Vesselin Petkov
Editor

FUNDAMENTAL THEORIES OF PHYSICS 153

Relativity and the Dimensionality of the World

Springer

Relativity and the Dimensionality of the World

Fundamental Theories of Physics

An International Book Series on The Fundamental Theories of Physics: Their Clarification, Development and Application

Volume 153

Relativity and the Dimensionality of the World

Edited by

Vesselin Petkov
Concordia University
Montreal, QC, Canada

Springer

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

ISBN 978-1-4020-6317-6 (HB)
ISBN 978-1-4020-6318-3 (e-book)

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

www.springer.com

Printed on acid-free paper

Preface

There is a story behind every book. The story of this book is about the open question of the nature of space-time and whether it is a question that physics should deal with.

All physicists would agree that one of the most fundamental problems of the twenty-first-century physics is the dimensionality of the world. But this seems to be the only thing that can be said with certainty when the dimensions of the world are discussed. On the one hand, physicists freely talk about multidimensional spaces and even the number of books they have written for the general public on topics such as hidden dimensions and parallel Universes has been steadily increasing. On the other hand, however, when it comes to the first higher-than-3D world predicted by special relativity – the 4D world of Minkowski (or Minkowski space-time) – there is no consensus among physicists on what such a world looks like.

The most challenging problem is the nature of the temporal dimension. In Minkowski space-time it is merely one of the 4D, which means that it is *entirely* given like the other three spacial dimensions. If the temporal dimension were not given in its entirety and only one constantly changing moment of it existed, Minkowski space-time would be reduced to the ordinary 3D space. But if the physical world, represented by Minkowski space-time, is indeed 4D with time being the forth dimension, then such a world is drastically different from its image based on our perceptions. Minkowski 4D world is a block universe. It is a frozen world in which nothing happens, since all moments of time are given "at once", which means that physical bodies are 4D worldtubes containing the whole histories in time of the 3D bodies of our everyday experience. The implications of a real Minkowski world for physics itself and especially for our world view are enormous. If the world turns out to be 4D:

- Even the language of physics should be changed to reflect the fact that it describes a static world. The presently used language would be inadequate and misleading in such a world since, for example, there are no moving particles and even no 3D particles in space-time; what space-time contains are only the particles' worldtubes. All interactions between particles are entirely realized in space-time and are represented by a forever-given network of the particles' worldtubes. The realization that what we interpret as interactions in our 3D language is, in fact, just geometry[1] in space-time had led Minkowski to the conclusion that "[T]he whole universe is seen to resolve itself into similar world lines, and I would fain anticipate myself by saying that in my opinion physical laws might find their most perfect expression as reciprocal relations between these world-lines" [1, p. 76]. Therefore, "[T]hree-dimensional geometry becomes a chapter in four-dimensional physics" [1, p. 80].

- The relativistic effects and the *experiments* that confirmed them should be rigorously analyzed and if it is found that they are indeed manifestations of the 4D of the world as Minkowski advocated that would be the end of the debate over the nature of space-time. Then no arguments based on quantum physics or any other arguments could change the *experimental* fact that reality is a 4D (block) world.
- Although the relativistic effects can be regarded as the manifestations of the 4D of the Minkowski world one can still try to determine whether the world reveals its dimensionality through other physical phenomena. One such manifestation is virtually obvious – it appears natural to expect that the deformed (non-geodesic) worldtube of a non-inertial particle would resist its deformation (like a deformed 3D rod); that resistance would be observed as the resistance a non-inertial particle offers whenever deviated from its geodesic path. Hence, inertia would be another manifestation of the 4D of the world since it would turn out to be a result of a 4D stress arising in the deformed worldtube of a non-inertial particle [3].
- The impact of the implications of a real 4D world for our entire world view would constitute perhaps the greatest intellectual challenge humankind has ever faced. It is sufficient to mention just one of those implications – that there is no free will in the frozen Minkowski world.

When we face such a challenge coming from a physical theory, I think physicists and especially relativists should lead the assault on it. So far, however, most relativists have avoided facing this challenge by either declaring that the question about the reality of the 4D world of relativity is a philosophical question or by pointing out that that question needs no answer since the 3D and 4D representations of relativity are just two equivalent descriptions. Unfortunately, neither of these two reasons makes the challenge disappear. Moreover, I wonder how many physicists, if directly asked, would agree that the question "What is the dimensionality of the world according to relativity?" is a philosophical one. The 3D and 4D representations of relativity may be considered equivalent in a sense that they adequately *describe* the relativistic effects but with respect to the dimensionality of the world they are radically different since the world is *either* 3D *or* 4D (at least at the macroscopic level where relativity is wholly applicable[2]). In this sense only one of the two representations of relativity is correct.

[1] I think Wheeler prematurely abandoned his geometrodynamics [2] and the whole idea of reducing physics to geometry. The difficulty to deal with fermions, for instance, could not be a crucial reason for that. By the same argument relativity should be also abandoned; see footnote 2. On the other hand, if it is established that reality is a 4D world, then the conclusion that (macroscopic) physics is just geometry is inescapable.

[2] The equations of motion of relativity do not describe adequately the motion of the quantum objects, which means that relativity is not wholly applicable at the quantum level.

As in 2008 we will be celebrating the 100th anniversary of Minkowski's talk on space-time I think we owe him a thorough examination of the physical meaning of his main contribution to physics – Minkowski space-time (and the other relativistic space-times since they are based on the same fundamental idea of uniting space and time into an inseparable 4D entity). To further stimulate the research on the nature of space-time was the main reason behind the idea to establish a society and to hold biennial conferences on space-time. The International Society for the Advanced Study of Spacetime (http://www.spacetimesociety.org) was founded during the First International Conference on the Ontology of Spacetime, which was held at Concordia University in Montreal from 11 to 14 May 2004. It was decided that at least for the time being the conferences would be held at Concordia University. The second conference took place from 9 to 11 June 2006 and the third, which will commemorate Minkowski's anniversary, will be held from 13 to 15 June 2008.

The Ontology of Spacetime [4] containing selected papers presented at the first space-time conference was published in August 2006. As the contributors were mostly philosophers I believed a second volume on the same issue should be written by physicists, preferably by relativists. Moreover, it is the relativists who will have the last and definite say on what the dimensionality of the world according to relativity is. For this reason I invited physicists actively working in relativity, cosmology, or relativity-related areas to contribute papers to a volume whose main focus is the question:

> Is space-time nothing more than a mathematical space (which describes the evolution in time of the ordinary three-dimensional world) or is it a mathematical model of a real four-dimensional world with time entirely given as the fourth dimension?

While contacting the potential authors I took the opportunity and included in the invitation a concise argument demonstrating why it is a valid question to ask what the physical meaning of space-time is. Many of the invited found that the "project sounds very interesting" and that "the time is probably ripe to discuss it," but not all of them decided to participate in it. The present volume contains 14 papers which either directly address the main question of the nature of space-time or explore issues related to it.

I would like to thank all invited colleagues who considered making contributions to this volume and especially those who did contribute.

Montreal *Vesselin Petkov*
12 December 2006

References

1. H. Minkowski: Space and Time. In: H.A. Lorentz, A. Einstein, H. Minkowski, and H. Weyl (eds.): *The Principle of Relativity: A Collection of Original Memoirs on the Special and General Theory of Relativity* (Dover, New York, 1952) pp 75–91

2. J.A. Wheeler: *Geometrodynamics* (Academic Press, New York, 1963)
3. V. Petkov: *Relativity and the Nature of Spacetime* (Springer, Berlin, Heidelberg, New York, 2005), Ch 10
4. D. Dieks (ed.): *The Ontology of Spacetime* (Elsevier, Amsterdam, 2006); Philosophy and Foundations of Physics Series

Contents

List of Contributors

R. Aldrovandi
Instituto de Fisica Teorica – UNESP, Säo Paulo, Brazil

H. Arcos
Instituto de Fisica Teorica – UNESP, Säo Paulo, Brazil

Carlos Barceló
Instituto de Astrofísica de Andalucía (CSIC), Spain
e-mail: carlos@iaa.es

Orfeu Bertolami
Departamento Física, Instituto Superior Técnico, Lisbon, Portugal
e-mail: orfeu@cosmos.ist.utl.pt

Martin Bojowald
Institute for Gravitation and the Cosmos
The Pennsylvania State University
e-mail: bojowald@gravity.psu.edu

Joy Christian
Perimeter Institute and University of Oxford
e-mail: joy.christian@wolfson.oxford.ac.uk

Michael Cifone
Department of Philosophy, University of Maryland

George F. R. Ellis
Department of Mathematics and Applied Mathematics
University of Cape Town
e-mail: Ellis@MATHS.uct.ac.za

Graham Hall
Department of Mathematical Sciences, University of Aberdeen
e-mail: g.hall@maths.abdn.ac.uk

Luca Lusanna
INFN Firenze, Italy
e-mail: LUSANNA@FI.INFN.IT

Hans C. Ohanian
Department of Physics, University of Vermont
e-mail: HansOhanian@cs.com

Massimo Pauri
I.N.F.N Sezione di Milano-Bicocca,
Gruppo Collegato di Parma
Department of Physics, University of Parma
e-mail: pauri@pr.infn.it

José Geraldo Pereira
Instituto de Fisica Teorica - UNESP, Säo Paulo, Brazil
e-mail: jpereira@ift.unesp.br

Vesselin Petkov
Science College, Concordia University
e-mail: vpetkov@alcor.concordia.ca

C. W. Rietdijk
Pinellaan 7, 2061LH Bloemendaal, The Netherlands
e-mail: cw.rietdijk@quicknet.nl

Michael Silberstein
Department of Philosophy, Elizabethtown College
e-mail: silbermd@etown.edu

W.M. Stuckey
Department of Physics, Elizabethtown College
e-mail: stuckeym@etown.edu

Paul Wesson
Physics and Applied Mathematics, University of Waterloo
e-mail: pswastro@hotmail.com

Chapter 1
The Meaning of Dimensions

Paul S. Wesson

Abstract We review the current status of dimensions, as the result of a long and controversial history that includes input from philosophy and physics. Our conclusion is that they are subjective but essential concepts which provide a kind of bookkeeping device, their number increasing as required by advances in physics. The world almost certainly has more than the four dimensions of space and time, but the introduction of the fifth and higher dimensions requires a careful approach wherein known results are embedded and new ones are couched in the most productive manner.

1.1 Introduction

Dimensions are both primitive concepts that provide a framework for mechanics and sophisticated devices that can be used to construct unified field theories. Thus the ordinary space of our perceptions (xyz) and the subjective notion of time (t) provide the labels with which to describe Newtonian mechanics, or with the introduction of the speed of light to form an extra coordinate (ct) the mechanics of 4D Einstein relativity. But used in the abstract, they also provide a means of extending general relativity in accordance with certain physical principles, as in 10D supersymmetry. As part of the endeavour to unify gravity with the interactions of particle physics, there has recently been an explosion of interest in manifolds with higher dimensions. Much of this work is algebraic in nature, and has been reviewed elsewhere. Therefore, to provide some balance and direction, we will concentrate here on fundamentals and attempt to come to an understanding of the meaning of dimensions.

Our main conclusion, based on 35 years of consideration, will be that dimensions are basically inventions, which have to be chosen with skill if they are to be profitable in application to physics.

This view may seem strange to some workers, but is not new. It is implicit in the extensive writings on philosophy and physics by the great astronomer Eddington,

V. Petkov (ed.), *Relativity and the Dimensionality of the World*, 1–11.

and has been made explicit by his followers, who include the writer. This view is conformable, it should be noted, with algebraic proofs and other mathematical results on many-dimensional manifolds, such as those of the classical geometer Campbell, whose embedding theorem has been recently rediscovered and applied by several workers to modern unified-field theory. Indeed, a proper understanding of the meaning of dimensions involves both history and modern physics.

There is a large literature on dimensions; but it would be inappropriate to go into details here, and we instead list some key works. The main philosophical/physical ones are those by Barrow (1981), Barrow and Tipler (1986), Eddington (1935, 1939), Halpern (2004), Kilmister (1994), McCrea and Rees (1983), Petley (1985), Price and French (2004) and Wesson (1978, 1992). The main algebraic/mathematical works are those by Campbell (1926), Green et al. (1987), Gubser and Lykken (2004), Seahra and Wesson (2003), Szabo (2004), Wesson (1999, 2006) and West (1986). These contain extensive bibliographies, and we will quote freely from them below.

The plan of this paper is as follows. Section 2 outlines the view of our group, that dimensions are inventions whose geometrical usefulness for physics involves a well-judged use of the fundamental constants. This rests on work by Eddington, Campbell and others; so in sections 3 and 4 we give accounts of the main philosophical and algebraic results (respectively) due to these men. Section 5 is a summary, where we restate our view that the utility of dimensions in physics owes at least as much to skill as to symbolism. We aim to be pedagogical rather than pedantic, and hope that the reader will take our comments in the spirit of learning rather than lecture.

1.2 Dimensions and Fundamental Constants

Minkowski made a penetrating contribution to special relativity, and our view of mechanics, when by the simple identification of $x^4 \equiv ct$ he put time on the same footing as the coordinates $\left(x^{123} = xyz\right)$ of the ordinary space of our perceptions. Einstein took an even more important step when he made the Principle of Covariance one of the pillars of general relativity, showing that the 4 coordinates traditionally used in mechanics can be altered and even mixed, producing an account of physical phenomena which is independent of the labels by which we choose to describe them. These issues are nowadays taken for granted; but a little reflection shows that insofar as the coordinates are the labels of the dimensions, the latter are themselves flexible in nature.

Einstein was in his later years also preoccupied with the manner in which we describe matter. His original formulation of general relativity involved a match between a purely geometrical object we now call the Einstein tensor ($G_{\alpha\beta}$, where α and $\beta = 0, 123$ for t, xyz), and an object which depends on the properties of matter which is known as the energy–momentum (or stress–energy) tensor ($T_{\alpha\beta}$, which contains quantities like the ordinary density ρ and pressure p of matter). The coefficient necessary to turn this correspondence into an equation is (in suitable units) $8\pi G/c^4$, where G is the gravitational constant. Hence, Einstein's field equations, $G_{\alpha\beta} = \left(8\pi G/c^4\right) T_{\alpha\beta}$, which are an excellent description of gravitating matter. In writing these equations, it is common to read them from left to right, so that the geometry of 4D space-time is governed by the matter it contains. However, this split is artificial. Einstein himself realized this, and sought (unsuccessfully) for some way to turn the "base wood" of $T_{\alpha\beta}$ into the "marble" of $G_{\alpha\beta}$. His aim, simply put, was to geometrize all of mechanical physics – the matter as well as the fields.

A potential way to geometrize the physics of gravity and electromagnetism was suggested in 1920 by Kaluza, who added a fifth dimension to Einstein's general relativity. Kaluza showed in essence that the apparently empty 5D field equations $R_{AB} = 0\,(A, B = 0, 123, 4)$ in terms of the Ricci tensor, contain Einstein's equations for gravity and Maxwell's equations for electromagnetism. Einstein, after some thought, endorsed this step. However, in the 1920s quantum mechanics was gaining a foothold in theoretical physics, and in the 1930s there was a vast expansion of work in this area, at the expense of general relativity. This explains why there was such a high degree of attention to the proposal of Klein, who in 1926 suggested that the fifth dimension of Kaluza ought to have a closed topology (i.e. a circle), in order to explain the fundamental quantum of electric charge (e). Klein's argument actually related this quantity to the momentum in the extra dimension, but in so doing employed the fundamental unit of action (h) or Planck's constant. However, despite the appeal of Klein's idea, it was destined for failure. There are several technical reasons for this, but it is sufficient to note here that the crude 5D gravity/quantum theory of Kaluza/Klein implied a basic role for the mass quantum $(hc/G)^{1/2}$. This is of order 10^{-5} g, and does *not* play a dominant role in the spectrum of masses observed in the real universe. (In more modern terms, the so-called hierarchy problem is centred on the fact that observed particle masses are far less than the Planck mass, or any other mass derivable from a tower of states where this is a basic unit.) Thus, we see in retrospect that the Klein modification of the Kaluza scheme was a dead end. This does not, though, imply that there is anything wrong with the basic proposition, which follows from the work of Einstein and Kaluza that matter can be geometrized with the aid of the fundamental constants. As a simple example, an astrophysicist presented with a problem involving a gravitationally dominated cloud of density ρ will automatically note that the free fall or dynamical timescale is the inverse square root of $G\rho$. This tells him immediately about the expected evolution of the cloud. Alternatively, instead of taking the density as the relevant physical quantity, we can form the *length* $\left(c^2/G\rho\right)^{1/2}$ and obtain an equivalent description of the physics in terms of a geometrical quantity.

The above simple outline, of how physical quantities can be combined with the fundamental constants to form geometrical quantities such as lengths, can be much developed and put on a systematic basis (Wesson 1999). The result is induced-matter theory, or as some workers prefer to call it, space–time–matter theory. The philosophical basis of the theory is to realize Einstein's dream of unifying geometry and matter (see above). The mathematical basis of it is Campbell's theorem, which ensures an embedding of 4D general relativity with sources in a 5D theory whose field equations are apparently empty (see below). That is, the Einstein equations $G_{\alpha\beta} = \left(8\pi G/c^4\right) T_{\alpha\beta}\,(\alpha,\beta = 0,123)$ are embedded perfectly in the Ricci-flat equations $R_{AB} = 0\,(A,B = 0,123,4)$. The point is, in simple terms, that we use the fifth dimension to model matter.

An alternative version of 5D gravity, which is mathematically similar, is membrane theory. In this, gravity propagates freely in 5D, into the "bulk"; but the interactions of particles are confined to a hypersurface or the "brane". It has been shown by Ponce de Leon and others that both the field equations and the dynamical equations are effectively the same in both theories. The only difference is that whereas induced-matter theory treats all 5 dimensions as equivalent, membrane theory makes spacetime a special (singular) hypersurface. For induced-matter theory, particles can wander away from the hypersurface at a slow rate governed by the cosmological constant; whereas for membrane theory, particles are confined to the hypersurface by an exponential force governed by the cosmological constant. Both versions of 5D general relativity are in agreement with observations. The choice between them is largely philosophical: Are we living in a universe where the fifth dimension is "open", or are we living an existence where we are "stuck" to a particular slice of a 5D manifold?

The fundamental constants available to us at the present stage in the development of physics allow us to geometrize matter in terms of *one* extra dimension. Insofar as mechanics involves the basic physical quantities of mass, length and time, it is apparent that any code for the geometrization of mass will serve the purpose of extending 4D space-time to a 5D space-time-mass manifold (the theory is covariant). However, not all parametizations are equally convenient, in regard to returning known 4D physics from a 5D definition of "distances" or metric. Thus, the "canonical" metric has attracted much attention. In it, the line element is augmented by a flat extra dimension, while its 4D part is multiplied by a quadratic factor (the corresponding metric in membrane theory involves an exponential factor, as noted above). The physics flows from this factor, which is $(l/L)^2$ where $x^4 = l$ and L is a constant, which by comparison with the 4D Einstein metric means $L = (3/\Lambda)^{1/2}$ where Λ is the cosmological constant. In this way, we weld ordinary mechanics to cosmology, with the identification $x^4 = l = Gm/c^2$ where m is the rest mass of a macroscopic object. If, on the other hand, we wish to study microscopic phenomena, the simple coordinate transformation $l \to L^2/l$ gives us a quantum (as opposed to classical) description of rest mass via $x^4 = h/mc$. In other words, the large and small scales are accommodated by choices of coordinates which utilize the available fundamental constants, labelling the mass either by the Schwarzchild radius or by the Compton wavelength.

It is not difficult to see how to extend the above approach to higher dimensions. However, skill is needed here. For example, electric charge can either be incorporated into 5D (along the lines originally proposed by Kaluza and Klein), or treated as a sixth dimension (with coordinate $x_q \equiv \left(G/c^4\right)^{1/2} q$ where q is the charge, as studied by Fukui and others). A possible resolution of technical problems like this is to "fill up" the parameter space of the lowest-dimensional realistic model (in this case 5D), before moving to a higher dimension. As regards other kinds of "charges" associated with particle physics, they should be geometrized and then treated as coordinates in the matching N-dimensional manifold. In this regard, as we have emphasized, there are choices to be made about how best to put the physics into correspondence with the algebra. For example, in supersymmetry, every integral-spin boson is matched with a half-integral-spin fermion, in order to cancel off the enormous vacuum or zero-point fields which would otherwise occur. Now, it is a theorem that any curved energy-full solution of the 4D Einstein field equations can be embedded in a flat and energy-free 10D manifold. (This is basically a result of counting the degrees of freedom in the relevant sets of equations: see section 4 below). This is the simplest motivation known to the writer for supersymmetry. However, it is possible in certain cases that the condition of zero energy can be accomplished in a space of less than 10 dimensions, given a skillful choice of parameters.

We as physicists have chosen geometry as the currently best way to deal with macroscopic and microscopic mechanics; and while there are theorems which deal with the question of how to embed the 4D world of our senses in higher-dimensional manifolds, the choice of the latter requires intuition and skill.

1.3 Eddington and His Legacy

In studying dimensions and fundamental constants over several decades, the writer has come to realize that much modern work on these topics has its roots in the views of Sir Arthur Stanley Eddington (1882–1944; for a recent interdisciplinary review of his contributions to physics and philosophy, see the conference notes edited by Price and French, 2004). He was primarily an astronomer, but with a gift for the pithy quote. For example: "We are bits of stellar matter that got cold by accident, bits of a star gone wrong." However, Eddington also thought deeply about more basic subjects, particularly the way in which science is done, and was of the opinion that much of physics is subjective, insofar as we necessarily filter data about the external world through our human-based senses. Hence, the oft-repeated quote: "To put the conclusion crudely – the stuff of the world is mind-stuff." The purpose of the present section is to give a short and informal account of Eddington's views, and thereby alert workers in fundamental physics to his influence.

His impact was primarily through a series of non-technical books, and his personal contacts with a series of great scientists who followed his path. These include Dirac, Hoyle and McCrea. In the preceding section, we noted that while it is possible to add an arbitrary number of extra dimensions to relativity as an exercise in mathematics, we need to use the fundamental constants to identify their relevance to physics. (We are here talking primarily about the speed of light c, the gravitational constant G and Planck's constant of action h, which on division by 2π also provides the quantum of spin-angular momentum.) To appreciate Eddington's legacy, we note that his writings contain the first logical account of the large dimensionless numbers which occur in cosmology, thereby presaging what Dirac would later formalize as the Large Numbers Hypothesis. This consists basically in the assertion that large numbers of order 10^{40} are in fact equal, which leads among other consequences to the expectation that G is variable over the age of the universe (see Wesson 1978). This possibility is now discussed in the context of field theory in $N > 4$ dimensions, where the dynamics of the higher-dimensional manifold implies that the coupling constants (like G) in 4D are changing functions of the spacetime coordinates (Wesson 1999). One also finds in Eddington's works some very insightful, if controversial, comments about the so-called fundamental constants. These appear to have influenced Hoyle, who argued that the c^2 in the common relativistic expression $\left(c^2t^2 - x^2 - y^2 - z^2\right)$ should not be there, because "there is no more logical reason for using a different time unit than there would be for measuring x, y, z in different units". The same influence seems to have acted on McCrea, who regarded c, G and h as "conversion constants and nothing more". These comments are in agreement with the view advanced in section 2, namely that the fundamental constants are parameters which can be used to change the physical units of material quantities to lengths, enabling them to be given a geometrical description.

