xz}}{\partial z} = X$$

$$\rho\left(\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + w\frac{\partial v}{\partial z}\right) + \frac{\partial p_{yx}}{\partial x} + \frac{\partial p_{yy}}{\partial y} + \frac{\partial p_{yz}}{\partial z} = Y$$

$$\rho\left(\frac{\partial w}{\partial t} + u\frac{\partial w}{\partial x} + v\frac{\partial w}{\partial y} + w\frac{\partial w}{\partial z}\right) + \frac{\partial p_{zx}}{\partial x} + \frac{\partial p_{zy}}{\partial y} + \frac{\partial p_{zz}}{\partial z} = Z.$$

10 See E. Cartan, *Leçons sur les Invariants intégraux*, Chapter VII, p. 65.

10. However, these equations are not complete. In effect, this is because we have not taken into account the theorem[11] of angular momentum which may be expressed in the present framework as follows:

$$\iiint y\Pi_z - z\Pi_y = \iiiint (yZ - zY)\,dtdxdydz$$
$$\iiint z\Pi_x - x\Pi_z = \iiiint (zX - xZ)\,dtdxdydz$$
$$\iiint x\Pi_y - y\Pi_x = \iiiint (xY - yX)\,dtdxdydz,$$

where the integrals on the right-hand side are taken over an arbitrary volume element of space-time and those on the left hand side, on the 3-dimensional boundary of this volume. These equations yield:

$$[dy\Pi_z] - [dz\Pi_y] = 0$$
$$[dz\Pi_x] - [dx\Pi_z] = 0$$
$$[dx\Pi_y] - [dy\Pi_x] = 0,$$

and they are satisfied trivially in absence of pressure. In the general case, they imply:

$$p_{zy} - p_{yz} = 0$$
$$p_{xz} - p_{zx} = 0$$
$$p_{yx} - p_{xy} = 0.$$

11. One can express the previous results using a simple vectorial notation. Let e_0, e_1, e_2, e_3 denote the 4-vectors whose components are, respectively, $(1, 0, 0, 0)$, $(0, 1, 0, 0)$, $(0, 0, 1, 0)$, and $(0, 0, 0, 1)$. The last three are spatial
[41] vectors. In this notation, the energy-momentum I of a particle of mass m is given by

$$m\left(\mathrm{e}_0 + \frac{dx}{dt}\mathrm{e}_1 + \frac{dy}{dt}\mathrm{e}_2 + \frac{dz}{dt}\mathrm{e}_3\right).$$

Let us now denote by m the space-time point (t, x, y, z). Then the derivative $d\mathrm{m}/dt$ of this point w.r.t. time is a space-time vector with components

$$\left(1,\ \frac{dx}{dt},\ \frac{dy}{dt},\ \frac{dz}{dt}\right).$$

Thus, the energy-momentum of the particle is given by

11 Note that the analytical formulation of this theorem does not require the restriction to rectangular axes.

$$m\frac{d\mathrm{m}}{dt}.$$

The points and the (free) vectors are “geometric forms” of order one. One can also consider geometric forms of second order which represent systems of *sliding vectors*. We shall denote by $[\mathrm{mm}']$ the sliding vector whose origin lies at the space-time point m and whose extremity is at the space-time point m'. This sliding vector has ten plückerian coordinates which are the 2×2 determinants constructed from the tableau

$$\begin{matrix} 1, & t, & x, & y, & z \\ 1, & t', & x', & y', & z'; \end{matrix}$$

clearly,

$$[\mathrm{mm}'] = -[\mathrm{m}'\mathrm{m}].$$

Similarly, we shall denote by $[\mathrm{me}]$ the sliding vector obtained from the vector which originates at the space-time point m and which is parallel to a given vector e. The plückerian coordinates of this vector are obtained from the tableau

$$\begin{matrix} 1, & t, & x, & y, & z, \\ 0, & \theta, & \xi, & \eta, & \zeta, \end{matrix}$$

where the second line contains the components of the vector e. Finally, | let us denote [42]
by $[\mathrm{ee}']$ the bivector whose ten coordinates are obtained from the tableau

$$\begin{matrix} 0, & \theta, & \xi, & \eta, & \zeta, \\ 0, & \theta', & \xi', & \eta', & \zeta', \end{matrix}$$

of the components of the two free vectors e and e'.

In each of these cases, the sliding vector or the bivector under consideration may be viewed as the (exterior) product of the two factors each of which is a first order form (a point or a free vector). The product is distributive and antisymmetric.

12. The sliding vector whose origin lies at the space-time point m representing the position of a point particle at a given instant of time and which carries the energy-momentum of the particle can be expressed as

$$m\left[\mathrm{m}\frac{d\mathrm{m}}{dt}\right].$$

Therefore, the equation

$$\frac{d}{dt}\left\{m\left[\mathrm{m}\frac{d\mathrm{m}}{dt}\right]\right\} = [\mathrm{F}],$$

where $[\mathrm{F}]$ is the "force" sliding vector, contains, at once, the fundamental principle of dynamics and the theorem of angular momentum. Indeed, it contains the ten equations

$$\frac{dm}{dt} = 0, \qquad \frac{d}{dt}\left(m\frac{dx}{dt}\right) = X,$$

$$\frac{d}{dt}\left(m\frac{dy}{dt}\right) = Y, \qquad \frac{d}{dt}\left(m\frac{dz}{dt}\right) = Z,$$

$$\frac{d}{dt}\left(mt\frac{dx}{dt} - mx\right) = tX, \qquad \frac{d}{dt}\left(mt\frac{dy}{dt} - my\right) = tY,$$

$$\frac{d}{dt}\left(mt\frac{dz}{dt} - mz\right) = tZ, \qquad \frac{d}{dt}\left(my\frac{dz}{dt} - mz\frac{dz}{dt}\right) = yZ - zY,$$

$$\frac{d}{dt}\left(mz\frac{dx}{dt} - mx\frac{dz}{dt}\right) = zX - xZ, \frac{d}{dt}\left(mx\frac{dy}{dt} - my\frac{dx}{dt}\right) = xY - yX.$$

[43] | 13. Let us now return to the mechanics of continuous media. Denote the energy-momentum by a sliding vector G and the force per unit element of a 3-dimensional volume by a sliding vector F. *Then the equations of mechanics are succintly captured in the single formula*

$$\mathrm{G}' = [dt\mathrm{F}]. \tag{6}$$

Note that

$$\mathrm{G} = [m\mathrm{e}_0]\Pi + [m\mathrm{e}_1]\Pi_x + [m\mathrm{e}_2]\Pi_y + [m\mathrm{e}_3]\Pi_z,$$

and

$$\mathrm{F} = [m\mathrm{e}_1]X dxdydz + [m\mathrm{e}_2]Y dxdydz + [m\mathrm{e}_3]Z dxdydz.$$

The equation

$$d\mathrm{m} = \mathrm{e}_0 dt + \mathrm{e}_1 dx + \mathrm{e}_2 dy + \mathrm{e}_3 dz$$

yields

$$\begin{aligned}\mathrm{G}' = {} & [m\mathrm{e}_0]\Pi' + [m\mathrm{e}_1]\Pi'_x + [m\mathrm{e}_2]\Pi'_y + [m\mathrm{e}_3]\Pi'_z \\ & + [\mathrm{e}_0\mathrm{e}_1][dt\Pi_x - dx\Pi] + [\mathrm{e}_0\mathrm{e}_2][dt\Pi_y - dy\Pi] + [\mathrm{e}_0\mathrm{e}_3][dt\Pi_z - dz\Pi] \\ & + [\mathrm{e}_2\mathrm{e}_3][dy\Pi_z - dz\Pi_y] + [\mathrm{e}_3\mathrm{e}_1][dz\Pi_x - dx\Pi_z] + [\mathrm{e}_1\mathrm{e}_2][dx\Pi_y - dy\Pi_x].\end{aligned}$$

It is easy to verify that the coefficients of $[\mathrm{e}_0\mathrm{e}_1]$, $[\mathrm{e}_0\mathrm{e}_2]$ and $[\mathrm{e}_0\mathrm{e}_3]$ vanish identically. If the elements of matter are subject *to a torque*, *in addition to the force*, one has simply to add terms of the following form to the expression of the force:

$$[\mathrm{e}_2\mathrm{e}_3]\mathrm{L}dxdydz + [\mathrm{e}_3\mathrm{e}_1]\mathrm{M}dxdydz + [\mathrm{e}_1\mathrm{e}_2]\mathrm{N}dxdydz.$$

This implies

$$p_{zy} - p_{yz} = \mathrm{L}, \qquad p_{xz} - p_{zx} = \mathrm{M}, \qquad p_{yx} - p_{xy} = \mathrm{N}.$$

Then the fundamental equation (6) continues to hold. Clearly, the basic equation of dynamics can be recovered from equation (6) under the assumption that the matter is contained in a very small spatial volume: in this approximation, one obtains

$$d\mathrm{G} = dt\mathrm{F}.$$

| 14. Equation (6) will enable us to obtain easily the equations of mechanics of continuous media. Let us attach a variable Galilean frame to each point of space-time. Denote by $\mathrm{e}_0, \mathrm{e}_1, \mathrm{e}_2, \mathrm{e}_3$ the free space-time vectors which define the Galilean frame attached to the point m. As we move from a point m to an infinitesimally nearby point m', these vectors will change. However, the time component of e_0 will be always equal to 1, and those of $\mathrm{e}_1, \mathrm{e}_2, \mathrm{e}_3$ will always vanish. One will therefore have the following formulae:[12] [44]