There is a corollary of this view which is pertinent to several modern versions of higher-dimensional physics. Whatever the size of the manifold, the equations of the related physics are homogeneous in their physical units (M, L, T) so they can always be regarded as equalities involving *dimensionless* parameters. It makes sense to consider the possible (say) time variation of such parameters; but it makes no sense to argue that the component *dimensionfull* quantities are variable. To paraphrase Dicke: Physics basically consists of the comparison of dimensionless parameters at different points in the manifold.

Views like this still raise the hackles of certain physicists who have not analysed the problem at a deep level. Eddington, in particular, was severely criticized by both physicists and philosophers when he presented his opinions in the 1930s. Fortunately, many workers – as a result of their studies of unified field theory – came to a sympathetic understanding of Eddington's opinions in the 1990s. However, there is an interesting question of psychology involved here.

Plato tells us of an artisan whose products are the result of experience and skill and meet with the praise of his public for many years. However, in later times he suddenly produces a work, which is stridently opposed to tradition and incurs widespread criticism. Has the artisan suffered some delusion, or has he broken through to an art form so novel that his pedestrian-minded customers cannot appreciate or understand it?

Eddington spent the first part of his academic career doing well-regarded research on stars and other aspects of conventional astronomy. He then showed great insight and mathematical ability in his study of the then-new subject of general relativity. In his later years, however, he delved into the arcane topic of the dimensionless numbers of physics, attempting to derive them from an approach which combined elements of pure reason and mathematics. This approach figures significantly in his book *Relativity Theory of Protons and Electrons*, and in the much-studied posthumous volume *Fundamental Theory*. The approach fits naturally into his philosophy of science, which latter argued that many results in physics are the result of *how* we do science, rather than direct discoveries about the external world (which, however, he admitted). Jeffreys succeeded Eddington to the Plumian Chair at Cambridge, but was a modest man more interested in geophysics and the formation of the solar system than the speculative subject of cosmology. Nevertheless, he developed what at the time was a fundamental approach to the theory of probability, and applied his skills to a statistical analysis of Eddington's results. The conclusion was surprising: according to Jeffreys' analysis of the uncertainties in the underlying data, which Eddington had used to construct his account of the basic physical parameters, the results agreed with the data *better than they ought to have done*. This raised the suspicion that Eddington had "cooked" the results. This author spent the summer of 1970 in Cambridge, having written (during the preceding summer break from undergraduate studies at the University of London) a paper on geophysics which appealed to Jeffreys. We discussed, among other things, the status of Eddington's results. Jeffreys had great respect for Eddington's abilities, but was of the opinion that his predecessor had *unwittingly* put subjective elements into his approach, which accounted for their unreasonable degree of perfection. The writer pointed out that there was another possible explanation: that Eddington was in fact right in his belief that the results of physics were derivable from first principles, and that his approach was compatible with a more profound theory which yet awaits discovery.

1.4 Campbell and His Theorem

Whatever the form of a new theory which unifies gravity with the forces of particle physics, there is a consensus that it will involve extra dimensions. In section 2, we considered mainly the 5D approach, which by the modern names of induced-matter and membrane theory is essentially old Kaluza–Klein theory without the stifling condition of compactification. The latter, wherein the extra dimension is "rolled up" to a very small size, answers the question of why we do not "see" the fifth dimension. However, an equally valid answer to this is that we are constrained to live close to a hypersurface, like an observer who walks across the surface of the Earth without being directly aware of what lies beneath his feet.

In this interpretation, 5D general relativity must be regarded as a kind of new standard. It is the simplest extension of Einstein's theory, and is widely viewed as the low-energy limit of more sophisticated theories which accommodate the internal symmetry groups of particle physics, as in 10D supersymmetry, 11D supergravity and 26D string theory. There is, though, no sancrosanct value of the dimensionality N. It has to be chosen with a view to what physics is to be explained. (In this regard, St. Kalitzin many years ago considered $N \to \infty$.) All this understood, however, there is a practical issue which needs to be addressed and is common to all higher-N theories: How do we embed a space of dimension N in one of dimension $(N+1)$? This is of particular relevance to the embedding of 4D Einstein theory in 5D Kaluza–Klein theory. We will consider this issue in the present section, under the rubric of Campbell's theorem. While it is central and apparently simple, it turns out to have a rather long history with some novel implications.

John Edward Campbell was a professor of mathematics at Oxford whose book *A Course of Differential Geometry* was published posthumously in 1926. The book is basically a set of lecture notes on the algebraic properties of ND Riemannian manifolds, and the question of embeddings is treated in the latter part (notably Chapters 12 and 14). However, what is nowadays called Campbell's theorem is there only sketched. He had intended to add a chapter dealing with the relation between abstract spaces and Einstein's theory of general relativity (which was then a recent addition to physics), but died before he could complete it. The book was compiled with the aid of Campbell's colleague, E.B. Elliot, but while accurate is certainly incomplete.

The problem of embedding an ND (pseudo-) Riemannian manifold in a Ricci-flat space of one higher dimension was taken up again by Magaard. He essentially proved the theorem in his Ph.D. thesis of 1963. This and subsequent extensions of the theorem have been discussed by Seahra and Wesson (2003), who start from the Gauss–Codazzi equations and consider an alternative proof which can be applied to the induced-matter and membrane theories mentioned above.

The rediscovery of Campbell's theorem by physicists can be attributed largely to the work of Tavakol and co-workers. They wrote a series of articles in the mid-1990s which showed a connection between the CM theorem and a large body of earlier results by Wesson and co-workers (Wesson 1999). The latter group had been using 5D geometry as originally introduced by Kaluza and Klein to give a firm basis to the aforementioned idea of Einstein, who wished to transpose the "base-wood" of the right-hand side of his field equations into the "marble" of the left-hand side.

That an effective or induced 4D energy–momentum tensor $T_{\alpha\beta}$ can be obtained from a 5D geometrical object such as the Ricci Tensor R_{AB} is evident from a consideration of the number of degrees of freedom involved in the problem (see below). The only requirement is that the 5D metric tensor be left general, and not be restricted by artificial constraints such as the "cylinder" condition imposed by Kaluza and Klein. Given a 5D line element $dS^2 = g_{AB}(x^\gamma, l)\, dx^A dx^B \, (A, B = t, xyz, l)$ it is then merely a question of algebra to show that the equations $R_{AB} = 0$ contain the ones $G_{\alpha\beta} = T_{\alpha\beta}$ named after Einstein. (In accordance with comments about the non-fundamental nature of the constants, and common practice, we in this section choose units which render $8\pi G/c^4$ equal to unity.) Many exact solutions of $R_{AB} = 0$ are now known (see Wesson 2006 for a catalog). Of these, special mention should be made of the "standard" 5D cosmological ones due to Ponce de Leon, and the 1-body and other solutions in the "canonical" coordinates introduced by Mashhoon et al. It says something about the divide between physics and mathematics that the connection between these solutions and the CM theorem was only made later, by the aforementioned work of Tavakol et al. Incidentally, these workers also pointed out the implications of the CM theorem for lower-dimensional $(N < 4)$ gravity, which some researchers believe to be relevant to the quantization of this force.

The CM theorem, which we will reprove below, is a *local* embedding theorem. It cannot be pushed towards solving problems which are the domain of (more difficult) *global* embeddings. This implies that the CM theorem should not be applied to initial-value problems or situations involving singularities. It is a modest – but still very useful – result, whose main implication is that we can gain a better understanding of matter in 4D by looking at the field equations in 5D.

The CM theorem in succinct form says: "Any analytic Riemannian space $V_n\,(s,t)$ can be locally embedded in a Ricci-flat Riemannian space $V_{n+1}\,(s+1,t)$ or $V_{n+1}\,(s,t+1)$."

We are here using the convention that the "small" space has dimensionality n with coordinates running 0 to $n-1$, while the "large" space has dimensionality $n+1$ with coordinates running 0 to n. The total dimensionality is $N=1+n$, and the main physical focus is on $N=5$.

The CM theorem provides a mathematical basis for the induced-matter theory, wherein matter in 4D as described by Einstein's equations $G_{\alpha\beta}=T_{\alpha\beta}$ is derived from apparent vacuum in 5D as described by the Ricci-flat equations $R_{AB}=0$ (Wesson 1999, 2006). The main result is that the latter set of relations satisfy the former set if

$$T_{\alpha\beta}=\frac{\Phi_{,\alpha;\beta}}{\Phi}-\frac{\varepsilon}{2\Phi^2}\left\{\frac{\Phi_{,4}g_{\alpha\beta,4}}{\Phi}-g_{\alpha\beta,44}+g^{\lambda\mu}g_{\alpha\lambda,4}g_{\beta\mu,4}\right.$$
$$\left.-\frac{g^{\mu\nu}g_{\mu\nu,4}g_{\alpha\beta,4}}{2}+\frac{g_{\alpha\beta}}{4}\left[g^{\mu\nu}_{,4}g_{\mu\nu,4}+\left(g^{\mu\nu}g_{\mu\nu,4}\right)^2\right]\right\}.$$

Here the 5D line element is $dS^2=g_{\alpha\beta}\,(x^\gamma,l)\,dx^\alpha dx^\beta+\varepsilon\Phi^2\,(x^\gamma,l)\,dl^2$, where $\varepsilon=\pm1$, a comma denotes the ordinary partial derivative and a semicolon denotes the ordinary 4D covariant derivative. Nowadays, it is possible to prove Campbell's theorem using the ADM formalism, whose lapse-and-shift technique has been applied extensively to derive the energy of 5D solutions. It is also possible to elucidate the connection between a smooth 5D manifold (as in induced-matter theory) and one containing a singular surface (as in membrane theory). We now proceed to give an ultra-brief account of this subject.

Consider an arbitrary manifold Σ_n in a Ricci-flat space V_{n+1}. The embedding can be visualized by drawing a line to represent Σ_n in a surface, the normal vector n^A to it satisfying $n\cdot n\equiv n^A n_A=\varepsilon=\pm1$. If $e^A_{(\alpha)}$ represents an appropriate basis and the extrinsic curvature of Σ_n is $K_{\alpha\beta}$, the ADM constraints read

$$G_{AB}n^A n^B=-\frac{1}{2}\left(\varepsilon R^\alpha_\alpha+K_{\alpha\beta}K^{\alpha\beta}-K^2\right)=0$$
$$G_{AB}e^A_{(\alpha)}n^B=K^\beta_{\alpha;\beta}-K_{,\alpha}=0.$$

These relations provide $1+n$ equations for the $2\times n\,(n+1)\,/\,2$ quantities $g_{\alpha\beta}$, $K_{\alpha\beta}$. Given an arbitrary geometry $g_{\alpha\beta}$ for Σ_n, the constraints therefore form an underdetermined system for $K_{\alpha\beta}$, so infinitely many embeddings are possible.

This demonstration of Campbell's theorem can easily be extended to the case where V_{n+1} is a de Sitter space or anti-de Sitter space with an explicit cosmological constant, as in some applications of brane theory. Depending on the application, the remaining $n\,(n+1)-(n+1)=\left(n^2-1\right)$ degrees of freedom may be removed by imposing initial conditions on the geometry, physical conditions on the matter, or conditions on a boundary.

The last is relevant to brane theory with the Z_2 symmetry, where $dS^2 = g_{\alpha\beta}(x^\gamma, l)\, dx^\alpha dx^\beta + \varepsilon dl^2$ with $g_{\alpha\beta} = g_{\alpha\beta}(x^\gamma, +l)$ for $l \geq 0$ and $g_{\alpha\beta} = g_{\alpha\beta}(x^\gamma, -l)$ for $l \leq 0$ in the bulk. Non-gravitational fields are confined to the brane at $l = 0$, which is a singular surface. Let the energy–momentum in the brane be represented by $\delta(l) S_{AB}$ (where $S_{AB} n^A = 0$) and that in the bulk by T_{AB}. Then the field equations read $G_{AB} = \kappa[\delta(l) S_{AB} + T_{AB}]$ where κ is a 5D coupling constant. The extrinsic curvature discussed above changes across the brane by an amount $\Delta_{\alpha\beta} \equiv K_{\alpha\beta}(\Sigma_{l>0}) - K_{\alpha\beta}(\Sigma_{l<0})$ which is given by the Israel junction conditions. These imply

$$\Delta_{\alpha\beta} = -\kappa \left(S_{\alpha\beta} - \frac{1}{3} S g_{\alpha\beta} \right).$$

But the $l = 0$ plane is symmetric, so

$$K_{\alpha\beta}(\Sigma_{l>0}) = -K_{\alpha\beta}(\Sigma_{l<0}) = -\frac{\kappa}{2} \left(S_{\alpha\beta} - \frac{1}{3} S g_{\alpha\beta} \right).$$

This result can be used to evaluate the 4-tensor

$$P_{\alpha\beta} \equiv K_{\alpha\beta} - K g_{\alpha\beta} = -\frac{\kappa}{2} S_{\alpha\beta}.$$

However, $P_{\alpha\beta}$ is actually identical to the 4-tensor $\left(g_{\alpha\beta,4} - g_{\alpha\beta} g^{\mu\nu} g_{\mu\nu,4}\right) / 2\Phi$ of induced-matter theory, where it figures in 4 of the 15-field equations $R_{AB} = 0$ as $P^{\beta}_{\alpha;\beta} = 0$ (Wesson 1999). That is, the conserved tensor $P_{\alpha\beta}$ of induced-matter theory is essentially the same as the total energy–momentum tensor in Z_2-symmetric brane theory. Other correspondences can be established in a similar fashion.

Thus while induced-matter theory and membrane theory are often presented as alternatives, they are in fact the same thing, and from the viewpoint of differential geometry both are rooted in the CM theorem. This theorem also has the wider implication that, given the physics in a given manifold, we can always derive the corresponding physics in a manifold of plus-or-minus one dimension. In other words, Campbell's theorem provides a kind of ladder which enables us to go up or down between manifolds of different dimensionality.

1.5 Summary

Dimensions are a delightful subject with which to dally, but we should remind ourselves that they need the cold scrutiny of common sense to be useful. This means, among other things, that we should have a physical identification of the extra coordinates, in order to understand the implications of their associated dimensions. In 5D, we have seen that the extra coordinate can profitably be related to rest mass, either as the Schwarzschild radius or the Compton wavelength, in the classical and quantum domains respectively. This implies that the fifth dimension is related to a scalar field, which is presumably the classical analog of the Higgs field by which particles acquire mass in quantum field theory. This interpretation depends on a judicial use of the fundamental constants (section 2). This approach owes much to the work of Eddington, who delved deeply into the meaning of the equations of physics (section 3). Our usage of dimensions also owes something to Campbell, whose theorem in its modern form shows how to go between manifolds whose dimensionality differs by one (section 4). The use of dimensions may seem in some respects to resemble a game of chess. But to be of practical importance, we need to ascribe the appropriate physical labels to the coordinates and the spaces. This requires skill.

Acknowledgements The views expressed above have been formed over the years by many colleagues, who include P. Halpern, the late F. Hoyle, J. Leslie, B. Mashhoon and R. Tavakol. This work was supported in part by N.S.E.R.C.

References

Barrow, J.D., 1981. Quart. J. Roy. Astron. Soc. **22**, 388.
Barrow, J.D., Tipler, F.J., 1986. The Anthropic Cosmological Principle. Oxford University Press, New York.
Campbell, J.E., 1926. A Course of Differential Geometry. Clarendon Press, Oxford.
Eddington, A.S., 1935. New Pathways in Science. Cambridge University Press, Cambridge.
Eddington, A.S., 1939. The Philosophy of Science. Cambridge University Press, Cambridge.
Green, M.B., Schwarz, J.H., Witten, E., 1987. Superstring Theory. Cambridge University Press, Cambridge.
Gubser, S.S., Lykken, J.D., 2004. Strings, Branes and Extra Dimensions. World Scientific, Singapore.
Halpern, P., 2004. The Great Beyond: Higher Dimensions, Parallel Universes, and the Extraordinary Search for a Theory of Everything. Wiley, Hoboken, New Jersey.
Kilmister, C.W., 1994. Eddington's Search for a Fundamental Theory. Cambridge University Press, Cambridge.
McCrea, W.H., Rees, M.J. (eds.), 1983. Phil. Trans. Roy. Soc. (London) A **310**, 209.
Petley, B.W., 1985. The Fundamental Constants and the Frontier of Measurement. Hilger, Bristol.
Price, K., French, S. (eds.), 2004. Arthur Stanley Eddington: Interdisiplinary Perspectives. Centre for Research in the Arts, Humanities and Social Sciences (10–11 March), Cambridge.
Seahra, S.S., Wesson, P.S., 2003. Class Quant. Grav. **20**, 1321.
Szabo, R.J., 2004. An Introduction to String Theory and D-Brane Dynamics. World Scientific, Singapore.
Wesson, P.S., 1978. Cosmology and Geophysics. Hilger/Oxford University Press, New York.
Wesson, P.S., 1992. Space Science Reviews, **59**, 365.
Wesson, P.S., 1999. Space, Time, Matter. World Scientific, Singapore.
Wesson, P.S., 2006. Five-Dimensional Physics. World Scientific, Singapore.
West, P., 1986. Introduction to Supersymmetry and Supergravity. World Scientific, Singapore.

Chapter 2
Some Remarks on the Space-Time of Newton and Einstein

Graham Hall

Abstract This paper presents an informal discussion of the space and time of classical Newtonian theory and of the space-time of Einstein's special relativity theory, together with a comparison of them. The essential reason for the (1+3)-dimensionality of classical theory and the 4D of special relativity is described.

2.1 Introduction

This paper is an attempt to lay down the foundations of special relativity in an informal and reasonably accessible way and in a manner that facilitates comparison with the classical theory of Galileo and Newton. Particular attention is paid to its 4D. It is not intended as as philosophical discussion and the approach ignores such questions as the *nature* of space or time. This omission could be interpreted either as the tacit assumption that space and time are sufficiently well behaved in the usual intuitive sense or, what more or less amounts to the same thing, that space and time are assumed to have those nice properties that lend themselves to a coherent mathematical treatment at the classical level. The paper is organised in the following way. In section 2, some remarks are given regarding the geometry and topology of space and time. These include a discussion of Hilbert's axioms for Euclidean geometry and some remarks on the topology of the plane. In section 3, the mechanics of Galileo and Newton will be described and the concepts of absolute space and time investigated. This will include a brief encounter with vector bundles, and their use in discriminating between Greek and Classical physics, and with the Galilean group of transformations. In section 4, Einstein's special theory of relativity will be introduced and compared and contrasted with the classical physics of the previous section. Here, the Lorentz group will be described and compared with the Galilean group. In section 5, the 4D approach to special relativity initiated by Minkowski and others will be indicated. This leads naturally to the incorporation of particle mechanics and Maxwell's electromagnetic theory into special relativity.

V. Petkov (ed.), *Relativity and the Dimensionality of the World*, 13–29.

2.2 Geometry

Both classical Newtonian physics and special relativity are dominated by 3D Euclidean geometry and thus some discussion of this discipline should be given here. Until about 150 years ago, Euclid's geometry was the "only" geometry and the writings of Kant had not helped this situation. This changed with the discovery of non-Euclidean geometry, independently and almost simultaneously, by the Russian, Nickolai Lobachevski, and the Hungarian, Janos Bolyai. (Lobachevski published first.) However, Euclidean geometry had essentially been a "visual" science, in spite of Euclid's attempts to describe it axiomatically, and this was still true until the close of the nineteenth century. Thus, Euclid's geometry was always associated with $\mathbb{R}^3$ together with its "straight" lines and "planes"(or $\mathbb{R}^2$ and its straight lines for the case of plane geometry). Then the work of David Hilbert [1] revolutionised the situation. He was able to axiomatise Euclid's geometry in a way that revealed its true identity. For Hilbert, Euclidean geometry was simply a collection of sets controlled by certain conditions (axioms) and which, in fact, lead uniquely (up to isomorphism, in a way that will be described more precisely later) to the standard Euclidean model. The freeing of the initial assumptions (axioms) from visual reference to $\mathbb{R}^3$ (or $\mathbb{R}^2$) laid this geometry bare and emancipated it from the prejudices of the past. In the next few paragraphs, a brief account of this axiomatisation will be given. It is beautifully intuitive and will be given for the 2D Euclidean plane for simplicity; the extension to the 3D case involving no further novelties. The actual exposition here is in a modern form and differs a little from the original.

Hilbert's axioms for 2D Euclidean geometry start with two sets, labelled $\mathcal{P}$ and $\mathcal{L}$ (and thought of as "points" and "lines", respectively, but only for ease of reference) and an *incidence* relation denoted by $\mathcal{I}$, so that $\mathcal{I}$ is a subset of $\mathcal{P} \times \mathcal{L}$. If $p \in \mathcal{P}$ and $L \in \mathcal{L}$ and $(p, L) \in \mathcal{I}$ then one says that p is *incident with* L or, more simply, p is *on* L. There then follow five groups of axioms.

The *first* group (the axioms of *incidence*) simply say that any $L \in \mathcal{L}$ has at least two distinct members of $\mathcal{P}$ incident with it, that given two distinct members of $\mathcal{P}$ there is a unique member $L \in \mathcal{L}$ with which they are each incident and that given any $L \in \mathcal{L}$ there is at least one $p \in \mathcal{P}$ which is not incident with L. These axioms immediately show that there exists an *injective* map $f : \mathcal{L} \to$ the power set of $\mathcal{P}$ which maps $L \in \mathcal{L}$ to the subset of $\mathcal{P}$ consisting of all points incident with L. Thus, one can ease the mental burden by making the natural interpretation of members of $\mathcal{L}$ as subsets of ("lines" in) $\mathcal{P}$.

The *second* group of axioms, sometimes called the "betweenness" axioms, introduces a subset $\mathcal{B}$ of $\mathcal{P} \times \mathcal{P} \times \mathcal{P}$ so that if $(p, q, r) \in \mathcal{B}$, there exists $L \in \mathcal{L}$ such that p, q and r are incident with L, i.e. $p, q, r \in L$ (and one says that p, q, and r are *collinear*) and which are subject to a series of axioms designed so that one can sensibly say that q is between p and r but that p is *not* between q and r and that r is *not* between p and q. This is the axiomatic way of ruling out "circular" lines. These axioms also enable one to make a sensible definition about a pair of points in $\mathcal{P}$ being on the "same" or "opposite" sides of a line $L \in \mathcal{L}$.