$$\begin{aligned} d\mathrm{e}_0 &= \omega_0^1\mathrm{e}_1 + \omega_0^2\mathrm{e}_2 + \omega_0^3\mathrm{e}_3 \\ d\mathrm{e}_1 &= \omega_1^1\mathrm{e}_1 + \omega_1^2\mathrm{e}_2 + \omega_1^3\mathrm{e}_3 \\ d\mathrm{e}_2 &= \omega_2^1\mathrm{e}_1 + \omega_2^2\mathrm{e}_2 + \omega_2^3\mathrm{e}_3 \\ d\mathrm{e}_3 &= \omega_3^1\mathrm{e}_1 + \omega_3^2\mathrm{e}_2 + \omega_3^3\mathrm{e}_3 \end{aligned} \tag{7}$$

where ω_i^j are linear combinations of the differentials of the four functions which label space-time points. Let us denote by

$$d\mathrm{m} = \omega^0\mathrm{e}_0 + \omega^1\mathrm{e}_1 + \omega^2\mathrm{e}_2 + \omega^3\mathrm{e}_3 \tag{8}$$

the free vector joining m and m'. Thus ω^0 is simply the infinitesimal time interval between m and m'. Now, one can again obtain the following expressions:

$$\mathrm{G} = [\mathrm{me}_0]\Pi + [\mathrm{me}_1]\Pi_x + [\mathrm{me}_2]\Pi_y + [\mathrm{me}_3]\Pi_z,$$

and

$$\mathrm{F} = [\mathrm{me}_1]X\omega^1\omega^2\omega^3 + [\mathrm{me}_2]Y\omega^1\omega^2\omega^3 + [\mathrm{me}_3]Z\omega^1\omega^2\omega^3.$$

It is only the expression of G' that becomes more complicated because the free vectors $\mathrm{e}_0, \mathrm{e}_1, \mathrm{e}_2, \mathrm{e}_3$ are no longer fixed. One has:

12 As in section no. 1, the axes of coordinate frames are not necessarily orthogonal here.

$$\begin{aligned} G' = {} & [me_0]\Pi' + [me_1][\Pi'_x + \omega_0^1\Pi + \omega_1^1\Pi_x + \omega_2^1\Pi_y + \omega_3^1\Pi_z] \\ & + [me_2][\Pi'_y + \omega_0^2\Pi + \omega_1^2\Pi_x + \omega_2^2\Pi_y + \omega_3^2\Pi_z] \\ & + [me_3][\Pi'_z + \omega_0^3\Pi + \omega_1^3\Pi_x + \omega_2^3\Pi_y + \omega_3^3\Pi_z] \\ & + [e_0e_1][\omega^0\Pi_x - \omega^1\Pi\ ] + [e_0e_2][\omega^0\Pi_y - \omega^2\Pi\ ] + [e_0e_3][\omega^0\Pi_z - \omega^3\Pi\ ] \\ & + [e_2e_3][\omega^2\Pi_z - \omega^3\Pi_y] + [e_3e_1][\omega^3\Pi_x - \omega^1\Pi_z] + [e_1e_2][\omega^1\Pi_y - \omega^2\Pi_x]. \end{aligned}$$

The required equations now follow immediately. ∎

[45] THE AFFINE CONNECTION OF SPACE-TIME AND CLASSICAL MECHANICS

15. Up to this point we have dealt with the usual notion of equality of space-time vectors. However, the formulae obtained above would continue to be valid for an arbitrary definition of equality of two space-time vectors *whose origins are infinitesimally close*: if a more general definition is used, equations (7) preserve their form *but with modified coefficients*.[13]

Let us assume, as is indeed the case in applications, that the only volume force present is the one due to gravity. If the definition of equivalence of two nearby Galilean frames—or, equivalently, the definition of equality of two 4-vectors whose origins are infinitesimally close—is so chosen as to cancel the gravitational forces, the equations of dynamics would reduce to $G' = 0$. Fix a Galilean frame and choose for e_0, e_1, e_2 and e_3 vectors which remain equal, in the usual sense, to the unit vectors of this frame. Set

$$\omega_0^1 = -X dt, \qquad \omega_0^2 = -Y dt, \qquad \omega_0^3 = -Z dt,$$

and

$$\omega_i^j = 0 \qquad (i, j = 1, 2, 3).$$

Then, the equations of mechanics become

$$\Pi' = 0, \quad \Pi'_x - X[dt\Pi] = 0, \quad \Pi'_y - Y[dt\Pi] = 0, \quad \Pi'_z - Z[dt\Pi] = 0$$

or, equivalently,

13 Note that, if we had attached a *different* Galilean frame, say,

$$\bar{e}_0 = e_0 + u e_1, \ \bar{e}_1 = e_1, \qquad \bar{e}_2 = e_2, \qquad \bar{e}_3 = e_3,$$

at each world point, the formula (7) would have to be modified. In particular, ω_0^1 would have to be replaced by $\bar{\omega}_0^1 = \omega_0^1 + u\omega_1^1 + du$.

$$\begin{aligned}\Pi' &= 0,\\ \Pi'_x &= \rho X[dtdxdydz],\\ \Pi'_y &= \rho Y[dtdxdydz],\\ \Pi'_z &= \rho Z[dtdxdydz].\end{aligned}$$

These are the equations of classical dynamics of a continuous medium subject to a volume force which is proportional to the mass. To geometrize gravity, it suffices to choose X, Y, Z to be the components of the acceleration due to gravity.

The result just obtained is completely analogous to the one which led directly to the dynamics of a point particle. Indeed, the formulas

$$d\mathrm{e}_0 = -Xdt\mathrm{e}_1 - Ydt\mathrm{e}_2 - Zdt\mathrm{e}_3, \qquad d\mathrm{e}_1 = d\mathrm{e}_2 = d\mathrm{e}_3 = 0$$

imply that two Galilean frames originating at t, x, y, z and $t+dt$, | $x+dx$, $y+dy$, [46]
$z+dz$ should be considered as equivalent if the corresponding triads T and T' are equivalent in the usual sense and T' undergoes a rectilinear and uniform translation of velocity (Xdt, Ydt, Zdt) w.r.t. T.

16. Let us adopt the convention that a given definition of the equivalence of Galilean frames with infinitesimally close origins gives rise to a space-time *affine connection*. It is now easy to see that the gravitational phenomena are compatible with several distinct affine connections on space-time. It is important to note first that, *although the affine connection depends on the matter distribution in space, it does not undergo a substantial change on introduction of a small mass in a given region of space-time*. If the entire system consists only of a small mass, the corresponding affine connection will not depend upon the state of this mass. Any possible modification in the affine connection has the effect that the following terms are added to the expression of G':

$$\left.\begin{aligned} &[m\mathrm{e}_1][\bar\omega_0^1\Pi + \bar\omega_1^1\Pi_x + \bar\omega_2^1\Pi_y + \bar\omega_3^1\Pi_z]\\ &+[m\mathrm{e}_2][\bar\omega_0^2\Pi + \bar\omega_1^2\Pi_x + \bar\omega_2^2\Pi_y + \bar\omega_3^2\Pi_z]\\ &+[m\mathrm{e}_3][\bar\omega_0^3\Pi + \bar\omega_1^3\Pi_x + \bar\omega_2^3\Pi_y + \bar\omega_3^3\Pi_z],\end{aligned}\right\} \tag{9}$$

where $\bar\omega_0^i, \bar\omega_i^j$ are the changes in the components ω_0^i, ω_i^j of the affine connection. Thus, the only possible modifications are the ones which make the three terms in parentheses vanish *irrespective of the numerical values of the quantities which characterize the state of the material medium*.

Let us first consider the most general situation in which the affine connection permits two Galilean frames, one with an orthogonal triad T and the other with a *non-orthogonal triad* T', to be equivalent. Let us assume—and it is permissible—that the vectors e_0, e_1, e_2, e_3 used in the equation (9) are still equal in the *usual sense,* i.e., when everything is referred back to a fixed Galilean frame. Set

$$\begin{aligned}\bar{\omega}_0^i &= \gamma_{00}^i dt + \gamma_{01}^i dx + \gamma_{02}^i dy + \gamma_{03}^i dz,\\ \bar{\omega}_i^j &= \gamma_{i0}^j dt + \gamma_{i1}^j dx + \gamma_{i2}^j dy + \gamma_{i3}^j dz.\end{aligned}$$

The coefficients of the forms Π, Π_x, Π_y, Π_z are

$$\begin{aligned}&\rho, \qquad \rho u, \qquad \rho v, \qquad \rho w,\\ &\rho u^2 + p_{xx}, \qquad \rho uv + p_{xy}, \qquad \rho uw + p_{xz}, \;\ldots, \qquad \rho w^2 + p_{zz}.\end{aligned}$$

[47] In order to cancel the three terms in parentheses in equation (7), | one can treat them as being independent and focus on one term at a time setting others equal to zero. This yields

$$\gamma_{00}^i = 0, \qquad \gamma_{0j}^i = \gamma_{j0}^i, \qquad \gamma_{ij}^k = \gamma_{ji}^k. \tag{10}$$