The *third* group of axioms are the axioms of *congruence*. Before these are introduced it is first remarked that at this stage one can, for $p, q \in \mathcal{P}$, define the *segment* $[p, q]$, which consists of the points p and q and all points on the (unique) line containing p and q and which lie "between" p and q, and the *ray* $\overrightarrow{pq}$ which consists of the segment $[p, q]$ together with all points r such that $(p,q,r) \in \mathcal{B}$. A second remark is that it is now possible to exhibit the "2-dimensionality" of this system in a natural and intuitive way. The axioms of congruence are divided into two sections; the former is an equivalence relation on segments in such a way that the members of one of the resulting equivalence classes are regarded as having the same "length" whilst the latter uses rays to construct a similar equivalence relation for "angular" measure.

So far the axioms have been beautifully intuitive. But now, a less obvious *fourth* group (in fact a single axiom) is introduced, that of *completeness*. This is brought into play to ensure that the "lines" in the model become copies of $\mathbb{R}$. At this point one can prove that if $p \in \mathcal{P}$ and $L \in \mathcal{L}$ and (p, L) is *not* in $\mathcal{I}$, then there exists $L' \in \mathcal{L}$ such that $(p, L') \in \mathcal{I}$ and $L \cap L' = \emptyset$. The existence of L' is, of course, only part of the parallel axiom; one requires its uniqueness also. The impossibility of the proof of this uniqueness at this stage is one of the beauties of this approach, because the geometry of Lobachevski and Bolyai can be shown to satisfy each of the axioms so far given and in this geometry there would be infinitely many such lines L'. However, it turns out that the model of Lobachevski and Bolyai and the Euclidean model are the *only* models satisfying the axioms thus far and so Euclidean geometry is then uniquely arrived at by the last (*fifth*) axiom group (again a single axiom, the *parallel* axiom) which declares the uniqueness of the above line L'. Thus, if the triplet $(\mathcal{P}, \mathcal{L}, \mathcal{I})$ satisfies all of the axioms so far laid down, and given the usual Euclidean structure on $\mathbb{R}^2$, there exists a bijective map $f : \mathcal{P} \to \mathbb{R}^2$ such that f maps members of $\mathcal{L}$ to "straight" lines in $\mathbb{R}^2$ and preserves the concepts of betweenness and congruence as given axiomatically for $(\mathcal{P}, \mathcal{L}, \mathcal{I})$ and as usually understood in $\mathbb{R}^2$. Such a map is called an *isomorphism*.

Now consider two copies of $\mathbb{R}^2$ and let the first copy be endowed with the standard Euclidean structure with its set of lines being determined by linear relations between the usual coordinates x and y and its incidence set $\mathcal{I}$ by the set membership criteria of points lying on lines (as described earlier at the end of the third paragraph of this section). Such a structure on $\mathbb{R}^2$ (or its equivalent on $\mathbb{R}^3$) will be said to be *compatible with* $\mathbb{R}^2$ (or $\mathbb{R}^3$). Now let f be *any* bijection from $\mathbb{R}^2$ to itself. Then f can be used to carry the Euclidean structure (including length and angle, once appropriate units have been established in the standard structure) of the first copy of $\mathbb{R}^2$ over to the second copy, in an obvious way, to give another Euclidean structure on $\mathbb{R}^2$ which will, in general, be different from the first (usual) one. However, mathematically, they are indistinguishable. How is it that we attach so much importance to the former? Physics may be brought in here in the form of, for example, the motion of free particles or light beams over a horizontal surface, or the lines of shortest distance, where this shortest distance is measured by, say, "minimal fuel consumption" for a motor vehicle moving over such a surface, for experience suggests that the original "straight line" structure will be recovered by such experiments. This observation will be incorporated into the assumptions for classical and special relativistic mechanics later.

Regarding the topology of the Euclidean plane, one may choose the natural topology that arises from the "usual metric" on $\mathbb{R}^2$ and which, in turn arises naturally from the geometrical axioms (up to a choice of units). This would, of course, yield the "standard" topology on $\mathbb{R}^2$. Also, if one regards $\mathbb{R}^2$ as a 2D manifold in the usual way, there is a natural manifold topology on $\mathbb{R}^2$ which is again the standard topology. Another possibility, bearing in mind the fact that the lines in the geometry are "primitive" elements, and turn out to be copies of the set $\mathbb{R}$, is to consider a topology for $\mathbb{R}^2$ which induces, as subspace topology, the standard real line topology on each line in $\mathbb{R}^2$. Clearly the standard topology for $\mathbb{R}^2$ does just this but it is not the only topology for $\mathbb{R}^2$ with this property. To see this consider a sequence $\{x_n\}$ which converges (in the standard topology on $\mathbb{R}^2$) to the point $(0,0)$ but is such that, for no positive integer n, is x_n equal to $(0,0)$ and no three points of $\{x_n\}$ are collinear. Then consider the collection of subsets of $\mathbb{R}^2$ given by the open sets in the standard topology together with the set U, which is the set $\mathbb{R}^2$ with the points of the above sequence removed. Taking this collection of subsets of $\mathbb{R}^2$ as a *subbasis* for a topology on $\mathbb{R}^2$ (i.e. the smallest topology on $\mathbb{R}^2$ containing these sets), this topology is different from (in fact, finer than) the standard one. This is because the point $(0,0)$ is not on the sequence and so U is not open in the standard topology on $\mathbb{R}^2$. To see this just note that $(0,0) \in U$ but no standard open ball containing $(0,0)$ is contained in U. However, this new topology (which is connected) clearly induces the standard topology of $\mathbb{R}$ on each line. This topology does not, however, reflect the homogeneity property satisfied by the Euclidean structure on $\mathbb{R}^2$. Given the importance of coordinatisation, it is usual to regard space or space and time as a manifold and so the manifold (that is the usual) topology for these spaces will be assumed.

2.3 Classical Mechanics

In Greek physics, a concept of time similar to the absolute time of Newtonian theory seems to have been assumed. Also, a preferred state of rest (with respect to the earth) was taken as "natural". Thus, observers (that is, coordinate systems) were at rest on earth. Within this theory, this is interpreted to mean that there are observer-independent concepts of *fixed points* in space and *simultaneity of events*. In this sense, space and time were absolute. To enable a precise description of the universe E of events one assumes that all time instants can be represented by the set T, assumed equal to $\mathbb{R}$, and that the *simultaneity* (or *instantaneous*) spaces $S(t)$ for each time $t \in T$ may, because of the fixed point concept, be naturally identified with each other and with some set S (taken here as $\mathbb{R}^3$) representing space. Then $E = S \times T$ with subsets of constant t being simultaneity spaces and subsets of constant $s \in S$ being fixed points enduring through time. To establish a concrete coordinate system on E, the time coordinate may be chosen by assuming that a "good" clock is present at each fixed point in space (By "good" is meant that if two of these clocks are brought together at any time they would be seen to measure the same time coordinate and so are synchronised and measure time using the same units. In this sense, moving the clocks about does not affect their time readings). Then one chooses a space-coordinate system by assuming a bijection $f : \mathbb{R}^3 \to S$, such that S has a Euclidean structure imposed on it in the obvious way through f, from the compatible natural Euclidean structure on $\mathbb{R}^3$ and which is perceived by experiment to be the "natural" Euclidean structure there (see section 2). Finally one has $E = \mathbb{R}^3 \times \mathbb{R} = \mathbb{R}^4$. It is noted how the fixed point assumption has the effect of "gluing together" the spaces $S(t)$ so that E is a product space and that this product structure is independent of the observer and the coordinate systems chosen on T and S.

In Galilean–Newtonian physics, one again has absolute time as described above, represented by T, and regarded as the set $\mathbb{R}$. Thus, if the Universe is taken as the set of events E, one has a map $p : E \to T$ attaching to each event its absolute time coordinate. The nature of absolute time means, just as in the case of Greek physics, that the existence of such a map (but not, of course, the map itself) is independent of the unit and choice of origin of time used. For each time instant $t \in T$, the subset $p^{-1}\{t\}$ is the instantaneous space for time t and is again assumed to be a set $S(t)$ of the same cardinality as $\mathbb{R}^3$ for each t. It is then assumed that on each such copy of $S(t)$ one can put the structure of a Euclidean geometry and a Cartesian coordinate system with coordinates x, y, and z which is compatible with it. The problem then is the "binding together" of these copies of $S(t)$, now regarded as copies of $\mathbb{R}^3$, since the nature of Newtonian–Galilean mechanics rules out the existence of absolute (observer-independent) fixed points. Such a binding is essentially an "observer" or space–time coordinate system (and was achieved in Greek physics by the assumption of the existence of fixed points).

One of the fundamental features (assumptions) of Newtonian theory is the concept of force which prevails in that theory and the theory's assumed ability to distinguish between "real" and "inertial" (or "accelerative") forces. The next assumption is the existence of a special collection of observers, called *inertial observers*, for whom the above space-time coordinate system can be chosen in such a way that the path of a "free" particle (one upon which no "real" forces act) satisfies the conditions that the functions $x(t)$, $y(t)$, and $z(t)$ are at least twice differentiable and that $dx/dt, dy/dt$, and dz/dt are all constants and also that any path satisfying these conditions is the potential path of such a particle. The triple $(dx/dt, dy/dt, dz/dt)$ is called the *velocity* of the particle at time t in this frame. In this sense the Galilean law of inertia (Newton's first law) is true by the very construction of inertial observers and gives rise, within any particular inertial frame, to the concept of a special type of free particle in that frame called a "fixed point" or a "point at rest" (a path where x, y, and z are each constant in that frame). Any free particle is a fixed point for some inertial observer. Of course, there are many such inertial observers depending on the binding together of the spaces $p^{-1}\{t\}$ (subject to the above conditions) and the relationship between them will be discussed later. Thus each inertial observer may view the universe E as a product space, as in the Aristotelian case, but there is no such *natural* product, since such a product is observer dependent. In this sense there is no unique absolute space (as in Greek physics) but many absolute spaces, but from the dynamical viewpoint they are still absolute. In mathematics, one thus has a *vector bundle* (E,T,p) with total space E, base space T and projection p and with local (in fact, global) trivialisations being the above product spaces provided by each inertial observer using the base space T (with the usual real-line topology) and his "identified" fibres $p^{-1}\{t\}$, each taken as the vector space $\mathbb{R}^3$ with its usual topology. A given inertial observer thus uses his "points at rest" to construct three independent global vector fields (cross sections) over T to achieve the global trivialisation of E. The paths of free particles pick out those lines (and planes) defined by linear relations between his Cartesian coordinates and hence, in this sense, fix the usual Euclidean structure on each set $p^{-1}\{t\}$. It is, of course, also assumed that, from the point of view of any such inertial observer (that is, in his instantaneous space $p^{-1}\{t\}$), any two points and any two "directions" (1D subspaces) are "equivalent" from the physical viewpoint (the *homogeneity* and *isotropy* assumptions). It is understood that no experiment using particles (subject to the above laws) can distinguish one inertial frame from another. This is the *Newtonian principle of relativity*.

The relationship between two inertial frames in classical Newtonian theory, that is the transformations relating the above trivialisations, can now be found. Let I and I' represent two inertial observers with coordinates (from their product representation of E) given by (x_1,x_2,x_3,t) and (x'_1,x'_2,x'_3,t), and spatial origins represented by their fixed points O and O', respectively. For I choose an orthonormal basis $(\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3)$ for his copy of $\mathbb{R}^3$ which are directed along his x_1- , x_2-, and x_3-axes from O and choose, similarly, a basis $(\mathbf{e}'_1,\mathbf{e}'_2,\mathbf{e}'_3)$ for I'. These basis vectors are taken, intuitively, to connect fixed points in the respective frames and so are constant vectors in these frames. Let $\mathbf{c}(t)$ denote the vector $\overrightarrow{OO'}$ at time t so that one may write $\mathbf{c}(\mathbf{t}) = \sum c_i\mathbf{e}_i$ for $c_i(t)$. Now suppose that at time t an event Q occurs which has coordinates (x_1,x_2,x_3,t) in I and (x'_1,x'_2,x'_3,t) in I'. One must find the expression for the primed coordinates in terms of the unprimed ones. Here, it will be assumed that the former are differentiable functions of the latter. Then $\sum x_i\mathbf{e}_i = \sum x'_i\mathbf{e}'_i + \sum c_i\mathbf{e}_i$. Now, since the bases are orthonormal, there exists an orthogonal matrix $\mathbf{A}$ such that $\mathbf{e}'_\mathbf{i} = \sum A_{ji}\mathbf{e}_\mathbf{j}$. Thus one finds that if $\mathbf{x}(t) = (x_1(t),x_2(t),x_3(t))$ and similarly for $\mathbf{x}'$,

$$\mathbf{x} = \mathbf{c}+\mathbf{A}\mathbf{x}' \qquad \mathbf{x}' = \mathbf{A}^{-\mathbf{1}}(\mathbf{x}-\mathbf{c}) \tag{2.1}$$

Equation (2.1) shows the instantaneous connection between $\mathbf{x}(t)$ and $\mathbf{x}'(t)$ at time t. As time changes, one must allow for the possibility that A and $\mathbf{c}$ may change with time. If one now uses the fact that I and I' are inertial so that if a particle represented by a succession of events such as Q above has constant $d\mathbf{x}/dt$, then $d\mathbf{x}'/dt$ is also constant in I', then a standard argument shows that $\mathbf{A}$ is a constant orthogonal matrix and that there exists constant vectors $\mathbf{v}$ and $\mathbf{p}$ such that $\mathbf{c} = \mathbf{v}t+\mathbf{p}$. Thus, the transformation between I and I' is determined by the orthogonal matrix A, and the vectors $\mathbf{v}$ and $\mathbf{p}$ and amounts to $3+3+3=9$ free parameters with A representing the "constant" orientation of the space coordinates of I with respect to those of I', $\mathbf{v} = d\mathbf{c}/dt$ representing the constant velocity of O' with respect to O and $\mathbf{p} = \mathbf{c}(0)$ representing the original vector displacement of O' from O. The collection of such transformations is the set of *Galilean transformations* (without time scalings or time translations). They can be represented by the scheme

$$\begin{pmatrix}\mathbf{x}\\ t\end{pmatrix} \longrightarrow \begin{pmatrix}\mathbf{x}\\ t\\ 1\end{pmatrix} \longrightarrow \begin{pmatrix}\mathbf{x}'\\ t\\ 1\end{pmatrix} = \begin{pmatrix}A^{-1} & -A^{-1}\mathbf{v} & -A^{-1}\mathbf{p}\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}\mathbf{x}\\ t\\ 1\end{pmatrix} \longrightarrow \begin{pmatrix}\mathbf{x}'\\ t\end{pmatrix} \tag{2.2}$$

A natural group property follows for the Galilean transformations by representing the above transformation by the 5×5 matrix

$$\begin{pmatrix}A^{-1} & -A^{-1}\mathbf{v} & -A^{-1}\mathbf{p}\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}$$

Then the combination of two such transformations is found by simply identifying their representative matrices, multiplying them together and reconstructing the transformation from the product (representative) matrix. For the special case that each transformation satisfies $\mathbf{p} = 0$ one achieves the subgroup of Galilean transformations (without time scalings or time or space translations) for which six free parameters are available and whose representative matrices can be taken as the 4×4 matrices $\begin{pmatrix}A^{-1} & -A^{-1}\mathbf{v}\\ 0 & 1\end{pmatrix}$ in an obvious way. These two Galilean groups are closed subgroups of the Lie groups $GL(5,\mathbb{R})$ and $GL(4,\mathbb{R})$, respectively, and so are each Lie groups (of dimensions nine and six, respectively) whose Lie algebras are easily found. In the latter case (the homogeneous Galilean group), the transformations in Eq. (2.1) reduce to

$$\mathbf{x}' = A^{-1}(\mathbf{x} - \mathbf{v}t) \tag{2.3}$$

The important and instructive special case when I and I' are in *standard configuration* occurs when A is the unit 3×3 unit matrix (so that the $x-, y-$ and $z-$ axes in I are parallel, respectively, to the $x'-, y'-$ and $z'-$ axes in I') and when $\mathbf{v}$ has components $(v,0,0)$ in I, $v \in \mathbb{R}$ (so that the space origin of I' moves along the x-axis in I with speed v). Thus one gets; $x' = x - vt$, $y' = y$, $z' = z$ (and $t' = t$). It is noted here that one may speak of the respective axes of I and I' being "parallel" since they are lines in the same Euclidean space of simultaneity.

Newtonian theory is then supplemented with Newton's second law which describes the path of a particle *in an inertial frame* when some force acts upon it. Here the particle has mass m (assumed constant) and the force is taken as a "real" force (assumed identifiable in Newtonian theory) and described in an inertial frame I and at a certain event E by the triplet $\mathbf{f} = (f_1, f_2, f_3)$ and which, of course, depends on E. This law is then written as

$$\mathbf{f} = m\left(\frac{d^2x}{dt^2}, \frac{d^2y}{dt^2}, \frac{d^2z}{dt^2}\right) \tag{2.4}$$

and is consistent with the Galilean law in the case when $\mathbf{f} = 0$ for free particles. The components of $\mathbf{f}$ change, under a change of inertial frame, in such a way as to be consistent with Eqs. (2.3) and (2.4). Thus if, at E, the force in I' is the triple $\mathbf{f}'$, then $\mathbf{f}' = A^{-1}\mathbf{f}$. Any deviation in the path of a particle from the uniform motion described by Galileo's law of inertia would, in an inertial frame, be interpreted as due to a real force. In this sense, inertial observers experience only real forces. Of course, since it appears that one cannot shield any particle from a gravitational force, it is clear that free particles cannot exist and so inertial observers must be regarded as an extrapolation from experiments made, in the 2D case, on horizontal surfaces (where gravity is assumed to be neglected) or from "armchair" experiments made on particles "far away" from gravitational fields. Modern cosmology denies the existence of this latter haven of escape but allows the possibility of regions of "weak" gravitational fields.

It is clear that the foundations of Newtonian theory are easily attacked. Yet, its success is undoubted. The concept of Newtonian absolute space (or absolute spaces) would appear to be the main problem and this was rightly criticised by the philosophers (see, for example [2] and [3]). The assumption of absolute space can be viewed in several different ways. It can be taken as the assumption of the existence of inertial observers (and the points at rest enduring through time for each such inertial observer then become marker points for an "absolute space", for that observer) or the essentially equivalent assumption of the ability to distinguish real forces from inertial ones. The multiplicity of these absolute spaces may be taken as a weakening of its absoluteness, but the facts that inertial observers exist and that real and inertial forces remain distinguishable in this theory are the main problem. Furthermore, mechanics cannot live in isolation and with the advent of the electromagnetic theory more serious problems arose. The so-called ether, postulated in the latter theory, or equivalently, the existence of a "preferred speed", that of light, and which was achieved by light (or any other electromagnetic wave) only in a frame at rest with respect to the ether, immediately gave a preferred inertial (ether) frame. Of course, there is no immediate contradiction provided the ether is part of physics and hence measurable. It would simply restrict the Newtonian principle of relativity to mechanical phenomena. This preferred speed, usually denoted by c, occurs naturally as a constant in Maxwell's equations and suggested the preference of the ether frame when electromagnetic phenomena are considered. Of course, this would reduce the mathematical structure of Newton's theory, when the electromagnetic theory is included, to the structure of Greek physics, with the ether frame playing the role of rest frame. However, the ether resisted many optical attempts to measure it, in particular, the celebrated experiment of Michelson and Morley in 1887. In 1899 Poincaré was of the opinion that optical experiments depended only on the relative velocity of the relevant bodies and, indeed, that absolute motion was undetectable by any means (for further details of this and other developments in this period, see [4]). This is, of course, a much stronger statement than the Newtonian principle of relativity and it is, in fact, not clear that the latter is acceptable as a statement in physics, since the existence of an experiment which does not involve electromagnetic phonomena is doubtful. Poincaré, in 1904, and Einstein, in 1905, enunciated what is now known as the *principle of relativity* and which states that no experiment whatsoever can distinguish one inertial frame from another. Hence, physical laws should reflect this principle in that their fundamental equations should be formally the same in all inertial frames. The clearest statement and consequent development of this idea came in 1905 with Einstein's fundamental paper on special relativity [5].

2.4 Special Relativity

In special relativity, the rejection of absolute time makes the starting point for an informal discussion of its foundations quite different. One assumes, first, as in Newtonian theory, that one can distinguish between "real" and "inertial" forces. Then one assumes the existence of observers (also called *inertial*) which are assumed to satisfy the following conditions. For such an observer, each point in the universe of events E can be designated a time coordinate t, with t drawn from the set of real numbers, such that for each t, (a) the subset $S(t)$ of points of E with a particular time coordinate $t \in \mathbb{R}$ is of the same cardinality as $\mathbb{R}^3$, (b) each set $S(t)$ can be given the structure of a Euclidean geometry together with a compatible Cartesian coordinate system such that Galileo's law of inertia now holds for free particles for this observer and with these space and time coordinates. Thus the path $\mathbf{x}(t)$ of such a free particle satisfies $d^2\mathbf{x}/dt^2 = 0$. Again, the assumptions of homogeneity and isotropy for the sets $S(t)$ are made. But now, an extra assumption is required; if a photon of light travels from an event P with time coordinate t_1 to an event Q with time coordinate t_2, it does so with a path $\mathbf{x}(t)$, which satisfies $d^2\mathbf{x}(t)/dt^2 = 0$ *and* if, according to the Cartesian coordinates given in (b), the distance PQ is d then $d = c(t_1 - t_2)$ where c is a constant which is *independent of the observer in question*. This extra restriction on the space and time–time coordinates is the critical difference between special relativity and Newtonian theory because, clearly, c is interpreted as the speed of light and this speed becomes independent of the inertial observer. Of course, in Newtonian theory, the addition laws for velocity show that no speed can have such a property.

The main point here is that absolute time is abandoned and each observer has its own "time" coordinate. Thus, whereas the time coordinates of the different inertial observers in classical (i.e. Newtonian) theory may be set equal to each other by a judicious synchronisation (and which leads to the bundle structure described in that theory), no such synchronisation (and no such bundle theory) is assumed in special relativity. The reason for the assumption regarding c may be argued, as stated above, from the occurrence of the constant c in Maxwell's equations or, by appeal to the experimental results of Michelson and Morley, which appear to contradict the Newtonian theory (or, at least, the addition laws for velocity that hold in that theory) but are consistent with this extra assumption in Einstein's special relativity theory. The rejection of absolute time means that, perhaps, the best one can say is that each inertial observer coordinatises space and time and that the universe should be viewed as the manifold $\mathbb{R}^4$ with each inertial observer supplying a coordinate system to its atlas of charts (but with none of these charts regarded in any way as preferred). This is, of course, analogous to the local trivialisations provided in Newtonian theory by inertial observers. Thus, although Newtonian theory can be regarded as a 4D (space and time) theory, it reduces to a vector bundle over a 1D base space with 3D fibres (observer-independent simultaneity sections). This reduction, caused by the absolute nature of time, does not occur in special relativity, (where the simultaneity sections are observer-dependent) and one must accept the 4D of the theory, together with the condition regarding the constancy of the speed of light.

The acceptance of special relativity as taking place on the 4D manifold $\mathbb{R}^4$ and for which each inertial observer contributes a global chart, now poses the question of the nature of the general transformations that arise on these overlapping charts. The concepts of homogeneity and isotropy used in Newtonian theory are taken over here without change, but the stronger principle of relativity described above will be assumed. A chart for such an observer will be supposed chosen such that the last coordinate is the time coordinate t of the event in question for that observer and the first three coordinates are the Cartesian coordinates of the event in the "space" of constant t. The general transformations required must have the following fundamental property. Let I and I' be inertial frames with the property that the events with coordinates (0, 0, 0, 0) in I and I' coincide (and they will jointly be called, O). Further, let P be an event in E with coordinates (x^1,x^2,x^3,t) in I and coordinates (x'^1,x'^2,x'^3,t') in I' (so that x, y and z have been replaced by x^1, x^2, and x^3, and similarly for the primed coordinates). Then, since the statement that a photon could pass through P and O is independent of the observer (and defining x^4 to be ct, and x'^4 to be ct'), one finds

$$(x^1)^2+(x^2)^2+(x^3)^2-(x^4)^2=0 \Leftrightarrow (x'^1)^2+(x'^2)^2+(x'^3)^2-(x'^4)^2=0 \qquad (2.5)$$

Thus, the relation between the coordinates (x^1,x^2,x^3,x^4) and the coordinates (x'^1,x'^2,x'^3,x'^4) must be such that the vanishing of the left-hand side of Eq. (2.5) is equivalent to the vanishing of the right-hand side of Eq. (2.5). Thus the "null cone" of points defined by the vanishing of either side of Equation (2.5) is preserved. Of course, more conditions are required to find the actual (homogeneous) transformations that are the special relativistic equivalent of Eq. (2.3). One option is to assume that the expressions for the primed space and time coordinates are *linear* expressions of the unprimed space and time coordinates. Thus, the transformations one is interested in (and recalling the insistence of fixing the space and time origin) are the (homogeneous) *Lorentz transformations* and are those linear maps $\mathbb{R}^4 \to \mathbb{R}^4$ which, from Eq. (2.5), preserve the metric η on $\mathbb{R}^4$ where η is the matrix given by diag $(+1+1+1-1)$. The set of all such transformations, $\mathcal{L}$, is then given by

$$\mathcal{L} = \{A \in GL(4,\mathbb{R}) : A\eta A^T = \eta\} \qquad (2.6)$$

where A^T denotes the *transpose* of A. This is a 6D closed subgroup of the Lie group $GL(4,\mathbb{R})$ and is hence, a Lie group. Its Lie algebra is L where $L = \{A \in M_4\mathbb{R} : A\eta + (A\eta)^T = 0\}$, that is, the Lie algebra of 4×4 matrices which are skew-self adjoint with respect to η.

Most derivations of the Lorentz transformations rely on the constancy of the speed of light (or at least on the existence of some fundamental constant speed). However, an account can be found in [6] which is based on the related assumption of the existence of a *causality* relation (but does not require even the assumption of the continuity or the linearity of the transformations). It is interesting to remark that in [7] no assumption of such a velocity is assumed (but the linearity of the transformations is). It then turns out that such a "limiting" velocity is, of necessity, present (and must, of course, be determined by experiment). Within this approach is the special case of infinite limiting velocity and these are the Galilean transformations shown in Eq. (2.3). It is noted that the Lorentz transformations corresponding to the classical situation when I and I' are in standard configuration can no longer be thought of in terms of parallel space axes, since such axes are not necessarily in the same instantaneous space. In special relativity, such frames I and I' are related by the conditions that the coordinates of any event satisfy $y = 0 \Leftrightarrow y' = 0$, $z = 0 \Leftrightarrow z' = 0$ and $x' = 0 \Leftrightarrow x = vt$ for some $v \in \mathbb{R}$, together with the assumption of the Einstein principle of relativity and the speed of light postulate. The resulting transformations are given by

$$x' = \gamma(v)(x - vt), \qquad y' = y, \qquad z' = z, \qquad t' = \gamma(v)(t - vx/c^2) \tag{2.7}$$

(or their equivalents in terms of the coordinates x^a and $x^{a'}$) where $\gamma(v) = (1 - v^2/c^2)^{-1/2}$. The theoretical link with the corresponding Galilean transformations, as c becomes arbitrarily large or for v/c, small, is clear.

2.5 The 4D Formulation of Special Relativity

Special relativity can be developed as an essentially "3+1"-dimensional theory or, perhaps more accurately, in the manner that one is accustomed to in Newtonian theory. This was, in fact, the standard approach taken in the early days. However, once the full impact of the Lorentz transformations and the associated Lorentz invariance was taken into account, a 4D theory was developed using the ideas of (Lorentz) vectors and tensors on the 4D manifold $\mathbb{R}^4$ equipped with the metric η. In Newtonian theory, the bundle structure resulting from the assumption of absolute space means that the 3D simultaneity slices of constant t are a well-defined part of the space and time structure whereas, in special relativity, they have no such significance, being observer dependent. These special slices in Newtonian theory essentially lead to the "3-dimensionality" of this theory. In special relativity theory, however, a manifestly 4D approach is more elegant. This 4D approach to special relativity is essentially the contribution of Hermann Minkowski. It relies on the description of space-time through (pseudo-orthonormal) inertial frames and so views the universe as $\mathbb{R}^4$ with the metric components always given by $\eta_{ab} = diag(+1,+1,+1,-1)$ (because such coordinate systems are related by Lorentz transformations and these preserve this matrix representation of η). This metric can then be used to raise and lower tensor indices in the usual way in Minkowski geometry.

A particularly illuminating example of the advantages of this 4D approach arises in the study of particle mechanics and electromagnetic theory. There are many approaches to the setting up of these topics in special relativity. Here, a reasonably convenient 4D one will be briefly summarised. Some historical remarks will be given at the end indicating how such developments came about chronologically. First, assuming that the Lorentz transformations are given, one can easily deduce the special relativistic addition laws for velocity for comparing a particle's so-called *3-velocity* in two different inertial frames. Here the 3-velocity of a particle P in a frame I with coordinates (x,y,z,t) is, as in Newtonian theory, given by first representing the particle's *space* path in I, as it evolves in time, by the function $t \to (x(t),y(t),z(t))$ and then defining the 3-velocity to be the triple $\mathbf{u} = (dx/dt, dy/dt, dz/dt)$. Such a description of a particle's path is appropriate in I but may be inconvenient for other inertial frames since the t coordinate will not necessarily be the time coordinate in the other frame (as would be the case in Newtonian theory). Thus, one should attempt to describe the path of a particle in the space and time of special relativity (now referred to as *space-time*) using some convenient parameter for this path which is the same (and meaningful) for all inertial observers. For this reason one introduces Minkowski's concept of the *proper time* τ of a particle P (the time reading on a good clock carried along by the particle) and this turns out to be such a parameter. [Here an extra assumption must be admitted; the time dilation formula, easily derived from the Lorentz transformations for a particle moving with constant velocity, is assumed to hold for arbitrary velocity. This is needed to relate derivatives with respect to t and those with respect to τ.] More precisely, one defines the *world line* of the particle to be the path in space-time $\mathbb{R}^4$ given in some coordinate system by

$$\tau \to (x^1(\tau), x^2(\tau), x^3(\tau), x^4(\tau)) \tag{2.8}$$

The tangent vector to this path is the particle's *4-velocity* U and is a contravariant vector in the space-time $\mathbb{R}^4$. It is written in the usual way, in component form, in the inertial frame I as $U^a \equiv dx^a/d\tau$. Now the 4-velocity U and 3-velocity $\mathbf{u}$ are easily checked to be related (in an obvious notation) by $U^a = \gamma(u)(\mathbf{u}, c)$ where $\gamma(u) \equiv (1-u^2/c^2)^{-1/2}$ and $u \equiv (\mathbf{u}.\mathbf{u})^{1/2}$, with a dot denoting the usual 3D inner product. In I', the 4-velocity has components given by $U'^a = dx'^a/d\tau$ (and on writing down the transformation law for U from I to I', one can achieve an alternative proof of the 3-velocity addition law). Now, in discussing the dynamics of a particle, one must introduce the particle mass. Here, no assumption that this mass is constant is made. Thus, a particle P is assumed to have mass m which may depend on its velocity (actually only on its speed, due to the isotropy assumption). Then one defines the *3-momentum* of P in I, as in Newtonian theory, by the triple $\mathbf{p} \equiv m\mathbf{u}$. and introduces the *conservation of momentum* for particles so that the usual "vector sum" of the triples $m\mathbf{u}$ is componentwise conserved in any frame during a collision. It is convenient to introduce the assumption of Newton's third law at this point, but with care, since it talks about the equality of two "simultaneous" forces (and, of course, simultaneity is not observer independent in special relativity). However, the forces referred to are at the same event and so simultaneity here is not a problem. Then one takes, as in Newtonian theory, the *3-force* on the particle as the triple $\mathbf{f} \equiv d(m\mathbf{u})/dt$. From the assumption of conservation of 3-momentum and the Lorentz transformations relating inertial frames one can deduce that the dependence of the mass m of P on its speed u is given by the *mass relation* $m = \gamma(u)m_0$ where m_0 is a constant ([8], see also [9]). One obviously interprets m_0 as the mass of P in that inertial frame where it is at rest (since $\gamma(0) = 1$), that is, as the *rest mass* of P. Next one defines the *4-momentum* of P as the space-time vector with components $P^a = m_0 U^a$. It can be checked from the assumption of conservation of 3-momentum and the consequent mass relation that the total mass (*not* the total rest mass) is the same before and after a collision and then that each of the four components of P is conserved in a collision in any frame. Conversely, the conservation of the 4-momentum, P, implies the conservation of 3-momentum, the conservation of total mass and the mass relation. [In fact one can reduce the conservation law assumption to a single conservation of a suitably defined energy [10]] One can then introduce the *4-force* $F^a = dU^a/d\tau$. Again, this can be checked to be a space-time vector under Lorentz transformations.

Now in order to accommodate the electromagnetic field into special relativity one assumes that Maxwell's equations hold in each inertial frame and that the charge on a particle is independent of the frame. To include the dynamics of this field, one must realise that Maxwell's theory is a linear theory in the sense that if one adds together two solutions of Maxwell's equations with certain source terms, another solution is obtained whose source term is the sum (in an obvious sense) of the original ones. Thus it follows that Maxwell's equations cannot contain, within themselves, the equations of motion of a charged particle. For example, if one adds together two static solutions of Maxwell's equations in some inertial frame, each representing a single stationary charged particle in that frame, one obtains a solution of Maxwell's equations where the two particles do not interact (one could think of the forces required to keep then fixed as being of non-electromagnetic origin and hence, not included in Maxwell's equations). Dynamics is introduced into Maxwell's theory by assuming the *Lorentz force law* which says that, in any inertial frame, if a particle of mass m, charge e and velocity $\mathbf{u}$ experiences an external electric and a magnetic field of magnitudes $\mathbf{E}$ and $\mathbf{H}$, respectively, in that frame, then the 3-*force* $\mathbf{f}$ on the particle due to these fields is given by

$$\mathbf{f}\left(=\frac{d(m\mathbf{u})}{dt}\right) = e(\mathbf{E} + c^{-1}(\mathbf{u} \times \mathbf{H})) \tag{2.9}$$

Now define a skew-symmetric quantity F with components $F_{ab}(= -F_{ba})$ given in any inertial frame by the matrix

$$\begin{pmatrix} 0 & H_3 & -H_2 & E_1 \\ -H_3 & 0 & H_1 & E_2 \\ H_2 & -H_1 & 0 & E_3 \\ -E_1 & -E_2 & -E_3 & 0 \end{pmatrix} \tag{2.10}$$

where $\mathbf{E} = (E_1, E_2, E_3)$ and $\mathbf{H} = (H_1, H_2, H_3)$ are the component expressions for $\mathbf{E}$ and $\mathbf{H}$ in that frame. The idea is to show that F is, in fact, a tensor (with indices as indicated) under Lorentz transformations. To do this, one first defines, using the Einstein summation convention, the components $G_a \equiv F_{ab}U^b$ where U is the 4-velocity of the particle. Using the expression $U^a = \gamma(u)(\mathbf{u}, c)$ given previously for the 4-velocity in terms of the 3-velocity $\mathbf{u}$, one evaluates G_a, using (2.10), to get

$$G_a = ce^{-1}\gamma(u)(\mathbf{f}, -c^{-1}(\mathbf{f}\cdot\mathbf{u})) \tag{2.11}$$

where $\mathbf{f}$ is given by Eq. (2.9). Next differentiate the mass relation to get $c^2 dm/dt = \gamma^2(u)\mathbf{u}\cdot d\mathbf{u}/dt$. Then compute $\mathbf{u}\cdot\mathbf{f} = \mathbf{u}\cdot((dm/dt)\mathbf{u} + md\mathbf{u}/dt)$ and substitute in the previous equation to find

$$\mathbf{u}\cdot\mathbf{f} = c^2\frac{dm}{dt} \tag{2.12}$$

Thus

$$G_a = ce^{-1}\gamma(u)(\mathbf{f}, -c\frac{dm}{dt}) \qquad (\Rightarrow G^a = ce^{-1}\gamma(u)(\mathbf{f}, c\frac{dm}{dt})) \tag{2.13}$$

It now follows from the above expressions for $\mathbf{f}$ and U and using the relation $d/d\tau = \gamma(u)d/dt$ linking the coordinate time t and the proper time τ that $G^a = ce^{-1}m_0 dU^a/d\tau$ and so the components G^a are the components of a space-time vector. Thus, since $G_a = F_{ab}U^b$, with U an arbitrary timelike space-time vector, it follows that the components F_{ab} are the components of a skew-symmetric covariant space-time tensor, which is the quantity F. Two things immediately follow from this. First, since F is a tensor under Lorentz transformations, the equations

$$\partial F^{ab}/\partial x^b = 0 \qquad \partial F_{ab}/\partial x^c + \partial F_{bc}/\partial x^a + \partial F_{ca}/\partial x^b = 0 \tag{2.14}$$

are such that if they are true in one inertial frame, they are true in them all. They are easily evaluated in any inertial frame, using Eq. (2.10), and turn out to be a rather elegant form of Maxwell's equations *in vacuo*

$$\text{div }\mathbf{E} = 0 \qquad \text{div }\mathbf{H} = 0 \qquad \text{curl }\mathbf{H} = \frac{1}{c}\frac{\partial\mathbf{E}}{\partial t} \qquad \text{curl }\mathbf{E} = -\frac{1}{c}\frac{\partial\mathbf{H}}{\partial t} \tag{2.15}$$

The extension of this result when electric charges and currents are present is straightforward. Second, the tensor-transformation law for the components F_{ab} under Lorentz transformations, together with Eq. (2.10), reveal the transformation laws for the components of the electric and magnetic fields from one inertial frame to another. Thus, in the special case when I, and I', are in standard configuration, one finds, using Eq. (2.7), that the electric and magnetic field components in I and those in I', denoted now with primes, are related by

$$\begin{array}{lll} E_1' = E_1, & E_2' = \gamma(v)(E_2 - v/cH_3), & E_3' = \gamma(v)(E_3 + v/cH_2) \\ H_1' = H_1, & H_2' = \gamma(v)(H_2 + v/cE_3), & H_3' = \gamma(v)(H_3 - v/cE_2) \end{array} \tag{2.16}$$

Thus, the unification of the electric and magnetic fields into a single *electromagnetic field* and which was achieved by the combination of the theories of Maxwell and Einstein, is given an elegant mathematical expression by identifying this electromagnetic field with Minkowski's (*Maxwell) tensor F* given in Eq. (2.10). This is one of the most impressive results of the 4D formulation of special relativity. When Maxwell's equations are written in the form Eq. (2.15) they are essentially written in a 3D language which still recognises the existence of separate electric and magnetic fields. However, in the form Eq. (2.14) they recognise only the (Minkowski–) Maxwell tensor and are formally much simpler. (In fact, by utilising the duality operator, one can rewrite the second equation in Eq. (2.14) in the same form as the first equation, but with F replaced by its dual F^*.) Thus if the sole electromagnetic source in I is a charge at rest, the field $\mathbf{H}$ is zero in I, but $\mathbf{H}'$ is not zero in I' (unless $v = 0$). [A semblance of order is restored by noting, from the transformation law for the tensor F, that for a Lorentz transformation consisting of a spatial rotation together with $t' = t$, the electric and magnetic fields transform as normal space vectors.] One of the main objections raised by Einstein at the beginning of his 1905 paper was that, in the Faraday induction experiments, the *explanation* for the current arising in a closed wire which was subject to a changing magnetic field depended on whether the wire or the magnet was regarded as being at rest. This asymmetry in the explanation contrasted somewhat awkwardly with the "symmetry" of the actual experiments. The reason is now clear. The explanation is given from the point of view of a *single* reference frame and in terms of the fields $\mathbf{E}$ and $\mathbf{H}$. If it is given in terms of the tensor F it is then in Lorentz invariant form and no such problem arises.

Finally a brief discussion will be given of the historical development of special relativity and, in particular, of particle dynamics and the electromagnetic theory. A more detailed account can be found, for example, in the books of Whittaker [4] and Pais [11] (even if the latter is not always in agreement with the former) and also in the book of Dugas [12]. Perhaps the first major statement regarding this subject came, following a suggestion by Maxwell, from the null result of the Michelson–Morley experiment in 1887, following a flawed first proposal by Michelson in 1881. Although it was a major experimental result, it was a "null" one and no positive conclusion was drawn from it by the experimenters. More positively Poincaré, in 1899, declared boldly that optical phonemena depended only on the relative motion of the bodies concerned and that absolute motion was undetectable by any experiment. Later, in 1904, he elevated this to what is now the well-known *principle of relativity* and foresaw a new dynamics with particle speeds restricted to being less than that of light. He also suggested, as a consequence of this, that the laws of physics should be able to be written in a way that reflected this inability to distinguish one inertial frame from another. In 1903, Lorentz, in considering those transformations which preserved Maxwell's equations, came across the transformations which, thanks to Poincaré, now bear his name (although they had been found earlier by Voigt [13] in his consideration of the wave equation). Whilst it is clear that Lorentz's contribution is fundamental at this point, his work carries the defects of the introduction of an (unclear) concept of "local time", an undetermined function in these transformations (shown to be identically equal to unity later by Einstein and Poincare), and an anomalous term in one of the Maxwell equations after the latter's transformation (again corrected by Einstein and Poincaré).

Einstein's famous paper of 1905 contains no references and is not easy to fit into the logical order of things. Although it is believed that Einstein was aware of the Michelson–Morley result, this is not absolutely clear. He does, however, make a reference in his paper to the failure to detect the Earth's movement through the ether. However, the experiment of Michelson and Morley was not the only such experiment. He was also, apparently, unaware of the recent work of Lorentz. It is not clear to the present author exactly how much of Poincaré's writings Einstein was familiar with. What Einstein's paper does contain is, perhaps, the first clear and natural physical derivation of the Lorentz transformations, where Lorentz's local time is now simply "ordinary" time but for another observer. He then goes on to find the well-known expressions for the length contraction and time dilation phenomena, the relativistic velocity addition laws and the relativistic formulae for the Doppler and aberration effects. This use of the Lorentz transformations to derive the so-called Lorentz–Fitzgerald contraction idea made clear for the first time that it was a pure (and natural) kinematic effect. (One should be careful here about the interpretation of this "contraction" result in special relativity. In its simplest form, it is a pure kinematical statement about the relationship between the space coordinates of, for example, a rigid rod in two inertial frames. It is not a statement of what would, in fact, be "seen" by an observer. This latter feature requires a more complicated consideration of those photons from the object being viewed which reach the eye simultaneously in the observer's frame.)

Einstein also used the Lorentz transformations to derive, directly, the transformations of the electric and magnetic fields given in Eq. (2.16) under a Lorentz transformation. This procedure is different from the one described earlier in this section. Einstein started with Maxwell's equations for the electric and magnetic fields $\mathbf{E}$ and $\mathbf{H}$ in the inertial frame I and given in Eq. (2.15) and, using the Lorentz transformations (2.7) between I and another inertial frame I', in standard configuration with I, together with the chain rule for differentiation, showed that the fields $\mathbf{E}'$ and $\mathbf{H}'$ given in Eq. (2.16) satisfy Maxwell's Eq. (2.15) (with all quantities in Eq. (2.15) now bearing primes). However, it is not entirely clear to the present author that $\mathbf{E}'$ and $\mathbf{H}'$, although clearly a solution of Maxwell's equations, constitute the required solution corresponding to the original electromagnetic field in I. The Lorentz force formula Eq. (2.9) must be invoked for this identification, since the electric and magnetic fields are essentially defined in terms of the force they impart to a charged particle with a certain velocity. But the relativistic form of Eq. (2.9) (that is, using the bracketed left-hand side in Eq. (2.9)) was not given until a little later by Planck. Granted Eq. (2.9), then Einstein's calculation proceeds satisfactorily. In fact, in 1906, Max Planck [14] investigated the dynamics of special relativity and by considering the equations of motion of a charged particle and assuming the Eq. (2.16) describing the transformations of the electric and magnetic fields, drew attention to the fact that if one considers the momentum of a particle of rest mass m_0 moving with 3-velocity $\mathbf{u}$ to be $m_0\gamma(u)\mathbf{u}$ then the electromagnetic force on the particle is $d/dt(m_0\gamma(u)\mathbf{u})$. It is a consequence of Planck's work that the above momentum law was developed and that the special relativistic equivalent of Newton's second law is taken as force equals “rate of change of momentum” rather than “mass times acceleration”, these being non-equivalent now because of the non-constancy of mass.

But perhaps the most important and far-reaching work on the 4D formulation of special relativity was accomplished by Hermann Minkowski [15]. It was he who introduced the metric which bears his name (even if he was inclined to use “imaginary time” to remove negative signs in the metric signature). He also introduced the Maxwell (skew–symmetric) tensor F and showed how the natural and simple transformation laws for such a tensor under Lorentz transformations lead to the transformation laws in Eq. (2.16) for the electric and magnetic fields. He was also responsible for extending Maxwell’s work on the 3D energy–momentum tensor and introducing the familiar 4D Maxwell energy–momentum tensor. In addition to this he also introduced the concepts of proper time, world line, the 4-velocity, the 4-momentum and the 4-acceleration and found the 4D version of the Lorentz force formula. His contribution to the mathematical formulation of special relativity was immense and his unique 4D way of looking at space-time has not only simplified the theory but put it into a clearer perspective. To Minkowski is due the most useful concept of the space-time diagram and the null cone. He would have frowned upon the breaking up of space-time into space and time and, in some sense, “foresaw” the philosophy of the space-time of general relativity. Certainly, his insistence on this 4D approach and the techniques he developed for it were important foundation stones for general relativity.