These equalities simply express the fact that the three quadratic differential forms

$$\bar{\omega}_0^i dt + \bar{\omega}_1^i dx + \bar{\omega}_2^i dy + \bar{\omega}_3^i dz \qquad (i = 1, 2, 3)$$

vanish identically. One can get the same result from the dynamics of a point particle: the equality

$$\frac{d}{dt}\left(e_0 + e_1\frac{dx}{dt} + e_2\frac{dy}{dt} + e_3\frac{dz}{dt}\right) = 0,$$

as well as equations (7) continue to hold provided one adds to the coefficients ω_0^i, ω_i^j the terms $\bar{\omega}_0^i, \bar{\omega}_i^j$ satisfying:

$$\bar{\omega}_0^i + \bar{\omega}_1^i\frac{dx}{dt} + \bar{\omega}_2^i\frac{dy}{dt} + \bar{\omega}_3^i\frac{dz}{dt} = 0,$$

for all values of the ratios of dx, dy, dz, dt. On the other hand, the results would be different if the components of pressure failed to be symmetric, as is the case when the material is subject to a torque.[14] In this case, expression (9) has to vanish even though

$$p_{xy} \neq p_{yx}, \qquad p_{yz} \neq p_{zy}, \qquad p_{zx} \neq p_{xz}.$$

which implies, as is seen easily, that all the coefficients γ_{ij}^k must vanish leaving only 9 undetermined coefficients instead of 18. In this case, and this case only, does the dynamics of continuous media impose conditions on the affine connection of space-time which are stronger than those imposed by the dynamics of a point particle.

---

14 This occurs for a magnet placed in a magnetic field.

17. Let us now suppose that the affine connection preserves the spatial metric, i.e., that a reference frame with an orthonormal triad cannot be equivalent to one with a non-orthonormal triad. In this case, we can restrict our field of Galilean frames such that the spatial vectors $\mathrm{e}_1, \mathrm{e}_2, \mathrm{e}_3$ are everywhere orthonormal. Then the relations (7) continue to hold but with additional restrictions

$$\omega_i^i = 0, \qquad \omega_i^j + \omega_j^i = 0,$$

| which are imposed by the conditions [48]

$$(\mathrm{e}_i)^2 = 1, \qquad \mathrm{e}_i\mathrm{e}_j = 0 \qquad (i \neq j), i, j = 1, 2, 3.$$

Furthermore, the three quantities $\omega_2^3 = -\omega_3^2, \omega_3^1 = -\omega_1^3$ and $\omega_1^2 = -\omega_2^1$ are now simply the components of the rotation necessary to make the triad T equivalent to T'. Thus, the permissible modifications of the affine connection are dictated by the same conditions as in section 16. However, since now

$$\bar{\omega}_i^i = \bar{\omega}_j^i + \bar{\omega}_i^j = 0,$$

the number of arbitrary coefficients is reduced to four: we have:[15]

$$\bar{\omega}_0^1 = r\,dy - q\,dz, \qquad \bar{\omega}_0^2 = p\,dz - r\,dx, \qquad \bar{\omega}_0^3 = q\,dx - p\,dy,$$
$$\bar{\omega}_2^3 = -\bar{\omega}_3^2 = p\,dt + h\,dx, \quad \bar{\omega}_3^1 = -\bar{\omega}_1^3 = q\,dt + h\,dy, \quad \bar{\omega}_1^2 = -\bar{\omega}_2^1 = r\,dt + h\,dz.$$

Furthermore, had the pressure not been assumed to be symmetric, the coefficient h would have vanished.

18. We shall see later on how, following the viewpoint adopted in section 16 or 17 above, one can select among all affine connections compatible with experiments, a specific one, which can be distinguished from others by its intrinsic properties. However, one may consider a theory in which the angular momentum of a typical element of matter about a point located within the element is not negligible compared to its linear momentum, or in which the stress within the medium manifests itself not only via *forces* but also through *torques*. Under these conditions, the analytic expression of G must contain terms such as $[\mathrm{e}_0\mathrm{e}_i]$ and $[\mathrm{e}_i\mathrm{e}_j]$ and hence *the precise affine connection of space-time would be determined only through experiments;* the experimental data from mechanics would be compatible with only one definition of the equivalence of two Galilean frames whose origins are infinitesimally close.

15 The geometric interpretation of these relations is straightforward.

SPACE-TIME OF SPECIAL RELATIVITY AND ITS AFFINE CONNECTION

19. The theory of special relativity admits the same Galilean frames of reference as classical mechanics. The essential difference lies in the transformation laws between
[49] coordinates (t, x, y, z) which label an | event in one Galilean frame and coordinates (t', x', y', z') which label it in another such frame. These laws are still linear, i.e., the special relativistic space-time continues to be an affine space. However, the time component t of a space-time vector is no longer an invariant. Instead, the invariant quantity now is:

$$c^2(t'-t)^2-(x'-x)^2-(y'-y)^2-(z'-z)^2,$$

where c is the velocity of light in vacuum. As a consequence, the *scalar product*, $c^2\theta\theta'-\xi\xi'-\eta\eta'-\zeta\zeta'$, of two vectors with components $(\theta, \xi, \eta, \zeta)$ and $(\theta', \xi', \eta', \zeta')$, respectively, is also an invariant. In particular, if e_0, e_1, e_2, e_3 denote, as above, unit vectors attached to a Galilean reference frame, the following relations hold:

$$\left.\begin{aligned}&(e_0)^2 = c^2, \qquad (e_1)^2 = (e_2)^2 = (e_3)^2 = -1,\\ &e_0e_i = 0, \qquad e_ie_j = 0 \qquad (i \neq j = 1, 2, 3).\end{aligned}\right\} \tag{11}$$