References

1. D. Hilbert. The Foundations of Geometry, Open Court Publishing, Chicago, 1902.
2. G. Berkeley. Principles of Human Knowledge, Collins, 1967.
3. E. Mach. The Science of Mechanics, Open Court Publishing, Illinois, 1960.
4. E. T. Whittaker. A History of the Theories of Aether and Electricity Two Volumes. Nelson, 1951.
5. A. Einstein. Annalen der Physik Vol 17, 1905, 891.
6. E. C. Zeeman. J.Math.Phys. Vol 5, 1964, 490.
7. L. A. Pars. Phil. Mag. 1921, 249.
8. G. N. Lewis and R. C. Tolman. Phil. Mag. 1909, 517.
9. M. Born. Einstein’s Theory of Relativity, Dover, 1965.
10. J. Ehlers, W. Rindler and R. Penrose. Am. J. Phys. Vol 33, 1965, 995.
11. A. Pais. Subtle is the Lord, Oxford, 1982.
12. R. Dugas. A History of Mechanics, Routledge & Kegan Paul, London, 1955.
13. W. Voigt. Gott. Nach 1887, 41.
14. M. Planck. Verh. d. Deutsch, Phys. Ges. 1906, 136.
15. H. Minkowski. Gott. Nach. 1908, 53.

Chapter 3
The Adventures of Space-Time

Orfeu Bertolami

3.1 Introduction

Since the nineteenth century, it is known, through the work of Lobatchevski, Riemann, and Gauss, that spaces do not need to have a vanishing curvature. This was for sure a revolution on its own, however, from the point of view of these mathematicians, the space of our day to day experience, the physical space, was still an essentially *a priori* concept that preceded all experience and was independent of any physical phenomena. Actually, that was also the view of Newton and Kant with respect to time, even though, for these two space-time explorers, the world was Euclidean.

As is well known, Leibniz held a very different opinion, as for him space and time were meaningless concepts if it were not for their relation with the material world. Starting with the concepts of space and time as quantities intrinsically related to matter, Hertz developed, between 1889 and 1894, a new formulation of mechanics, which culminated in the posthumous publication in 1894 of the book *Die Prinzipien der Mechanik in neuem Zusammenhange dargestellt*. But of course, it was only through the *General Theory of Relativity*, in 1915, that it was understood that space-time cannot be considered independently of matter at all.

Stepping a bit backwards, it was through special relativity that it was understood that the independence of the laws of physics in inertial frames from the velocity of the frame of reference requires that space and time are treated on same foot. It was the mathematician Hermann Minkowski who, in 1908, realized that the unity of the laws of physics could be more elegantly described via the fusion of space and time into the concept of *space-time*. Hence, the space-time is 4D. For Minkowski however, the similarity between space and time was not complete, as he defined the time coordinate using the imaginary unity so to preserve the Euclidean signature of the space-time metric. This description is not, as we know today, very satisfactory; one rather uses the Lorentzian signature for the space-time metric.

V. Petkov (ed.), *Relativity and the Dimensionality of the World*, 31–47.

Not much later, in 1909, the Finish Physicist Gunnar Nordström speculated that space-time could very well have more than four dimensions. A concrete realization of this idea was put forward by Theodor Kaluza in 1919 and Oskar Klein in 1925, who showed that an unified theory of gravity and electromagnetism could be achieved through a 5D version of general relativity and the idea that the extra dimension was compact and very small, and could hence, have passed undetected. This idea was very dear to Einstein, and this lead has been widely followed in further attempts to unify all known four interactions of nature. These developments, and most particularly general relativity, represented a fundamental departure from the way nineteenth-century mathematicians viewed space and also changed the attitude of physicists with respect to the physical world. Space-time is not a passive setting for physics as it is the solution of the field equations for the gravitational field for a given matter distribution, and the former evolves along with space-time. This methodology led physicists to describe nature along the lines of Cézanne's principle, that is *through the cylinder, the sphere, the cone ...*, i.e. through a geometrical or metrical description. Moreover, research in physics is now closely related with the "spacetime adventures" as, depending on the imposed conditions, space-time can expand, shrink, be torn, originate "baby" space-times and so on. And it is through physics that space-time acquires quite specific features. Let us introduce some examples.

The requirement of chiral fermions in 4D demand that, if there exist more than 4 space-time dimensions, then the total number of dimensions, d, must be even if all extra ones are compact [1]. To obtain a consistent effective 4D model arising from a d-dimensional Einstein–Yang–Mills theory, one should consider multidimensional universes of the form $M_d = \mathbf{R} \times G^{\text{ext}}/H^{\text{ext}} \times G^{\text{int}}/H^{\text{int}}$, where $G^{\text{ext(int)}}$, and $H^{\text{ext(int)}}$ are respectively the isometry groups in $3(d)$ dimensions. This technique, known as *coset space dimensional reduction* [2] (see Ref. [3] for an extensive discussion), is quite powerful and has been used in various branches of theoretical physics. In cosmology, when considering homogeneous and isotropic models (a 1D problem) it can be used, for instance, to obtain effective models arising from 4D [4] and d-dimensional Einstein–Yang–Mills–Higgs theories [5]. For the latter case, one considers for instance, $G^{\text{ext(int)}} = SO(4)\ (SO(d+1))$ and $H^{\text{ext(int)}} = SO(3)\ (SO(d))$ as the homogeneity and isotropy isometry groups in $3(d)$ dimensions.

The demand that supersymmetry, a crucial property of the 10D superstring theory, is preserved in 4 dimensions requires that 6 dimensions of the world are compact, have a complex structure, no Ricci curvature and an $SO(3)$ holonomy group. That is, this compact space must correspond to a Calabi–Yau manifold [6]. Connecting all string theories through S and T dualities suggests the existence of an encompassing theory, M-theory, and that space-time is 11D [7]. In another quantum approach to space-time, loop quantum gravity, it is suggested that space-time has, at its minutest scale, presumably the Planck scale, $L_P \simeq 10^{-35}$ m, a discrete structure [8].

Actually, the discussion on the number of space (n) and time (m) dimensions is not a trivial one, as it is related with the predictive power of solutions of the partial differential equations (PDEs) that describe nature. Indeed, given the importance of second-order PDEs for physics, it is natural to draw some general conclusions about this type of PDEs [9]. Consider a second-order PDE in $\mathbf{R}^d$, with $d = n+m$:

$$\left[\sum_{i=1}^{d}\sum_{j=1}^{d} A_{ij}\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j} + \sum_{i=1}^{d} v_i \frac{\partial}{\partial x_i} + f\right] u = 0 \,, \tag{3.1}$$

where the matrix A_{ij}, which can be taken without loss of generality to be symmetric, vector v_i and the function f are differential functions of d coordinates. Depending on the signs of the eigenvalues of A_{ij}, the PDE is said to be:

(i) Elliptic in some region of $\mathbf{R}^d$, if all eigenvalues are negative or all positive.
(ii) Hyperbolic in some region of $\mathbf{R}^d$, if one eigenvalue is negative and the remaining ones are positive (or vice-versa).
(iii) Ultrahyperbolic in some region of $\mathbf{R}^d$, if at least two eigenvalues are negative and at least two are positive.

The crucial issue about PDEs is that only hyperbolic equations allow for a well-posed *boundary value problem*, that is, for an unique solution, and a well-posed *initial value problem*, that is, initial data that lead to future predictions on regions beyond the boundary data, excluding singular points. Elliptic PDEs, on the other hand, allow for a well-posed boundary value problem, but an ill-posed initial value problem, so that no predictions about the future on regions beyond the boundary data, that is beyond local observations, can be made. Ultrahyperbolic PDEs, on their turn, have, for both space-like and timelike directions in a hypersurface, an ill-posed initial-value problem.

Hence, one sees that if $n = 0$ for any m or $m = 0$ for any n, the resulting PDEs are elliptic and hence no predictions can be made. If, on the other hand, $m \geq 1$ and $n \geq 1$, the PDEs are ultrahyperbolic and hence, lead to unpredictability.

One can advance with other reasons for excluding certain combinations of m and n. For instance, in a world where $n < 3$, there is no gravitational force in general relativity [10]. Moreover, one should expect weird "backward causality" if $m > 1$. It has has been pointed long ago by Ehrenfest [11], that if $n > 3$, neither atoms nor planetary orbits can be stable. This feature is associated with the fact that solutions of the Poisson equation give rise to electrostatic and gravitational potentials for a point-like particle that are proportional to r^{2-n} for $n > 2$ and to forces that are proportional to r^{1-n}. For $n > 3$ the two-body problem has no stable orbit solutions. The conclusion is that the choice $m = 1$ and $n = 3$ has quite desirable features and would be the selected one if one has reasons to think that the dimensionality of the world is chosen by selection arguments. We shall return to the issue of selection of "worlds" later on.

In what follows, we shall elaborate on how contemporary high-energy physics has changed our view of the space-time structure; however, before that we shall make a detour and discuss some mathematical properties of spaces.

3.2 Mathematical Space-Time

The historical development of the general theory of curved spaces,Riemannian geometry, has been guided and strongly influenced by the *General Theory of Relativity*. It is relevant to stress that, besides its mathematical interest, general relativity has passed all experimental tests so far and is believed to be a sound description of the space-time dynamics [12, 13].

Let us present some of the basic features of mathematical space in view of their relevance to physics. A d-dimensional differentiable manifold M endowed with a symmetric, non-degenerate second-rank tensor, the metric, g, is called a pseudo-Riemannian manifold, (M,g). A pseudo-Riemannian manifold whose metric has signature $(+,...,+)$ is said to be Riemannian. The metric of a pseudo-Riemannian manifold has a Lorentzian signature $(-,+,...,+)$. A condition for a differentiable manifold to admit a Lorentzian signature is that it is noncompact or has a vanishing Euler characteristic. A well-known theorem due to the mathematician Tulio Levi-Civita, states that a pseudo-Riemannian manifold has a unique symmetric affine connection compatible with the metric, being hence equipped with geodesics.

Some spaces are of particular importance for physics, since they correspond to solutions of the Einstein equations with a cosmological constant,[1] Λ:

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = 8\pi G T_{\mu\nu} + \Lambda g_{\mu\nu}, \tag{3.2}$$

where $R_{\mu\nu}$ is the Ricci curvature of M, R its trace, G is Newton's constant and $T_{\mu\nu}$ is the energy–momentum tensor of matter in (M,g). A minimal list includes:

1. The *de Sitter* (dS) space[2] which corresponds to the $(d+1)$-dimensional hyperboloid

$$-(x^0)^2 + (x^1)^2 + \cdots + (x^{d+1})^2 = r_0^2 \tag{3.3}$$

in a $(d+1)$-dimensional Minkowski space, which for the arbitrary constant r_0, satisfies the vacuum Einstein equations with a cosmological constant

$$\Lambda = \frac{d(d-1)}{2r_0^2} \; . \tag{3.4}$$

2. The *anti-de Sitter* (AdS) space which corresponds to the universal cover of the $(d+1)$-dimensional hyperboloid, that is

$$(x^1)^2 + \cdots + (x^d)^2 - (x^{d+1})^2 - (x^{d+2})^2 = -r_0^2 \tag{3.5}$$

in a $(d+2)$-dimensional space, which satisfies the vacuum Einstein equations with a cosmological constant

[1] We use units where $c = \hbar = k = 1$.

[2] After the Dutch Astronomer Willem de Sitter, who in 1917 first described this space.

$$\Lambda = -\frac{d(d-1)}{2r_0^2} \, . \tag{3.6}$$

3. The *Robertson–Walker* space[3] which corresponds to an homogeneous and isotropic spacetime. If (M_4, g) is a 4D manifold of constant curvature, corresponding to Euclidean $\mathbf{R}^3$ $(k=0)$, spherical $\mathbf{S}^3$ $(k=1)$, or hyperbolic $\mathbf{H}^3$ $(k=-1)$ spaces or quotients of these by discrete groups of isometries, then $M_4 = \mathbf{R} \times M_3$ and

$$ds^2 = -dt^2 + a^2(t)\left[d\chi^2 + f^2(\chi)(d\theta^2 + \sin^2\theta d\phi^2)\right] \, , \tag{3.7}$$

where $f(\chi) = (\chi\, , \sin\chi$, or $\sinh\chi)$, depending on the value of the constant spatial curvature $(k = 0, 1, -1)$. This metric is a solution of the Einstein equations for matter that can be described as a perfect fluid with velocity $\mathbf{u} = \frac{\partial}{\partial t}$ and energy density

$$\rho = \frac{\rho_0}{a^\gamma} \, , \tag{3.8}$$

where ρ_0 is a constant, $\gamma = (3)4$ corresponds to (non)relativistic matter and the scale factor, $a(t)$, satisfies the Friedmann equation, a constraint equation,

$$\frac{\dot{a}^2}{2} - \frac{4\pi G\rho_0}{3a^{\gamma-2}} = -\frac{k}{2} \, , \tag{3.9}$$

which can be easily recognized as the first integral of motion of a unit mass particle in the potential $V(a) = -4\pi G\rho_0/3a^{\gamma-2}$.

Notice that, since ρ_0 is positive, then so is the energy density and hence, the scale factor blows up in a finite time, which corresponds to a curvature singularity, the *Big Bang* or the *Big Crunch*. This means that time-like geodesics of the integral curves of $\frac{\partial}{\partial t}$ are incomplete. Actually, this is a fairly general feature of space-time, a result known as Hawking–Penrose singularity theorem, according to which physically meaningful Lorentzian manifolds are singular, i.e. are geodesically incomplete. A pedagogical and comprehensive introduction to the simplest singularity theorem of Hawking and Penrose can be found in Ref. [14].

It is relevant to point out that an important condition in the Hawking–Penrose singularity theorem is the one which concerns the physical nature of a manifold. A Lorentzian manifold (M, g) is said to be physically reasonable when it satisfies the *strong energy condition*:

$$R_{\mu\nu}V^\mu V^\nu \geq 0 \, , \tag{3.10}$$

for any timelike vector field, V^μ. From Einstein's equations this statement is equivalent, for $d \geq 2$, to the condition on the energy–momentum tensor and its trace, T,

$$T_{\mu\nu}V^\mu V^\nu \geq \frac{T}{d-1}V_\mu V^\mu \, , \tag{3.11}$$

---

[3] After the American and British mathematicians, who in 1930s showed the generality of this space.

which is satisfied by the vacuum, the cosmological constant, if $\Lambda \geq 0$, and by a perfect fluid if $\rho + 3p \geq 0$. Note that this condition is not respected during the inflationary period and by the present state of the Universe.

Another important feature of space-time concerns its topology. Locally, a topology is induced by the distance function $d(P,Q)$ between points P and Q in $\mathbf{R}^d$ through the definition of open sets, that is, sets for which $d(P,Q) < r_0$, where r_0 is an arbitrary quantity. The properties of open sets follow from the Hausdorff's condition or separation axiom, according to which points P and Q in $\mathbf{R}^d$ have non-intersecting neighborhoods U and V such that $U \ni P$ and $V \ni Q$. It follows that the intersection of open sets is an open set and that the union of any number of open sets is also an open set. Topology also concerns the global structure of a space and can be classified by the differential forms it admits. The topology of low-dimensional spaces ($d \leq 3$) is fully characterized by its genus.

It is rather remarkable that, on the largest scale, space-time can be modeled by a 4D manifold M_4 which is decomposed into $M_4 = \mathbf{R} \times M_3$, and is endowed with a locally homogeneous and isotropic Robertson–Walker metric, Eq. (3.7). As we have seen, the spatial section M_3 is often taken to be one of the following simply connected spaces: Euclidean $\mathbf{R}^3$, spherical $\mathbf{S}^3$, or hyperbolic $\mathbf{H}^3$ spaces. However, M_3 may be actually a multiply connected quotient manifold $M_3 = \widetilde{M}/\Gamma$, where Γ is a fixed point freely acting group of isometries of the covering space $\widetilde{M} = (\mathbf{R}^3, \mathbf{S}^3, \mathbf{H}^3)$.

It is known that for the Euclidean geometry, besides $\mathbf{R}^3$ there are 10 classes of topologically distinct compact 3D spaces consistent with this geometry, while for the spherical and hyperbolic geometries there are actually an infinite number of topologically inequivalent compact manifolds with nontrivial topology [15].

It is no less remarkable that the space-time topology can, at least in principle, be tested via the study of multiple images in the cosmic microwave background radiation (CMBR). A quite direct strategy to test the putative nontrivial topology of the spatial sections of the Universe is the "circles-in-the-sky" method. It relies on the search of multiple images of correlated circles in the CMBR maps [16]. Thus, in a nontrivial topology, the sphere of last scattering intersects a particular set of images along pairs of circles of equal radii, centered at different points on the last scattering sphere with the same distribution of temperature fluctuations.

It has been argued that an important evidence for a nontrivial topology arises from the fact that the Poincaré dodecahedral and the binary octahedral spaces can account for the observed low value of the CMBR quadrupole and octopole moments measured by the WMAP team [17–19]. However, a more recent search for the circles-in-the-sky, down to apertures of about 5° using WMAP 3-years data have not been successful in confirming this possibility [20].

Of course, a topologically nontrivial space can be only detected if the Universe is not exceedingly larger than the size of the last scattering surface, which is clearly a quite restrictive condition and consistent with a rather modest period of inflation in the early Universe. Even though, it is worth mentioning that through the circles-in-the-sky method one can obtain, besides the constraints arising from the usual astrophysical observational methods (supernova, baryon acoustic oscillations, CMBR bounds, etc), additional limits to the cosmological models. This can be shown to be particularly relevant for the ΛCDM model [21], for the unified model of dark energy and dark matter, the generalized Chaplygin gas model [22], characterized by the equation of state $p = -A/\rho^{\alpha}$, where p is the pressure, ρ, the energy density and A and α are positive constants [23], and for modified gravity models inspired in braneworld constructions [24].

Another relevant issue about the property of spaces concerns their boundaries. In $d = 4$ the theory of *cobordism* guarantees that for all compact 3-surfaces there always exists a compact 4D manifold such that S^3 is the only boundary, or equivalently, all 3D *compact* hypersurfaces are cobordant to zero [25]. This question is particularly relevant when considering the sum of histories in Quantum Cosmology. In these approach, the quantum state of a $d = 4$ Universe is described by a wave function $\Psi[h_{ij}, \Phi]$, which is a functional of the spatial 3-metric, h_{ij}, and matter fields generically denoted by Φ on a compact 3D hypersurface Σ. The hypersurface Σ is then regarded as the boundary of a compact 4-manifold M_4 on which the 4-metric $g_{\mu\nu}$ and the matter fields Φ are regular. The metric $g_{\mu\nu}$ and the fields Φ coincide with h_{ij} and Φ_0 on Σ and the wave function is then defined through the path integral over 4-metrics, ^4g, and matter fields:

$$\Psi[h_{ij}, \Phi_0] = \int_C D[^4g]D[\Phi] \exp\left(-S_E[^4g, \Phi]\right), \tag{3.12}$$

where S_E is the Euclidean action and C is the class of 4-metrics and regular fields Φ defined on Euclidean compact manifolds and with *no other* boundary than Σ. This wave function is the solution of the Wheeler–DeWitt equation, and it has been argued that this formalism allows for a theory of the initial conditions for the Universe [26]. Indeed, in this proposal, the wave functions are associated with a probability distribution and the most likely observational features of the Universe correspond to the peak of the solution of the Wheeler–DeWitt equation. Explicit solutions of this equation in the so-called *minisuperspace* approximation are known for some cases of interest, such as for a Universe dominated by a massless conformally coupled scalar field [26] and by radiation [27].

An extension of this proposal for Universes with $d > 4$ dimensions has some complications. In these d-dimensional models, the wave function would be a functional of the $(d-1)$ spatial metric, h_{IJ}, and matter fields, Φ, on a $(d-1)$-hypersurface, Σ_{d-1}, and is defined as the result of performing a path integral over all compact d-metrics and regular matter fields on M_d, that match h_{IJ} and the matter fields on Σ_{d-1}. One starts assuming that the $(d-1)$-surface Σ_{d-1} does not possess any disconnected parts. Is there always a d-dimensional manifold M_d such that Σ_{d-1} is the only boundary? In higher dimensional manifolds, this is actually not guaranteed. There exist compact $(d-1)$-hypersurfaces Σ_{d-1} for which there is no compact d-dimensional manifold such that Σ_{d-1} is the only boundary. This seems to indicate that in $d > 4$ dimensions there are configurations which cannot be attained by the sum over histories in the path integral. The wave function for such configurations would therefore be zero. However, this situation can be circumvented so as to obtain nonvanishing wave functions for such configurations, namely by dropping the assumption that the $(d-1)$-surface Σ_{d-1} does not possess any disconnected parts (see e.g. [28, 29] and references therein).

Indeed, if one assumes that the hypersurfaces Σ_{d-1} consist of any number $n > 1$ of disconnected parts $\Sigma^{(n)}_{d-1}$, then one finds that the path integral for this disconnected configuration involves terms of two types. The first type consists of disconnected d-manifolds, each disconnected part of which closes off the $\Sigma^{(n)}_{d-1}$ surfaces separately. These will exist only if each of the $\Sigma^{(n)}_{d-1}$ are cobordant to zero, but this may not always be the case. There will indeed be a second type of term which consists of connected d-manifolds joining some of the $\Sigma^{(n)}_{d-1}$ together. This second type of manifold will always exist in any number of dimensions, providing the $\Sigma^{(n)}_{d-1}$ are topologically similar, i.e. have the same *characteristic numbers*. The wave function of any $\Sigma^{(1)}_{d-1}$ surface which is not cobordant to zero would be nonvanishing and can be obtained by assuming the existence of other surfaces of suitable topology and then summing over all compact d-manifolds which join these surfaces together. Thus, given a compact $(d-1)$ hypersurface Σ_{d-1} which is not cobordant to zero, a nonzero amplitude can be found by assuming it possesses disconnected parts.

However, the above considerations for disconnected pieces and generic Σ_{d-1} surfaces would spoil the Hartle–Hawking prescription, since the manifold would have more than one boundary. In other words, the general extension discussed above would imply in a description in terms of propagation between such generic Σ_{d-1} surfaces. The wave function would then depend on every piece and not on a single one. Nevertheless, if one restricts oneself to the case of a truncated model with a global topology given by a product of a 3D manifold to a d-dimensional one, then the spacelike sections always form a boundary of a d-dimensional manifold with no other boundaries. Since hypersurfaces $S^3 \times S^d$ are always cobordant to zero, it implies that for spacetimes with topology $\mathbf{R} \times S^3 \times S^d$ the Hartle–Hawking proposal can be always implemented [29].

Let us close this section, introducing a notion that has been recently quite useful in physics, namely the idea of an *orbifold*. From the mathematical point of view, an orbifold is a generalization of the concept of manifold which includes the presence of the points whose neighborhood is diffeomorphic to the coset $\mathbf{R}^d/\Gamma$, where Γ is a finite group of isometries. In physics, an orbifold usually describes an object that can be globally written as a coset M/G, where G is the group of its isometries or symmetries. The best known case of an orbifold corresponds to a manifold with boundary since it carries a natural orbifold structure, the $\mathbf{Z}_2$-factor of its double. Thus, a factor space of a manifold along a smooth $\mathbf{S}^1$-action without fixed points carries an orbifold structure.

In what follows we shall describe the properties that physical theories require for the physical space-time.

3.3 Physical Space-Time

Within the framework of general relativity, the dynamics of the physical space-time is actually related with the history and evolution of the Universe. The mathematical description of space-time does allow for a wide range of scenarios; however, recent developments in observational cosmology do indicate that our Universe is well described by a flat Robertson–Walker metric, meaning that the energy density of the Universe is fairly close to the critical one, $\rho_c \equiv 3H_0^2/8\pi G \simeq 10^{-29} g/cm^3$, where $H_0 \simeq 73\ km\ s^{-1} Mpc^{-1}$ is the Hubble expansion parameter at present. Furthermore, CMBR, Supernova, and large-scale structure data are consistent with each other if and only if the Universe is dominated by a smoothly distributed energy that does not manifest itself in the electromagnetic spectrum – dark energy. Moreover, it is found that the large scale structure of the Universe, as well as the dynamics of galaxies, requires matter that like dark energy, does not manifest electromagnetically – dark matter. More exactly, in the cosmic budget of energy, dark energy corresponds to about 73% of the critical density, while dark matter to about 23% and baryonic matter, the matter that we are made of, to only about 4% [30].

Actually, the dominance of dark energy at the present does have deep implications for the evolution of space-time. For instance, if dark energy remains the dominant component in the energy budget in the future, then geometry is no longer the determinant factor in the destiny of the Universe. As is well known, in a Universe where dark energy is subdominant, flat, and hyperbolic geometries give origin to infinity universes in the future; in opposition, a spherical universe does eventually recollapse and undergoes a Big Crunch in a finite time. If however, dark energy is the dominant component, the fate of the Universe is determined by the way it evolves. If its energy density is decreasing, the Universe will eventually be dominated by matter and its destiny is again ruled by its geometry as described above. If, on the other hand, the energy density remains constant, then the Universe expansion will continue to grow and the Universe will be quite diluted of matter. That is to say that, in the remote future the Universe will correspond to a dS space with a future horizon. This means that the world will have features similar to an isolated thermal cavity with finite temperature and entropy. A more drastic fate is expected if the energy density of dark energy continues to grow. This growth will eventually cause a *Big Rip*, that is, the growing velocity of the space-time expansion will eventually disrupt its very fabric and all known structures will be ripped off.

Actually, an ever accelerating Universe might not be compatible with some fundamental physical theories. For instance, an eternally accelerating Universe poses a challenge for string theory, at least in its present formulation, as it requires that its asymptotic states are asymptotically free, which is inconsistent with a space-time that exhibits future horizons [31–33]. Furthermore, it is pointed out that theories with a stable supersymmetric vacuum cannot relax into a zero-energy ground state if the accelerating dynamics is guided by a single scalar field [31, 32]. This suggests that the accelerated expansion might be driven by at least two scalar fields. It is interesting that some two-field models allow for solutions with an exit from a period of accelerated expansion, implying that decelerated expansion is resumed (see e.g. Ref. [34]). Hence, a logical way out of this problem is to argue that the dS space is unstable. This might also occur, for instance, due to quantum tunneling, if the cosmological constant is not too small.

Another significant feature about our Universe is that only if it has undergone a period of quite rapid and accelerated expansion in its early history, one can understand why its spatial section is so close to flat and why it is so homogeneous and isotropic on large scale [35–37]. This *inflationary* phase of accelerated expansion, a tiny fraction of a second after the big bang, about 10^{-35} seconds, corresponds to a period where the geometry of the Universe is described by a dS space. It is quite remarkable that a rather brief period of inflation, a quite generic behavior of most of the anisotropic Bianchi-type spaces [1], Kantowski–Sachs spaces [39] and inhomogeneous spaces [40] dominated by a cosmological constant, drives a microscopic Universe into a large one, whose features closely resemble ours. Moreover, in this process, small quantum fluctuations of the field responsible for inflation, the *inflaton*, are amplified to macroscopic sizes and are ultimately responsible for the formation of large-scale structure (see e.g. Ref. [41] for an extensive discussion). It is a great achievement of modern cosmology that the broad lines of this mechanism are corroborated by the observed features of the CMBR, such as its main peak, whose position is consistent with the size of the scalar density fluctuations that first reentered the horizon, as well as the nearly scale invariant and Gaussian nature of these fluctuations.

However, in what concerns space-time, the stock of surprises arising from physics is far from over. Indeed, recent developments on the understanding of string theory have led to speculations that may be regarded as somewhat disturbing for those who believe that the laws of nature can be described by an *action*, which encompasses the relevant underlying fundamental symmetries, and from which an unique vacuum arises and the spectrum of elementary objects, particles, can be found. These view has been recently challenged by a quite radical set of ideas. The genesis of these can be traced from the understanding that the initial outlook concerning the original five distinct string theories was not quite correct. It is now understood that there is instead a continuum of theories, that includes M-theory, interpolating the original five string theories. One rather speaks of different solutions of a master theory than of different theories. The space of these solutions is often referred to as the *moduli space of supersymmetric vacua* or *supermoduli-space*. These moduli are fields, and their variation allows moving in the supermoduli-space. The moduli vary as one moves in the space-time, as moduli have their own equations of motion.

However, the continuum of solutions in the supermoduli-space are supersymmetric and have all a vanishing cosmological constant. Hence, in order to describe our world, there must exist some non-supersymmetric "islands" in the supermoduli-space. It is believed that the number of these discrete vacua is huge, googles, $G = 10^{100}$, or googleplexes 10^{G}, instead of unique [42]. If the cause of the accelerated expansion of the Universe is due to a small cosmological constant, then the state of our Universe corresponds to moduli values some of the non-supersymmetric islands in the supermoduli-space. The fact that the magnitude of the cosmological constant is about 10^{120} smaller than its natural value M_P^4, where $M_P = 1.2 \times 10^{19}$ GeV is the Planck mass, makes it highly unlikely to find such a vacuum, unless there exists a huge number of solutions with every possible value for the cosmological constant. The space of all such string theory vacua is often referred to as the *landscape* [43].

From the landscape proposal springs a radical scenario. In principle, vacua of the landscape do not need to correspond to actual worlds, however, very much on the contrary, it is argued that the string landscape suggests a *multiuniverse*. According to this proposal, the multiple vacua of string theory is associated to a vast number of "pocket universes" in a single large mega-universe. These pocket universes, like the expanding universe we observe around us, are all beyond any observational capability, as they lie beyond the cosmological horizon. In the words of Susskind, a vociferous proponent of the multiuniverse idea [44], "According to classical physics, those other worlds are forever completely sealed off from our world". Clearly, the implications of these ideas are somewhat disturbing. First, the vacuum that corresponds to our world must arise essentially form a selection procedure, to be dealt with via anthropic or quantum cosmological considerations. Thus, it seems that somehow our existence plays an important role in the selection process. Second, the vast number of vacua in the landscape ensures the reality of our existence; one refers to this scenario as the *anthropic landscape*, when based on anthropic arguments. For sure, this interpretation is not free from criticism. It has been pointed out, for instance, that the impossibility of observing a multiuniverse implies that its scientific status is questionable. It is in the realm of metaphysics, rather than of physics [45]. It has also been argued that selection criteria like the anthropic landscape must be necessarily supplemented by arguments based on dynamics and symmetry, as only these lead to a real "enlightenment", the former are actually a "temptation" [46]. Indeed, Weinberg argues that the anthropic reasoning makes sense for a given constant whenever the range over which it varies is large compared with the anthropic allowed range. That is to say, it is relevant to know what constants actually "scan". The most likely include the cosmological constant, and the particle masses set by the electroweak symmetry-breaking mechanism. The possibility that the later is anthropically fixed is regarded as an interesting possibility, given that it renders an alternative solution for the hierarchy problem, such as technicolor or low-energy supersymmetry, that are not fully free of problems [47]. In any case, we feel that we cannot close this discussion without some words of caution. For instance, Polchinski has recently pointed out as the landscape picture requires a higher level of theoretical skepticism given that it suggests that science is less predictive. Furthermore, he remarks that the current scenario is tentative at best, as a nonperturbative formulation of string theory is still missing [48].

Let us close this discussion with a couple of remarks. The first concerns the possibility that the topology of the landscape is nontrivial. This hypothesis would imply that multiuniverses are not causally exclusive, meaning that within our Universe one might observe pocket subuniverses where the laws of physics are quite different from the ones we know. Since it is natural to assume the in these subuniverses the fundamental constants assume widely different values, one might expect to observe oddities such as quantum phenomena on macroscopic scales and relativistic effects at quite mundane velocities. Of course, this possibility would imply in a further loss of the ability to predict the properties of the cosmos.

Another relevant investigation on the selection of landscape vacua concerns the understanding of up to which extent the problem can be addressed in the context of the quantum cosmology formalism. As already discussed, this formalism allows for a theory of initial conditions, which seems to be particularly suitable to deal with the problem of vacua selection. It is quite interesting that this problem can be addressed as an N-body problem for the multiple scattering among the N-vacua sites of the landscape [49]. The use of the Random Matrix Theory methods shows that the phenomenon of localization on a lattice site with a well-defined vacuum energy, the so-called Anderson localization, occurs. It is found that the most probable universe with broken supersymmetry corresponds to a dS universe with a small cosmological constant. Furthermore, it is argued that the relevant question on why the Universe started in a low-entropy state can only be understood via the interplay between matter and gravitational degrees of freedom and the inclusion of dynamical back-reaction effects from massive long wavelength modes [49]. It is interesting to speculate whether these features remain valid beyond scalar field models, the case considered in Refs. [49]. Massive vector fields with global $U(1)$ and $SO(3)$ symmetries seem to be particularly suitable to generalize these results, given that the reduced matrix density and Wigner functional of the corresponding *midisuperspace* model [50] exhibit properties that closely resemble the localization process induced by the back-reaction of the massive long wavelength modes discussed in [49].

Let us describe some recent developments involving the AdS space, introduced in the previous section.

In the so-called *braneworlds*, one can admit two 3-branes at fixed positions along the 5th dimension, such that the *bulk*, the 5D spacetime is AdS, with a negative cosmological constant, $\Lambda = -3M_5^3k^2$, where M_5 is the 5D Planck mass and k a constant with dimension of mass. In this setup, compactification takes place on a $\mathbf{S^1}/\mathbf{Z^2}$ orbifold symmetry. Einstein equations admit a solution that preserves Poincaré invariance on the brane, and whose spatial background has a nonfactorisable geometry with an exponential warp form

$$ds^2 = e^{-2k|z|} g_{\mu\nu} dx^\mu dx^\nu + dz^2 \ , \tag{3.13}$$

$g_{\mu\nu}$ being the 4D metric.

The action of the model is given by Randall and Sundrum [51]

$$S_5 = 2\int d^4x \int_0^{z_c} dz\sqrt{-g_5}\left[\frac{M_5^3}{2}R_5 - 2\Lambda\right] - \sigma_+ \int d^4x\sqrt{-g_+} - \sigma_- \int d^4x\sqrt{-g_-} \ , \tag{3.14}$$

corresponding to the bulk space with metric, g_{5MN}, a with a positive tension, σ_+, brane with metric, $g_{+\mu\nu}$, sitting at $z = 0$, and a negative tension, σ_-, brane with metric, $g_{-\mu\nu}$, sitting at $z = z_c$. The standard model (SM) degrees of freedom lie presumably on the brane at $z = z_c$.

In order to ensure a vanishing cosmological constant in 4D one chooses:

$$\sigma_+ = -\sigma_- = 3M_5^3 k \ . \tag{3.15}$$

This is quite interesting, as a cancellation involving different dimensions is actually not possible within the context of the Kaluza–Klein compactification mechanism [52]. Another pleasing feature of this proposal is that the hierarchy between the Planck and the SM scales can be dealt with in a geometrical way, actually via the warp e^{-2kz_c} factor. Finally, integration over $\mathbf{S^1}$ allows obtaining Planck's constant in 4D:

$$M_P^2 = \frac{M_5^3(1-e^{-2kz_c})}{4k}, \tag{3.16}$$

which clearly exhibits a low dependence on kz_c.

The literature on braneworlds is quite vast and it is not our aim to review here the most important proposals, however it is interesting to mention that the type of cancellation mechanism described above can be also considered to understand why Lorentz invariance is such a good symmetry of nature in 4 dimensions [53]. Furthermore, it is worth mentioning that evolving 3-branes can be regarded as solutions of an effective theory that arises from the fundamental M-theory. Indeed, the $d = 11$ M-theory, compactified on an $\mathbf{S^1}/\mathbf{Z^2}$ orbifold symmetry with E_8 gauge multiplets on each of the 10D orbifold-fixed planes, can be identified with the strongly coupled $E_8 \otimes E_8$ heterotic theory [54]. An effective theory can be constructed via the the reduction of the $d = 11$ theory on a Calabi–Yau threefold space, $\mathbf{K}$, that is, $M_{11} = \mathbf{R}^4 \times \mathbf{K} \times \mathbf{S^1}/\mathbf{Z^2}$. It is shown that this effective theory admits evolving *cosmological domain-wall solutions* corresponding to a pair of 3-branes [55].

Furthermore, it is relevant to realize that an AdS space is the natural background for supergravity and M-theory, given that in the weak coupled limit the latter theory corresponds to a $N = 1$ supergravity theory in 11D. Thus, the AdS space is intimately related with string theory. Actually, this background space is quite crucial in the so-called Maldacena or AdS/CFT conjecture (see Ref. [56] for an extensive discussion), according to which a supergravity theory in d-dimensions on a AdS space is equivalent to a conformal field theory (CFT) defined on the $(d-1)$-dimensional boundary of that theory.

Before drawing an end to our brief discussion on some of the properties of space-time, let us discuss one last striking development concerning the nature of space-time. It has been suggested that at the most fundamental level, the underlying geometry of spacetime is noncommutative. This feature arises from the discovery in string theory that the low-energy effective theory of a D-brane in the background of a NS–NS B field lives in a noncommutative space [57–59] where the configuration variables satisfy the commutation relation:

$$[x^\mu, x^\nu] = i\theta^{\mu\nu}. \tag{3.17}$$

where $\theta^{\mu\nu}$ is a constant antisymmetric matrix. Naturally, this set of numbers do not transform covariantly, which implies in the breaking of Lorentz invariance down to the stability subgroup of the noncommutative parameter [60]. Approaches where $\theta^{\mu\nu}$ is regarded as a Lorentz tensor were considered, for instance, in the context of a noncommmutative scalar field coupled to gravity in homogeneous and isotropic spaces [61].

Naturally, if space-time has a noncommutative structure one should expect important implications in field theory (see Refs. [62, 63] for extensive discussions) and even in the nonrelativistic limit, that is, at Quantum Mechanics level. In the first case, one finds a host of new effects including the violation of translational invariance (see [64] and references therein). In the nonrelativistic limit, versions of noncommutative Quantum Mechanics (NCQM) have been recently the subject of many studies. Although in string theory only the space coordinates exhibit a noncommutative structure, some authors have suggested NCQM models in which noncommutative geometry is defined in the whole phase space [65–67]. Implications for the gravitational quantum well, recently realized for ultra-cold neutrons from the research reactor of the Laue–Langevin Institute in Grenoble [68], have been examined for the NCQM models with a phase space noncommutative geometry [66, 69].

3.4 Concluding Remarks

Physics has unquestionably made untenable the philosophical thinking according to which space and time are *a priori* concepts, independent of the physical world. Physics has also immensely stretched the notions of space and time, expanding reality to limits that were thought to be beyond imagination. Indeed, the physical world was, according to Aristotle, compact and locked within the sublunar realm. Galileo's observations and the universality of Newton's mechanics have fundamentally changed that. Nineteenth-century physics was rather modest about the timescale of the world, based on thermodynamical considerations about dissipation of heat and the conversion of gravitational energy into heat. Indeed, estimates by Lord Kelvin and Helmholtz suggested a few hundred millions of years for the ages of the Sun and Earth. Geologists were actually the first to understand that this could not be possible. Earth had to be at least a billion years old to be consistent with the transformations that are in operation at present. Paleontologists followed suit, given the tight correlation between fossils and the geological strata they are found. On its hand, astronomy has open up space and time, providing us with impressive estimates of the size and the distance of astronomical objects, having ultimately shown us that space itself is expanding – in fact in an accelerated fashion according to the most recent observations. The ticket to fully exert the freedom to expand space and time was conquered when Einstein understood that general relativity was a theory of the space-time at large. Since then, scrutinizing the ways space-time might exist is, in a way, the very essence of physics. Physics has thus given substance to the pioneering work of scores of brilliant mathematicians who speculated on the geometry and topology of spaces.

According to the Swiss painter, Paul Klee, "L'art ne reproduit pas le visible, il rendre le visible," and, in a broader sense, the same can be said about physics. Indeed, from its original goal of describing nature, physics has created a picture of the world that is much richer than the one that meets the eye, and it turns out that, in this process, space-time has acquired a quite rich structure. However, the adventure is by no means over. On a quite fundamental level, we do not understand how to reconcile our picture of the macroscopic space-time with the rules of Quantum Mechanics, a theory that successfully describes all, but gravitational phenomena. This is an unbearable gap in our knowledge. Moreover, this difficulty has quite severe and concrete implications, the most evident being, as we have seen, that we cannot explain the smallness of the cosmological constant without paying a quite heavy toll. In fact, the cosmological constant problem is such a formidable challenge that it is tempting to go around it and compare it with Wittggenstein's suggestion, according to which all problems of philosophy are actually problems of language. Indeed, our expectation that the cosmological constant is immensely greater than the observed value on cosmological scales is based on the "language" of quantum field theory. We do not expect and we have not seen a breakdown of the quantum field theory formalism down to scales of about 10^{-18} m, but this still a long away from the typical quantum gravity length scale, $L_P \simeq 10^{-35}$ m. In fact, it is relevant to bear in mind that the cosmological constant problem is intimately related with supersymmetry, duality symmetries, and the spacetime dimensions [70]. When put together, these ingredients may imply in an important "language" shift, as is the case of the landscape scenario. Most likely, the ultimate landslide is still to come. In any case, unraveling the ultimate structure of space-time down to the smallest scale, and then back up to the largest one will remain, as it is nowadays, an exciting quest for many generations to come.

References

1. E. Witten, Nucl. Phys. B **186**, 412 (1981).
2. E. Witten, Phys. Rev. Lett. **38**, 121 (1977); P. Forgacs and N.S. Manton, Comm. Math. Phys. **72**, 15 (1980).
3. Yu.A. Kubyshin, J.M. Mourão, G. Rudolph and I.P. Volobujev, Dimensional Reduction of Gauge Theories, Spontaneous Compactification and Model Building, Lecture Notes in Physics **349** (Springer, Berlin, Heidelberg 1989).
4. O. Bertolami, J.M. Mourão, R.F. Picken and I.P. Volobujev, Int. J. Mod. Phys. A **6**, 4149 (1991).
5. O. Bertolami, Yu.A. Kubyshin and J.M. Mourão, Phys. Rev. D **45**, 3405 (1992).
6. P. Candelas, G. Horowitz, A. Strominger and E. Witten, Nucl. Phys. B **258**, 46 (1985).
7. E. Witten, Nucl. Phys. B **443**, 85 (1995).
8. A. Ashtekar, C. Rovelli and L. Smolin, Phys. Rev. Lett. **69**, 237 (1992).
9. M. Tegmark, Class. Quantum Grav. **14**, L69 (1997).
10. S. Deser, R. Jackiw and G. t'Hooft, Ann. Phys. **152**, 220 (1984).
11. P. Ehrenfest, Proc. Amsterdam Acad. **20**, 200 (1917).
12. C. Will, The Confrontation between General Relativity and Experiment, gr-qc/0510072.
13. O. Bertolami, J. Páramos and S. Turyshev, General Theory of Relativity: Will it survive the next decade?, gr-qc/0602016.
14. J. Natário, Relativity and Singularities – A Short Introduction for Mathematicians, math.dg/0603190.
15. J.-P. Luminet, L'Univers chiffonné (Gallimard 2005).
16. N.J. Cornish, D.N. Spergel and G.D. Starkman, Class. Quantum Grav. **15**, 2657 (1998).
17. J.-P. Luminet, J. Weeks, A. Riazuelo, R. Lehoucq and J.-P. Uzan, Nature **425**, 593 (2003).