20. Consider a variable Galilean frame which depends on one or more parameters. For any infinitesimal variation of these parameters, one has:

$$\left.\begin{aligned}de_0 &= \omega_0^0 e_0 + \omega_0^1 e_1 + \omega_0^2 e_2 + \omega_0^3 e_3,\\ de_1 &= \omega_1^0 e_0 + \omega_1^1 e_1 + \omega_1^2 e_2 + \omega_1^3 e_3,\\ de_2 &= \omega_2^0 e_0 + \omega_2^1 e_1 + \omega_2^2 e_2 + \omega_2^3 e_3,\\ de_3 &= \omega_3^0 e_0 + \omega_3^1 e_1 + \omega_3^2 e_2 + \omega_3^3 e_3,\end{aligned}\right\} \tag{12}$$

where the ω_i^j are linear in the differentials of the parameters and are constrained due to equations (11). On differentiating (11) one easily obtains:

$$\omega_0^0 = 0, \qquad \omega_0^i = c^2\omega_i^0, \qquad \omega_i^j + \omega_j^i = 0 \qquad (i, j = 1, 2, 3). \tag{13}$$

Thus, we are left with six independent quantities, which is precisely the number of parameters required to fix the *orientation* of a Galilean frame.

In the above equations, the coefficients ω_0^1, ω_0^2 and ω_0^3 represent, after change of sign, the (infinitesimally small) uniform, translational velocity of the axes of the second Galilean frame w.r.t. those of the first. Note that, in the limit as c tends to infinity, equations (13) reduce to:

$$\omega_0^0 = 0, \qquad \omega_i^0 = 0, \qquad \omega_i^j + \omega_j^i = 0,$$

| so that one recovers the law relating two infinitesimally close Galilean frames[16] in classical mechanics. [50]

21. *The particle dynamics*. The notion of energy-momentum vector continues to underlie the dynamics of a point particle in special relativity. This vector is now given by

$$m\left(\mathrm{e}_0 + \mathrm{e}_1\frac{dx}{dt} + \mathrm{e}_2\frac{dy}{dt} + \mathrm{e}_3\frac{dz}{dt}\right).$$

The *rest mass* μ of the particle is, up to a multiplicative constant, the square root of the scalar product of the energy-momentum vector with itself. More precisely, we have:

$$\mu = m\sqrt{1 - \frac{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2}{c^2}} = m\sqrt{1 - \frac{v^2}{c^2}}.$$

This μ is a number attached to each point particle like the usual mass in classical mechanics.
The mathematical expression of the energy-momentum vector becomes more symmetric if one introduces the *proper time* τ of the point particle, given by

$$d\tau = \sqrt{dt^2 - \frac{dx^2 + dy^2 + dz^2}{c^2}} = \sqrt{1 - \frac{v^2}{c^2}}dt = \frac{\mu}{m}dt.$$

For, the energy-momentum vector can now be written as

$$\mu\left(\frac{dt}{d\tau}\mathrm{e}_0 + \frac{dx}{d\tau}\mathrm{e}_1 + \frac{dy}{d\tau}\mathrm{e}_2 + \frac{dz}{d\tau}\mathrm{e}_3\right),$$

where μ and $d\tau$ are independent of the reference frame.

The fundamental principle of mechanics can be now stated as follows: The derivative of the "energy-momentum space-time vector" w.r.t. the proper time equals the "hyperforce" space-time vector

$$R\mathrm{e}_0 + X\mathrm{e}_1 + Y\mathrm{e}_2 + Z\mathrm{e}_3.$$

The hyperforce vector has an intrinsic significance, independent of the choice of a reference frame: we have |

16 Here, as in section no. 17, the Galilean frames have orthogonal triads.

$$\frac{dm}{dt} = \frac{dm}{d\tau}\frac{d\tau}{dt} = \sqrt{1-\frac{v^2}{c^2}}R,$$

[51]

$$\frac{d}{dt}\left(m\frac{dx}{dt}\right) = \frac{d}{d\tau}\left(m\frac{dx}{dt}\right)\frac{d\tau}{dt} = \sqrt{1-\frac{v^2}{c^2}}X,$$

$$\frac{d}{dt}\left(m\frac{dy}{dt}\right) = \frac{d}{d\tau}\left(m\frac{dy}{dt}\right)\frac{d\tau}{dt} = \sqrt{1-\frac{v^2}{c^2}}Y,$$

$$\frac{d}{dt}\left(m\frac{dz}{dt}\right) = \frac{d}{d\tau}\left(m\frac{dz}{dt}\right)\frac{d\tau}{dt} = \sqrt{1-\frac{v^2}{c^2}}Z.$$