18. R. Aurich, S. Lustig and F. Steiner, Class. Quantum Grav. **22**, 3443 (2005).
19. D.N. Spergel et al., Astrophys. J. Suppl. **148**, 175 (2003).
20. J. Shapiro Key, N.J. Cornish, D.N. Spergel and G.D. Starkman, Astro-ph/0604616.
21. M.J. Rebouças, J.S. Alcaniz, B. Mota and M. Makler, Astron. Astrophys. **452**, 803 (2006).
22. M.C. Bento, O. Bertolami, M.J. Rebouças and P.T. Silva, Phys. Rev. D **73**, 043504 (2006).
23. M.C. Bento, O. Bertolami and A.A. Sen, Phys. Rev. D **66**, 043507 (2002).
24. M.C. Bento, O. Bertolami, M.J. Rebouças and N.M.C. Santos, Phys. Rev. D **73**, 103521 (2006).
25. R.E. Stong, Notes on Cobordism, Mathematical Notes (Princeton University Press, Princeton 1968).
26. J. Hartle and S.W. Hawking, Phys. Rev. D **28**, 2960 (1983).
27. O. Bertolami and J.M. Mourão, Class. Quant. Grav. **8**, 1271 (1991).
28. J.J. Halliwell, Nucl. Phys. B **266**, 228 (1986).
29. O. Bertolami, P.D. Fonseca and P.V. Moniz, Phys. Rev. D **56**, 4530 (1997).
30. D.N. Spergel et al., Wilkinson Microwave Anisotropy Probe (WMAP) Three Year Results: Implications for Cosmology, astro-ph/0603449.
31. S. Hellerman, N. Kaloper and L. Susskind, JHEP **0106**, 003 (2001).
32. W. Fischler, A. Kashani-Poor, R. McNess and S. Paban, JHEP **0107**, 003 (2001).
33. E. Witten, The Cosmological Constant From The Viewpoint Of String Theory, hep-ph/0002297.
34. M.C. Bento, O. Bertolami and N.M.C. Santos, Phys. Rev. D **65**, 067301 (2002).
35. A. Guth, Phys. Rev. D **23**, 347 (1981).
36. A.D. Linde, Phys. Lett. B **108**, 389 (1982).
37. A. Albrecht and P. Steinhardt, Phys. Rev. Lett. **48**, 1220 (1982).
1. R.M. Wald, Phys. Rev. D **28**, 2118 (1983); L.G. Jensen and J.A. Stein-Schabes, Phys. Rev. D **34**, 931 (1986); T. Rothman and G.F.R. Ellis, Phys. Lett. B **180**, 19 (1986).
39. M.S. Turner and L.M. Widrow, Phys. Rev. Lett. **57**, 2237 (1986); J.D. Barrow, Nucl. Phys. B **296**, 697 (1988).
40. T. Piran and R.M. Williams, Phys. Lett. B **163**, 331 (1985); D.S. Goldwirth and T. Piran, Phys. Rev. Lett. **64**, 2852 (1990).
41. K. Olive, Phys. Rep. **190**, 302 (1990).
42. L. Susskind, The Anthropic Landscape of String Theory, hep-th/0302219.
43. R. Bousso and J. Polchinski, JHEP **0006**, 006 (2000).
44. L. Susskind, The Cosmic Landscape: String Theory and the Illusion of Intelligent Design (Little, Brown, New York 2005).
45. G.F.R. Ellis, On horizons and the cosmic landscape, Astro-ph/0603266.
46. F. Wilczek, Enlightenment, Knowledge, Ignorance, Temptation, hep-ph/0512187.
47. S. Weinberg, Living in the Multiverse, hep-th/0511037.
48. J. Polchinski, The Cosmological Constant and the String Landscape, hep-th/0603249.
49. R. Holman and L. Mersini-Houghton, Why the Universe Started from a Low Entropy State, hep-th/0511102; Why did the Universe Start from a Low Entropy State?, hep-th/0512070.
50. O. Bertolami and P.V. Moniz, Nucl. Phys. B **439**, 259 (1995).
51. L. Randall and R. Sundrum, Phys. Rev. Lett. **83**, 3370 (1999); 4690 (1999).
52. V.A. Rubakov and M.E. Shaposhnikov, Phys. Lett. B **125**, 139 (1983).
53. O. Bertolami and C. Carvalho, Lorentz symmetry derived from Lorentz violation in the bulk, gr-qc/0607043 to appear in Phys. Rev. D.
54. P. Horava and E. Witten, Nucl. Phys. B **460**, 506 (1996).
55. A. Lukas, B.A. Ovrut and D. Waldram, Phys. Rev. D **60**, 086001 (1999).
56. O. Aharony, S.S. Gubner, J. Maldacena, H. Ooguri and Y. Oz, Phys. Rep. **323**, 183 (2000).
57. A. Connes, M.R. Douglas and C. Hull, JHEP **9802**, 008 (1998).
58. N. Seiberg and E. Witten, JHEP **9909**, 032 (1999).
59. V. Schomerus, JHEP **9906**, 030 (1999).
60. S.M. Carroll, J.A. Harvey, V.A. Kostelecký, C.D. Lane and T. Okamoto, Phys. Rev. Lett. **87**, 141601 (2001).

61. O. Bertolami, L. Guisado, Phys. Rev. D **67**, 025001 (2003); O. Bertolami, Mod. Phys. Lett. A **20**, 1359 (2005).
62. R. J. Szabo, Quantum Field Theory on Noncommutative Spaces, hep-th/0109162.
63. M.R. Douglas and N.A. Nekrasov, Rev. Mod. Phys. **73**, 977 (2001).
64. O. Bertolami and L. Guisado, JHEP **0312**, 013 (2003).
65. Jian-zu Zhang, Phys. Rev. Lett. **93**, 043002 (2004).
66. O. Bertolami, J.G. Rosa, C.M.L. Aragão, P. Castorina and D. Zappalà, Phys. Rev. D **72**, 025010 (2005).
67. A.E.F. Djemai and H. Smail, Commun. Theor. Phys. **41**, 837 (2004).
68. V.V. Nesvizhevsky et al., Nature **415**, 297 (2002) 297; Eur. Phys. J.C **40**, 479 (2005).
69. O. Bertolami, J. G. Rosa, C.M.L. Aragão, P. Castorina and D. Zappalà, Mod. Phys. Lett. A **21**, 795 (2006).
70. E. Witten, Mod. Phys. A **10**, 2153 (1995).

Chapter 4
Physics in the Real Universe: Time and Space-Time

George F. R. Ellis[1]

Abstract The block universe idea, representing space-time as a fixed whole, suggests the flow of time is an illusion: the entire Universe just is, with no special meaning attached to the present time. This paper points out that this view, in essence represented by usual space-time diagrams, is based on time-reversible microphysical laws, which fail to capture essential features of the time-irreversible macro-physical behaviour and the development of emergent complex systems, including life, which exist in the real Universe. When these are taken into account, the unchanging block Universe view of space-time is best replaced by an evolving block Universe which extends as time evolves, with the potential of the future continually becoming the certainty of the past; space-time itself evolves, as do the entities within it. However this time evolution is not related to any preferred surfaces in space-time; rather it is associated with the evolution of proper time along families of world lines.

4.1 The Block Universe

The standard space-time diagrams used in representing the nature of space and time present a view of the entire space-time, with no special status accorded to the present time; indeed the present ("now") is not usually even denoted in the diagram. Rather all possible "present times" are simultaneously represented in these diagrams on an equal basis. This is the usual space-time view associated both with special relativity (when gravity is negligible, see e.g. Ellis and Williams 2000) and with general relativity (when gravity is taken into account, see e.g. Hawking and Ellis 1973). When the Einstein field equations have as the source of curvature either a vacuum (possibly with a cosmological constant) or simple matter (e.g. a perfect fluid, an electromagnetic field, or a scalar field), everything that occurs at earlier and later times is locally known from the initial data at an arbitrary time, evolved according

[1] Email address: ellis@maths.uct.ac.za

V. Petkov (ed.), *Relativity and the Dimensionality of the World*, 49–79.

to time-reversible local physics; hence there is nothing special about any particular time. In a few cases time irreversible physics is taken into account (e.g., nucleosynthesis in the early Universe), but the notion of the present as a special time is still absent.

This view can be formalised in the idea of a *block universe* (Mellor 1998; Savitt 2001; Davies 2002)[2]: space and time are represented as merged into an unchanging space-time entity, with no particular space sections identified as the present and no evolution of space-time taking place. The Universe just is: a fixed space-time block. In effect this representation embodies the idea that time is an illusion: it does not "roll on" in this picture. All past and future times are equally present, and there is nothing special about the present ("now"). There are Newtonian, special relativity, and general relativity versions of this view (see Figs. 4.1–4.4), the latter being most realistic as it is both relativistic and includes gravity.[3]

The warrant for this view in the case of special relativity is the existence and uniqueness theorems for the relevant fields on a fixed Minkowski background space-time; for example, the existence and uniqueness theorems for fluid flows, for Maxwell's equations, or for the Klein Gordon equation (see Hadamard 1923; Wald 1984: 243–252). In the case where gravity is significant, the warrant is the

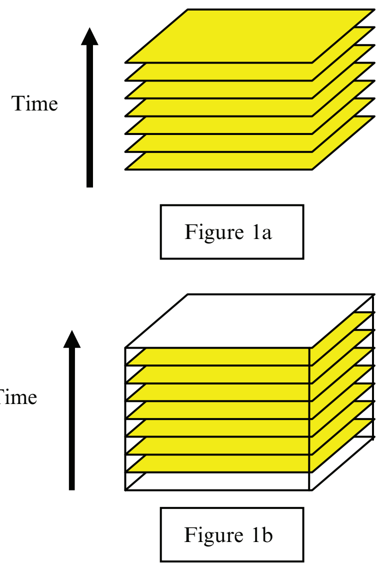

Fig. 4.1 Newtonian spaces (Fig. 4.1a) stacked together to make a Newtonian space-time (Fig. 4.1b) (see Ellis and Williams 2000). Thus this is like plywood: the grain out of which it is constructed remains as preferred space sections (i.e. surfaces of constant time) in space-time. Data on any of these surfaces determine the evolution of physics to the whole space-time (through time-reversible microphysics).

[2] And see Wikipedia: http://en.wikipedia.org/wiki/Block_time for a nice introduction.

[3] We do not consider here the possible variants when quantum gravity is taken into account.

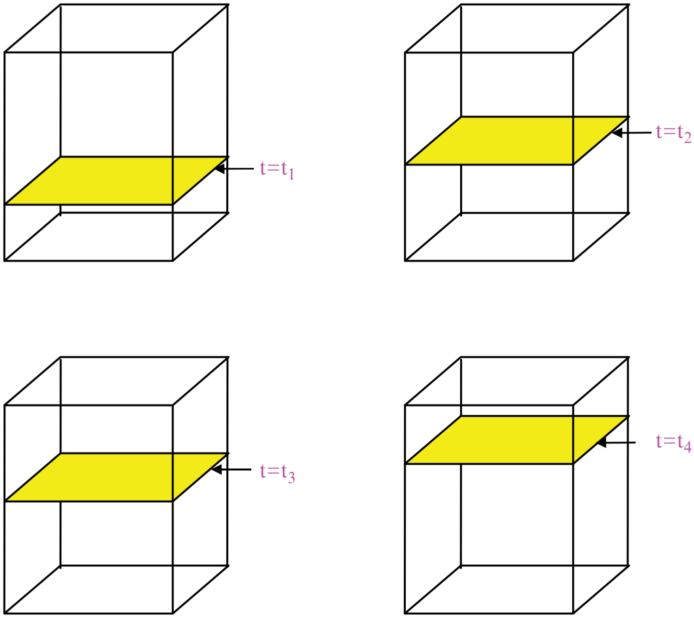

Fig. 4.2 Different parallel time surfaces in a special relativity "block space-time" where space-time is a block with a set of spatial sections indicating different specific times. These are just different slicings of the same immutable space-time, but they are not engrained in its structure; it is like a block of glass with no preferred sections. Data on any of these time surfaces determines the evolution of physics to the whole space-time (through time-reversible microphysics).

existence and uniqueness theorems of general relativity for suitable matter fields (Hawking and Ellis 1973: 226–255; Wald 1984: 252–267). They show that for such matter, initial data at an arbitrary time determine all physical evolution, including that of the space-time structure, to the past and the future equally, because we can predict and retrodict from that data up to the Cauchy horizon. The present time has no particular significance; it is just a convenient time surface we chose on which to consider the initial data for the Universe. We could have equally chosen any other such surface.

4.2 The Unfolding of Time

This block view is however an unrealistic picture because it does not take complex physics or biology seriously; and they do indeed exist in the real Universe. The irreversible flow of time is one of the dominant features of biology, as well as of the physics of complex interactions and indeed our own human experience (Le Poidevin 2004). Its associated effects are very significant on small scales (cells to ecosystems), though they are probably unimportant on large scales (galaxies and above).

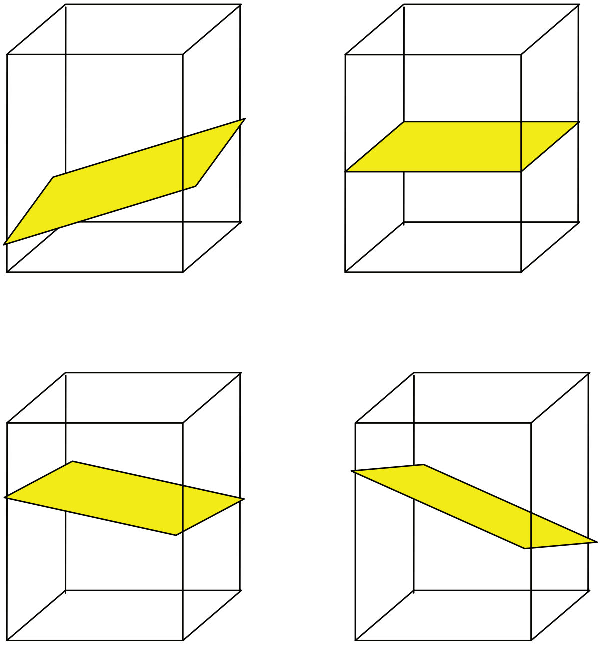

Fig. 4.3 Time surfaces in a special relativity "block space-time" are not unique (Ellis and Williams 2000; Lockwood 2005): they depend on the motion of the observer. Many other families can be chosen, and "space" has no special meaning. Space-time is the basic object. This is one of the justifications for the block universe picture: we can slice this immutable space-time in many ways.

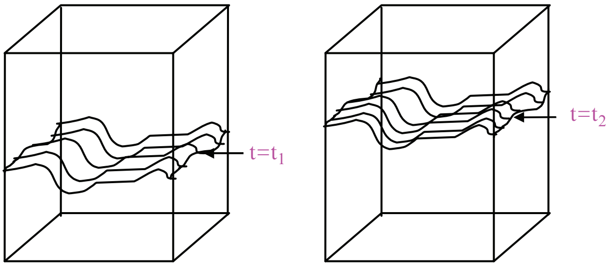

Fig. 4.4 Different time surfaces in a curved block space-time. General relativity allows any "time" surfaces that intersect all world lines locally. The space-time itself is also curved. Future and past physics, including the space-time itself, are locally determined from the data on any such surface.

This scale-dependence is a key feature, intimately related to the question of *averaging scales* in physics: every description used in any physical theory, including every space-time description, involves an explicit or implicit averaging scale (Ellis 1984; Ellis and Buchert 2005). To deal adequately with complex structures in a space-time, one must be very clear what averaging scale is being used. The flow of time is very apparent at some scales (e.g. that of biology), and not apparent at others (e.g. that of classical fundamental physics).

Classical microphysics is time reversible: detailed predictability to the past and future is in principle possible. It is in this case that "the present" may be claimed to have no particular meaning. However, the numbers of interactions involved, together with the existence of chaotic systems, can make detailed prediction impracticable in practice, leading to the use of statistical descriptions: we can predict the kinds of things that will happen, but not the specific outcome. Time-irreversible macrophysics and biology is based in the microphysics, but with emergent properties that often involve an overt "flow of time" and associated increase of entropy. The past is fixed forever, and can in principle be largely known; the future is unknown and mostly unpredictable in detail. The present is more real than the undetermined future, in that it is where action is now taking place: it is where the uncertain future becomes the immutable past. Various views are possible on how to relate these different aspects:

> There have been three major theories of time's flow. The first, and most popular among physicists, is that the flow is an illusion, the product of a faulty metaphor. The second is that it is not an illusion but rather is subjective, being deeply ingrained due to the nature of our minds. The third is that it is objective, a feature of the mind-independent reality that is to be found in, say, today's scientific laws, or, if it has been missed there, then in future scientific laws … Some dynamic theorists argue that the boundary separating the future from the past is the moment at which that which was undetermined becomes determined, and so "becoming" has the same meaning as "becoming determined". (Fieser and Dowden 2006, Section 7.)

The first and second views are those associated with the block universe picture and usual space-time diagrams. Can one find a space-time view supporting the third position? Yes one can; this is what we present below. Before doing so we first look more closely at a realistic view of the physics involved.

4.2.1 A Broken Wine Glass, Coarse-graining

A classic example of an irreversible process is the breaking of a wine glass when it falls from a table to the floor (Penrose 1989: 304–309). Now the key point for my argument here is that the precise outcome (the specific set of glass shards that result and their positions on the ground) is unpredictable: as you watch it fall, you cannot foretell what will be the fragmentation of the glass. You cannot predict what will happen at this level of detail because a macro-description of the situation (the initial shape, size, position, and motion of the glass) does not have enough detail of its microproperties (e.g. defects in its structure) to work this out.

The underlying physics is deterministic but our classical predictive model of what happens is not. It just might be possible to determine the fracturing that will occur if you have a detailed description of the crystalline structure of the glass, but that data – which in any case would be extremely difficult to obtain – is not available in a macro-description. From the macro-viewpoint what happens is random; you can give a statistical prediction of the likely outcome, but not a detailed definite prediction of the unique actual outcome. From a classical micro-viewpoint it is deterministic – you just need enough data and computing power to find out what will happen; but the macro-description and associated space-time picture does not contain that detailed information. *You can only find out what happens by watching it happen*; the physical result (e.g. the specific shapes and positions of the shards) unfolds as time progresses. Furthermore, this lack of predictability holds in both time directions. Considering the backward direction of time, you cannot reconstruct the details of the process of destruction (what happened when) from the fragments on the ground, because you cannot tell when the glass fell by looking at the resulting fragments. Even if we accessed all the micro-data available at late times, that uncertainty would remain: no amount of data collection will resolve it, once the thermal traces of the fall have dissipated and merged into the background noise.

Similar results will hold for example, for the explosion of a bomb: the distribution of fragments of a bomb that will occur is not predictable from macro-data available by external observation (see Fig. 4.5).

Generally in relating the description of a physical system at different scales, the microscopic data $(\mathrm{x_i},\mathrm{p_j})$ needed for the detailed phase space ($\mathrm{x_i}$ are position coordinates and $\mathrm{p_i}$ the corresponding momenta) is coarse-grained to give macroscopic variables $(\mathrm{X_i},\mathrm{P_j})$ characterising the phase space associated with the averaged macro-description of the system. If the averaging operator is A, then

$$A : (\mathrm{x_i}, \mathrm{p_j}) \rightarrow (\mathrm{X_i}, \mathrm{P_j}). \tag{4.1}$$

Many different micro-states correspond to the same macro-state, which is the source of entropy (Penrose 1989: 309–314). The micro-dynamics, usually given by Hamilton's equations (Penrose 1989:175–184), is governed by an operator φ_t giving the change over the time interval t:

$$\varphi_\mathrm{t} : (\mathrm{x_i}, \mathrm{p_j}) \rightarrow (\mathrm{x'_i}, \mathrm{p'_j}) = (\varphi_\mathrm{t}\mathrm{x_i}, \varphi_\mathrm{t}\mathrm{p_j}). \tag{4.2}$$

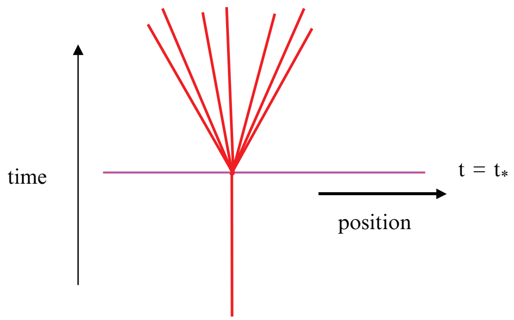

Fig. 4.5 The impossibility of prediction in the real world: space-time diagrams of the explosion of a bomb at time $t = t*$. The macro-state discription for $t < t*$ does not detemine the number of fragments and their motion for $t > t*$.

and the macro-dynamics by a corresponding operator Φ_t. The micro-variables will be subject to dynamic and structural constraints:

$$\mathrm{C}_1(\mathrm{x}_\mathrm{i}, \mathrm{p}_\mathrm{j}) = 0 \quad (\partial \mathrm{C}_\mathrm{i}/\partial \mathrm{p}_\mathrm{j} \neq 0), \quad \mathrm{C}_2(\mathrm{x}_\mathrm{i}) = 0, \tag{4.3}$$

the latter describing the structural relations of the system (e.g. the crystal structure of the glass). These constraints are preserved by the dynamics in Eq. (4.2):

$$\mathrm{C}_1(\mathrm{x}'_\mathrm{i}, \mathrm{p}'_\mathrm{j}) = 0, \quad \mathrm{C}_2(\mathrm{x}'_\mathrm{i}) = 0. \tag{4.4}$$

However, in some cases, different initial micro-states that correspond to the same initial macro-state result in different final macro-outcomes, because the dynamics and averaging do not commute:

$$\mathrm{A}\varphi_\mathrm{t} \neq \Phi_\mathrm{t}\, \mathrm{A}. \tag{4.5}$$

Then detailed micro-level predictability determines the macro-level outcomes but does not lead to reliable emergent macro-level behaviour: indeed then Φ_t is not well defined (see Ellis 2006b: Fig. 4.5). This will be the case for example, when chaotic behaviour occurs (Thomson and Stewart 1987). Furthermore the structural constraints only hold for a limited range of values of the dynamic variables: if these bounds are exceeded, the constraints will be violated:

$$\mathrm{C}_2(\mathrm{x}'_\mathrm{i}) \neq 0 \tag{4.6}$$

(a glass breaks or bomb goes off as in the examples above, or a phase change takes place). The way this happens is not described by the dynamics of Eq. (4.2), which assume these constraints are preserved, and is invalid otherwise. A dynamical analysis is required that covers the change of constraints and resulting new dynamics; but even phase changes from water to ice are not fully predictable at present (Laughlin 2005: Chapter 4).

4.2.2 Friction, Coarse-graining

In general, friction effects mean we have an inability to retrodict if we lose information below some level of coarse graining. The simplest example is a block of mass m sliding on a plane, slowing down due to constant limiting friction $F = -\mu R$ where μ is the coefficient of friction and $R = mg$ is the normal reaction, where g is the acceleration due to gravity (Spiegel 1967). The motion is a uniform deceleration; if we consider the block's motion from an initial time $t = 0$, it comes to rest at some later time $t_* > 0$. For $t < t_*$ the velocity v and position x of the object are given by

$$v_1(t) = -\mu g t + v_0, \quad x_1(t) = -{}^1\!/\!_2\,\mu g t^2 + v_0 t + x_0 \tag{4.7}$$

where (v_0, x_0) are the initial data for (v, x) at the time $t = 0$. This expression shows that it comes to rest at $t_* = \mu g / v_0$. For $t > t_*$, the quantities v and x are given by

$$v_2(t) = 0, \quad x_2(t) = \mathrm{X}\ (\text{constant}), \tag{4.8}$$

where $\mathrm{X} = -\,{}^1\!/_2 \mu g t_*^2 + v_0 t_* + x_0$.

The key point now is that from the later data of Eq. (4.8) at any time $t > t_*$ you cannot determine the initial data (v_0, x_0), nor the time t_* when the object came to rest, thus you cannot reconstruct the trajectory of Eq. (4.7) from that data. You cannot even tell if the block came from the left or the right (see Fig. 4.6).

If you could measure the distribution of heat in the table top soon enough after the block stopped, through thermal imaging for example, you could work out what had happened (because it's motion will have been converted into heat). Thus, the inability to retrodict soon after the block comes to rest is again a result of using a macro picture that does not include all the detailed data: here, the thermal motion of particles in the table top (which will then dissipate away and be lost in the environment; this data too will soon become irretrievably lost). This is of course the essence of successful physics models: using a simplified picture that throws most of the detailed data away, and concentrating on essentials. The cost is that the ability of the model to predict is strictly limited.

Similar results hold for any viscous processes or dissipative system: the final state will generically be an attractor; once it has settled down you cannot tell from macro data what initial state the system came from – it could have been any point in the basin of attraction (Thompson and Stewart 1987).