Thus, the force, in the usual sense of the term, is given by the spatial component of the hyperforce times $\sqrt{1-\frac{v^2}{c^2}}$. On the other hand, the constancy of the rest mass introduces constraints among R, X, Y and Z: since

$$c^2 m\frac{dm}{dt} - m\frac{dx}{dt}\frac{d}{dt}\left(m\frac{dx}{dt}\right) - m\frac{dy}{dt}\frac{d}{dt}\left(m\frac{dy}{dt}\right) - m\frac{dz}{dt}\frac{d}{dt}\left(m\frac{dz}{dt}\right) = 0,$$

we have

$$c^2 R dt = X dx + Y dy + Z dz.$$

This relation expresses the fact that the infinitesimal work done by the force equals the change in the quantity mc^2. The quantity

$$mc^2 = \frac{\mu c^2}{\sqrt{1-\frac{v^2}{c^2}}}$$

is the *energy* of the point particle. Indeed, if V is small compared to c, in the first approximation, mc^2 equals

$$\mu c^2 + \frac{1}{2}\mu v^2,$$

or, equivalently,

$$\mu c^2 + \frac{1}{2}m v^2.$$

22. *The dynamics of continuous media*. If we restrict ourselves to the special case
[52] in which all volume forces are absent, the equations | of dynamics of continuous media are essentially the same as in classical dynamics. Thus, we can introduce the sliding vector representing the energy-momentum of an element of matter,

$$\mathrm{G} = [\mathrm{me}_0]\Pi + [\mathrm{me}_1]\Pi_x + [\mathrm{me}_2]\Pi_y + [\mathrm{me}_3]\Pi_z,$$

and set its exterior derivative, G', to be identically zero. If the medium is equipped with a fixed Galilean frame, the components Π, Π_x, Π_y, Π_z can be again expressed as:

$$\begin{aligned}
\Pi &= \rho dxdydz - \rho u dydzdt - \rho v dzdxdt - \rho w dxdydt, \\
\Pi_x &= u\Pi - p_{xx}dydzdt - p_{xy}dzdxdt - p_{xz}dxdydt, \\
\Pi_y &= v\Pi - p_{yx}dydzdt - p_{yy}dzdxdt - p_{yz}dxdydt, \\
\Pi_z &= w\Pi - p_{zx}dydzdt - p_{zy}dzdxdt - p_{zz}dxdydt.
\end{aligned}$$

The density ρ_0 of the matter in its rest frame is given by:

$$\rho_0[dtdxdydz] = [dt\Pi] - \frac{1}{c^2}[dx\Pi_x] - \frac{1}{c^2}[dy\Pi_y] - \frac{1}{c^2}[dz\Pi_z],$$

where the right-hand side is essentially the scalar product of the vector (dt, dx, dy, dz) with the vector $(\Pi, \Pi_x, \Pi_y, \Pi_z)$. On simplifying, one obtains[17]

$$\rho_0 = \rho\left(1 - \frac{u^2 + v^2 + w^2}{c^2}\right) - \frac{1}{c^2}(p_{xx} + p_{yy} + p_{zz}).$$

Thus, in absence of external forces, the equations of dynamics of continuous media are identical to those of classical mechanics. Indeed, if a variable Galilean frame is attached to each point of space-time, one would obtain

$$\begin{aligned}
\mathrm{G}' = \; & [\mathrm{me}_0][\Pi' + \omega_1^0\Pi_x + \omega_2^0\Pi_y + \omega_3^0\Pi_z] \\
& + [\mathrm{me}_1][\Pi'_x + \omega_0^1\Pi + \omega_2^1\Pi_y + \omega_3^1\Pi_z] \\
& + [\mathrm{me}_2][\Pi'_y + \omega_0^2\Pi + \omega_1^2\Pi_x + \omega_3^2\Pi_z] \\
& + [\mathrm{me}_3][\Pi'_z + \omega_0^3\Pi + \omega_1^3\Pi_x + \omega_2^3\Pi_y] \\
& + [\mathrm{e}_0\mathrm{e}_1][\omega^0\Pi_x - \omega^1\Pi\;] + [\mathrm{e}_0\mathrm{e}_2][\omega^0\Pi_y - \omega^2\Pi\;] + [\mathrm{e}_0\mathrm{e}_3][\omega^0\Pi_z - \omega^3\Pi\;] \\
& + [\mathrm{e}_2\mathrm{e}_3][\omega^2\Pi_z - \omega^3\Pi_y] + [\mathrm{e}_3\mathrm{e}_1][\omega^3\Pi_x - \omega^1\Pi_z] + [\mathrm{e}_1\mathrm{e}_2][\omega^1\Pi_y - \omega^2\Pi_x].
\end{aligned}$$

23. Let us investigate whether or not several distinct affine connections can be compatible with experiments. In the passage from one | connection to another, the [53] components ω_0^i and ω_j^i undergo variations, $\bar{\omega}_0^i$ and $\bar{\omega}_0^j$, satisfying, of course, the relations $\bar{\omega}_0^i = c^2\bar{\omega}_i^0$, and $\bar{\omega}_i^j + \bar{\omega}_j^i = 0$. Furthermore, these variations should be such that the four terms

17 Note that, since the volume element $[dtdxdydz]$ is independent of the choice of the reference frame, the quantity ρ_0 has an absolute, frame-independent significance. It is of course not so for the apparent density ρ.

$$[\overline{\omega}_1^0\Pi_x] + [\overline{\omega}_2^0\Pi_y] + [\overline{\omega}_3^0\Pi_z],$$
$$[\overline{\omega}_0^1\Pi\ \ ] + [\overline{\omega}_2^1\Pi_y] + [\overline{\omega}_3^1\Pi_z],$$
$$[\overline{\omega}_0^2\Pi\ \ ] + [\overline{\omega}_1^2\Pi_x] + [\overline{\omega}_3^2\Pi_z],$$
$$[\overline{\omega}_0^3\Pi\ \ ] + [\overline{\omega}_1^3\Pi_x] + [\overline{\omega}_2^3\Pi_y].$$

must vanish identically, *irrespective of the state of the element of matter under consideration*. If the material medium is described using a fixed Galilean frame, one finds, as in section 16, that the four quadratic forms

$$\overline{\omega}_0^i dt + \overline{\omega}_1^i dx + \overline{\omega}_2^i dy + \overline{\omega}_3^i dz = 0 \qquad (i = 0, 1, 2, 3)$$

must vanish identically. One obtains the same expressions for $\overline{\omega}_0^i$ and $\overline{\omega}_j^i$ in terms of four arbitrary coefficients p, q, r, and h as in section 17.

Finally, had we enlarged our notion of mechanics of continuous media by allowing terms of the form $[e_0 e_i]$ and $[e_i e_j]$ in the expression of the energy-momentum density, there would have remained no arbitrary coefficients: *the affine connection of space-time would have been completely determined experimentally*.[18]

24. *Gravitation in special relativity.* In classical mechanics, the equations

$$\omega_0^1 = -X dt, \qquad \omega_0^2 = -Y dt, \qquad \omega_0^3 = -Z dt, \qquad \omega_i^j = 0 \quad (i, j = 1, 2, 3)$$

defining the affine connection which enables geometrization of gravity, preserve their form under the change of the Galilean frame of reference. In special relativity, one may postulate that Newton's law of gravity holds in the Galilean frame whose axes point towards the fixed stars and have the centre of mass of the solar system as origin. However, the law would not have the same form in other Galilean frames. On the other hand, we may follow Einstein and postulate that the law of gravity should have an invariant expression irrespective of the Galilean frame which is used.[19] We are then forced to modify the law itself. Nevertheless, let us note here that the resulting
[54] geometrical | formulation of gravity due to Einstein is essentially similar to the one mentioned in the beginning of this chapter.

25. *The viewpoint of general relativity.* Up to now, we have worked under the assumption that there exist Galilean frames which can label points in the *entire* space-time. At this point, however, it is clear how one can get rid of this assumption.

---

18 This is so if we simply allow torques to act on the elements of matter. For, in that case, the coefficient h is necessarily zero, and, since p, q, r, h transform into one another as components of a 4-vector under Galilean transformations, *one is forced to conclude that* p, q, r *vanish*.

19 The precise meaning of this phrase will become clear later on.

Indeed, to formulate physical laws, it is sufficient that the following two conditions are satisfied:

1) To measure quantities of physical interest, one has available a local reference frame which plays the role of a true Galilean frame[20] in a patch of space-time immediately surrounding the observer, and

2) One knows the space-time connection, i.e., one knows how to compare the observations carried out in two Galilean frames whose origins are infinitesimally close. One may reformulate this condition by saying that one has to know the Lorentz-Minkowski transformation required to make two frames coincide. Analytically, this means that one should know the coefficients in equations (8) and (12).

We shall now go on to the theory of manifolds with an affine connection. Application of this theory to general relativity will follow. We shall also examine the way in which the laws of electromagnetism serve to determine the affine connection of space-time.

20 This is obviously not the place to enter into a discussion of practical difficulties which may arise in assimilating a given reference system to a Galilean frame.

INDEX

Explanatory entries are marked in boldface

G

K

L

T

U

V

W

Z