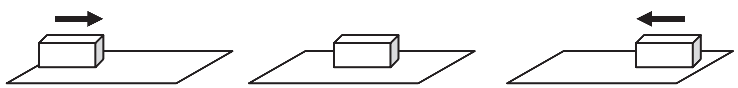

Fig. 4.6a The impossibility of retrodiction in the real world: a block sliding on a surface. The stationary block in the centre might have come from the left, and stopped under friction, or from the right. Observing it at rest does not tell us which was the case.

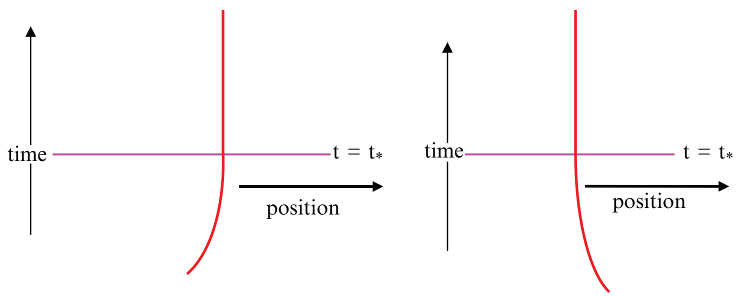

Fig. 4.6b Space-time diagrams of the position of the block for the two cases illustrated above. The state for $t > t*$ does not determine the state for $t > t*$, or even determine the value $t*$.

4.2.3 Quantum Uncertainty

In these examples, our inability to predict is associated with a lack of detailed information. So if we fine-grained to the smallest possible scales and collected all the available data, could we then determine uniquely what is going to happen? No, we cannot predict to the future in this way because of foundational quantum uncertainty relations (see e.g. Feynman 1985; Penrose 1989; Isham 1997), apparent for example, in radioactive decay (we cannot predict precisely when a nucleus will decay and what the velocities of the resultant particles will be) and the motion of a stream of particles through a pair of slits onto a screen (we cannot predict precisely where a photon or electron will end up on the screen). It is a fundamental aspect of quantum theory that this uncertainty is unresolvable: *it is not even in principle possible to obtain enough data to determine a unique outcome of quantum events.* This unpredictability is not a result of a lack of information: it is the very nature of the underlying physics.

More formally: if a measurement of an observable $\boldsymbol{A}$ takes place at time $t = t_*$, initially the wave function $\psi(x)$ is a linear combination of eigenfunctions $\mathrm{u}_n(x)$ of the operator $\tilde{\boldsymbol{A}}$ that represents $\boldsymbol{A}$: for $t < t_*$, the wave function is

$$\psi_1(x) = \Sigma_\mathrm{n} \psi_\mathrm{n} \mathrm{u}_\mathrm{n}(x). \tag{4.9}$$

(see e.g. Isham 1997: 5–7). But immediately after the measurement has taken place, the wave function is an eigenfunction of $\tilde{\boldsymbol{A}}$: it is

$$\psi_2(x) = \mathrm{a_N u_N}(x) \tag{4.10}$$

for some specific value N. The data for $t < t_*$ do not determine the index N; they just determine a probability for the choice N. One can think of this as due to the probabilistic time-irreversible collapse of the wave function (Penrose 1989: 260–263). Invoking a many-worlds description (see e.g. Isham 1997) will not help: in the actually experienced Universe in which we make the measurement, N is unpredictable. Thus the initial state Eq. (4.9) does not uniquely determine the final state of Eq. (4.10); and this is not due to lack of data, it is due to the foundational nature of quantum interactions. You can predict the statistics of what is likely to happen but not the unique actual physical outcome, which unfolds in an unpredictable way as time progresses; you can only find out what this outcome is after it has happened. Furthermore, in general *the time t_* is also not predictable from the initial data*: you do not know when "collapse of the wave function" (the transition from Eq. (4.9) to Eq. (4.10)) will happen (you cannot predict when a specific excited atom will emit a photon, or a radioactive particle will decay).

We also cannot retrodict to the past at the quantum level, because once the wave function has collapsed to an eigenstate we cannot tell from its final state what it was before the measurement. You cannot retrodict uniquely from the state of Eq. (4.10) immediately after the measurement takes place, or from any later state that it then evolves to via the Schrodinger equation at later times $t > t_*$, because knowledge of these later states does not suffice to determine the initial state of Eq. (4.9) at times $t < t_*$: the set of quantities ψ_n are not determined by the single number a_N.

The fact that such events happen at the quantum level does not prevent them from having macro-level effects. Many systems can act to amplify them to macro-levels, including photomultipliers[4] (whose output can be used in computers or electronic control systems). Quantum fluctuations can change the genetic inheritance of animals (Percival 1991) and so influence the course of evolutionary history on Earth. Indeed that is in effect what occurred when cosmic rays[5] – whose emission processes are subject to quantum uncertainty – caused genetic damage in the distant past:

> The near universality of specialized mechanisms for DNA repair, including repair of specifically radiation induced damage, from prokaryotes to humans, suggests that the earth has always been subject to damage/repair events above the rate of intrinsic replication errors ….. radiation may have been the dominant generator of genetic diversity in the terrestrial past. (Scalo et al. 2001)[6]

Consequently *the specific evolutionary outcomes on life on Earth (the existence of dinosaurs, giraffes, humans) cannot even in principle be uniquely determined by causal evolution from conditions in the early Universe, or from detailed data at the start of life on Earth.* Quantum uncertainty prevents this, because it significantly affected the occurrence of radiation-induced mutations in this evolutionary history. The specific outcome that actually occurred was determined as it happened, when quantum emission of the relevant photons took place: the prior uncertainty in their trajectories was resolved by the historical occurrence of the emission event, resulting in a specific photon emission time and trajectory that was not determined beforehand, with consequent damage to a specific gene in a particular cell at a particular time and place that cannot be predicted even in principle.

4.2.4 Space-Time Curvature: Time-dependent Equations of State

So far we have considered unpredictability of the evolution of local systems in a fixed space-time; this needs to be taken into account in our space-time pictures of such interactions. But can this uncertainty affect the nature of space-time itself? Yes indeed; in general relativity theory, matter curves space-time, and the curvature of space-time then affects the motion of matter (Hawking and Ellis 1973; Misner et al. 1973; Wald 1984). We can have unpredictability at both stages of the non-linear interaction that determines the future space-time curvature.

First, as regards matter determining space-time curvature: can unpredictable local processes have gravitational effects that in turn affect space-time curvature? Hermann Bondi (1965) posited a pair of orbiting massive objects ("Tweedledum"

---

[4] See Wikipedia: http://en.wikipedia.org/wiki/Photomultiplier for their physical realisation.

[5] See http://www.chicos.caltech.edu/cosmic_rays.html for a brief summary of their origin.

[6] See also, for example, (Babcock and Collins 1929; Rothschild 1999; National Academy 2005).

and "Tweedledee") where internal batteries drive motors slowly altering the shape of the bodies from oblate to prolate spheroids and back, thereby changing their external gravitational field. This enables exchange of energy and information between them via gravitational induction: time-dependent terms in the field equations and Bianchi identities are negligible. Time variation in the source term in the constraint equations conveys energy from one body to the other by altering the electric part E_{ab} of the Weyl conformal curvature tensor (the "free gravitational field") in the intervening space-time.[7] This then changes the tidal source term E_{ab} in the deviation equation[8] for matter in the second body, altering its shape. Information can be conveyed between them by altering the time pattern of these variations: computer control of the motors allows an arbitrary signal to be transferred between Tweedledum and Tweedledee, that cannot be predicted from the initial data for the gravitational field (it is specified by the computer programme). This unpredictability is a result of the implicit coarse-grained description of the physical system: changes in space-time curvature occur that cannot be predicted from external view of the objects because that description does not include details of the internal mechanisms, including the specific bits making up the stored computer programme (these would be represented at a much finer level of description). One can have similar processes involving gravitational radiation. Consider two identical spherical masses at the end of a strong rod, able to turn about a vertical axis. An electric motor rotates the rod, and is controlled via a computer to turn the rod at high angular speeds ω_i for a series of time intervals T_i separated by stationary intervals t_i. This creates a time-dependent gravitational dipole that will emit gravitational waves according to standard formulae (Misner et al. 1973: Chapter 36), with oscillations in the electric and magnetic parts of the Weyl tensor carrying energy and information from one place to another during each interval T_i. This time-dependent field can in principle be detected by a distant gravitational wave detector.

Second, the motion of matter can be affected in a similar manner. Suppose we attached a large number of massive rocket engines to one side of the Moon and fired them simultaneously. This would change the orbit of the Moon (for a while its motion would be non-geodesic) in a way that is cumulative with time. This then would affect the way it curved space-time in the future, for its future position relative to the Earth would be different from what it would otherwise have been. The local Weyl tensor will have been altered and so tides on the Earth would be altered. Thus such engineering efforts can change the future space-time curvature and its physical effects. Again computer control allows an arbitrary time evolution to be specified.

Generically the point is that *explicitly time-dependent equations of state can affect the future development of space-time, and how this will work out is unpredictable from the macro initial data at any specific time*. If the spacetime description were detailed enough to include the classical mechanisms involved in such physical causation (the clocks, computers, motors, rocket engines, etc. that caused such changes, including the computer programme) then they might be predictable; but a

[7] See Ellis (1971) for the relevant equations (the "div E" Bianchi identities).

[8] The generalisation of the geodesic deviation equation to non-geodesic motion.

standard space-time picture does not include this detailed data. Furthermore, *such mechanisms could include a random element making such prediction impossible even in principle*. One might for example arrange for the computer programme to use as input, signals from either a particle detector that detects the particles emitted through radioactive decay of unstable atoms, or a photon detector that responds to individual photons from a distant quasar. Then quantum uncertainty in the particle emission process would prevent precise prediction of the future space-time curvature, even in principle.[9] In effect we would in these cases be amplifying quantum uncertainty to astronomical scales.

Human intentionality underlies the unpredictable functioning of the mechanisms (motors, computers, etc.) considered above, as they would be the result of human agency (implied by their supposed existence as designed objects). However this kind of effect can occur in other contexts without human intervention, indeed it has already happened in the expanding Universe at very early times. According to the standard inflationary model of the very early Universe, we cannot predict the specific large-scale structure existing in the Universe today from data at the start of the inflationary expansion epoch, because density inhomogeneities at later times have grown out of random quantum fluctuations in the effective scalar field that is dominant at very early times:

> Inflation offers an explanation for the clumpiness of matter in the universe: quantum fluctuations in the mysterious substance that powered the [inflationary] expansion would have been inflated to astrophysical scales and therefore served as the seeds of stars and galaxies. (Hinshaw 2006)[10]

Thus *the existence of our specific Galaxy, let alone the planet Earth, was not uniquely determined by initial data in the very early Universe*. The quantum fluctuations that are amplified to galactic scale by this process are unpredictable in principle.

4.2.5 Emergent Complexity and Human Intentions

The very nature of emergent complexity means that microphysical effects do not by themselves always uniquely determine the future at a macrophysical level, because their outcomes are critically context-dependent in a way that can even alter the nature of the lower-level interactions (Ellis 2006a). Higher levels of emergent complexity are causally effective, in the biological case their outcome being determined by stored information, feedback control loops, and associated goals that get

[9] It has been suggested to me that in view of these outcomes, one should query the uncertainty principle of quantum theory, seeking a deterministic version instead, as suggested for example by David Bohm and David Gross. Such a theory may indeed become widely accepted one day, and then what is presented here would have to be revised; until this happens, the prudent view is to provisionally accept one of the major outcomes of physics last century, namely that quantum uncertainty is indeed real.

[10] See Kolb and Turner 1990 or Dodelson 2003 for details.

developed over time via Darwinian evolutionary processes (Roederer 2005). This functionality is not encapsulated in any macro-description of such systems (because the non-commutativity described by Eq. (4.5) above holds), and so enables physical behaviour to occur that is not predictable from available physical initial data. Physics by itself cannot predict either plant development or animal behaviour, for they depend crucially on biological information and context.

Furthermore human intentions are causally effective (Martin et al. 2003; Kane 2005). They are the result of conscious brain processes embodying higher-level intentions and abstract concepts, some being socially determined (Ellis 2005a,b, 2006a). In many contexts, it is our choices that determine which possible future is realised as "now" becomes "then". Indeed this is one of the most fundamental features of our lives (Le Poidevin 2004): intention changes the future; the past is fixed forever and cannot be changed, as stated poignantly in *The Rubaiyat of Omar Khayyam* (Fitgerald 1989):

> *The Moving Finger writes; and having writ,*
> *Moves on; nor all your Piety nor Wit*
> *Shall lure it back to cancel half a Line,*
> *Nor all your Tears wash out a Word of it.*

This aspect is missing in the reversible microphysics picture and associated block universe descriptions, which do not represent adequately the nature of macrophysics or biology. One can claim that our inability to predict the outcome of human actions is not just a question of being unable to do so for computational reasons. Rather, one can argue that *the results of human agency are unpredictable even in principle from initial physical data* (Ellis 2005a,b, 2006a), *and this uncertainty should be reflected in representations of the evolution of physical systems from the initial data.*

This is clearly a contentious claim, and the overall results of this paper are not critically dependent on the claims of the present section; nevertheless in my view this is an important issue in the overall causal nexus, reinforcing in a different way the arguments of the previous sections. However, we do note here that *it is this freedom that underlies the very existence of physics as a science*, for it is our own choice that enables us firstly to devise theories such as quantum electrodynamics or general relativity, secondly to set up experiments to test such theories, and thirdly to analyse the results of the experiments and see if they support the theories or not (Ellis 2006a: section 7.1). This is the unseen factor that is taken for granted in all the physical sciences: the ability of the experimenter to experiment. It is taken for granted in the examples above: in the existence of a wine glass or a bomb (section 2.1) and in the motion of the block (section 2.2), for example. In the latter case, the block would have been stationary initially and then set in motion by the experimenter at some time t_{i}. By observing the block at some earlier time $t_1 < t_{\mathrm{i}}$, you cannot predict the time t_{i} or the subsequent motion (Fig. 4.7), because that is the result of human volition. The same kind of issue arises in the very existence as well as the operational use of CERN, the Hubble Space Telescope, and every other scientific enterprise.

This kind of unpredictability is not related only to human actions: it also occurs for animals that are not self-conscious in the way humans are, for example, spiders

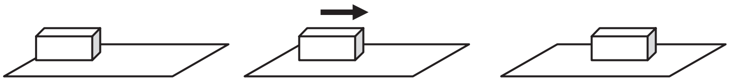

Fig. 4.7a The impossibility of prediction in the real world: a block sliding on a surface and then coming to rest. The stationary block in the left will be made to move by an impulse at a time t_{i} chosen by an experimenter. Observing it at rest at earlier times does not tell us when this will happen – or indeed that it will happen at all.

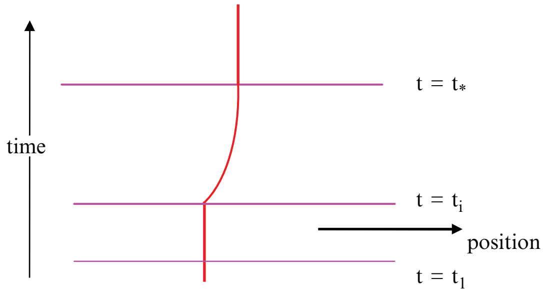

Fig. 4.7b Space-time diagrams of the position of the block for the case illustrated above. The state for any time $t_1 < t_{\mathrm{i}}$ does not determine the state for $t > t_{\mathrm{i}}$, or even determine the value t_{i}.

building webs and bees dancing to signal food sources to other bees. Indeed jumping spiders demonstrate mental representations and planning abilities (Prete 2004: 5–40), and bees can follow symbolic cues (Prete 2004: 41–74). In none of these cases does the initial physics data alone at some previous space-time epoch, say the time of decoupling of matter and radiation in the early Universe, determine the results of these later acts of living beings. That initial data set a possibility space for what will happen in the future but does not uniquely determine the outcome (Ellis 2005a,b), for example, the existence of neither specific beaver dams in Canada nor the Hubble Space Telescope was somehow uniquely encoded into the initial data at that time. Even statistical prediction does not apply in these cases, because emerging order and information are not statistically determined when they embody a purposeful direction: it is goals rather than statistics that determines the outcome (Ellis 2006a).

The physical freedom allowing this emergent order to have autonomous causal powers at the macro-level is probably related to the existence of quantum uncertainty at the micro level, which breaks the tight grip of Laplacian physical causation on the future and allows information to be selected by Darwinian evolutionary processes and then become causally effective (Roederer 2005; Ellis 2006a).

Whether or not one accepts the argument given above in this section, the very existence as a human artefact of the article you are presently reading certainly cannot be predicted from initial data in the very early Universe because neither the existence of the specific planet Earth, nor of any human beings at all on Earth, is guaranteed by the details of that initial data (see sections 2.3 and 2.4 above). Consequently *the* *specific outcomes of the actions of any particular human being on the planet Earth–such as the words of this article–certainly cannot be uniquely implied by that data.*

4.2.6 Overall: A Lack of Predictability in the Real Universe

In summary, *the future* is not determined till it happens because of

1. Time-dependent equations of state, which can be information driven
2. Quantum uncertainty, which can be amplified to macro-scales
3. Emergent complexity, including animal and human agency, and also in practice by
   a. Statistics/experimental errors/classical fluctuations, amplified by
      i. Chaotic dynamics/occurrence of catastrophes.

Essentially all realistic models of the Universe except for very large-scale cosmology are non-deterministic. Sufficient reasons for this are:

1. Coarse-graining by its very nature introduces a statistical element
2. Quantum processes occur on the small scale, and can be amplified to macro scales, so there is no deterministic microscopic model from which fully predictive classical macroscopic models can always be derived.

Because of these effects, we cannot predict uniquely to the future from present-day data; indeed their detailed features remain open and causally undetermined by initial conditions. Statistical prediction is however possible in contexts where emergent complexity, and in particular biological agency, is unimportant (it is not necessarily useful when Darwinian selection effects or human agency are active, for example, predicting the probability of existence of giraffes or ostriches at present, or of specific-endangered species in the future). This determines the kinds of thing that will happen, but not the specific outcome that actually occurs.

The past has happened and is fixed, so the nature of its existence is quite different than that of the indeterminate future. However, we cannot causally retrodict uniquely to the past from present day initial data by using the appropriate evolution equations for the matter, because of friction and other dissipative effects (see section 2.2) and the quantum measurement process ("collapse of the wave function", see section 2.3). What we can do, is observe present-day features resulting from past events (geological and archaeological data, photographic images, written records of past events, etc.), and thereby attempt to determine what in fact occurred by analysing these observations in conjunction with the dynamic projection of local physics from present data to the past (Lockwood 2005: 233–256).[11]

[11] There is uncertainty as regards both the future and the past, but its nature is quite different in these two cases. The future is uncertain because it is not yet determined: it does not yet exist in a physical sense. Thus this uncertainty has an ontological character. The past is fixed and unchanging because it has already happened, and the time when it happened cannot be revisited; but our knowledge about it is incomplete, and can change with time. Thus, this uncertainty is epistemological in nature.

4.3 A Realistic Space-Time Picture

The time-reversible picture of fundamental physics underlying the block universe viewpoint simply does not take these kinds of phenomena into account, specifically because it does not take cognisance of how complex phenomena arise from the underlying microphysics, with the emergence of the arrow of time. It does not take seriously the physics and biology of the real world but rather represents an idealised view of things which is reasonably accurate on certain (very large) scales where very simplified descriptions are successful.

In order to take the physical situations considered in the previous section into account, we need to modify the block universe pictures so as to adequately represent causation in these contexts. How do we envisage space-time and the objects in it as time unrolls? A way to do this is to consider an *evolving block universe* ("EBU") model of reality, with space-time ever growing and incorporating more events as time evolves along each world line.

To motivate this, consider the following scenario (Fig. 4.7): A massive object has rocket engines attached at each end to make it move either left or right. The engines are controlled by a computer that decides what firing intervals are utilised alternately by each engine, on the basis of a non-linear time-dependent transformation of signals received from a detector measuring particle arrivals due to random decays of a radioactive element. These signals at each instant determine what actually happens from the set of all possible outcomes, thus determining the actual space-time path of the object from the set of all possible paths (Fig. 4.8). This outcome is not determined by initial data at any previous time, because of quantum uncertainty in

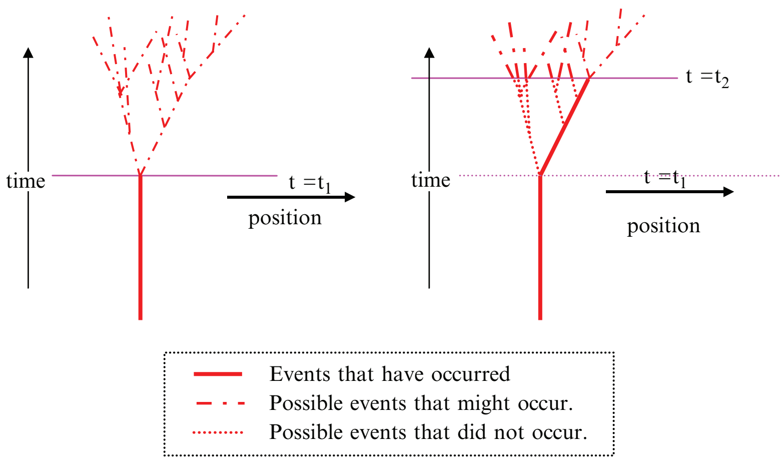

Fig. 4.8 Motion of a particle world line controlled in a random way, so that what happens is determined as it happens. On the left events are determined till time t_1 but not thereafter; on the right, events are determined till time $t_2 > t_1$, but not thereafter.[12]

[12] See Lockwood (2005), Figure 1.1 (page 12).

the radioactive decays.[13] As the objects are massive and hence, cause space-time curvature, the space-time structure itself is undetermined until the object's motion is determined in this way. Instant by instant, the space-time structure changes from indeterminate (i.e. not yet determined out of all the possible options) to definite (i.e. determined by the specific physical processes outlined above). Thus a definite space-time structure comes into being as time evolves. It is unknown and unpredictable before it is determined.

The evolving block universe model of space-time represents this kind of situation, showing how time progresses, events happen, and history is shaped. Things could have been different, but second by second, one specific evolutionary history out of all the possibilities is chosen, takes place, and gets cast in stone. This idea was proposed many years ago (Broad 1923),[14] but has not caught on in the physics community.[15]

We now consider it successively in the contexts of Newtonian theory, special relativity, and general relativity.

4.3.1 The Newtonian Case

Here we consider events in space-time as evolving from indefinite to determinate as time passes; the past is fixed and immutable, and hence, has a completely different status than the future, which is still undetermined and open to influence. The kinds of "existence" they represent are quite different: the future only exists as a potentiality rather than an actuality. The existential nature of the present is indeed unlike that of the past or future, for it is the set events we can (according to Newtonian theory) actually influence at any instant in our history.

We can represent this through a growing space-time diagram with unique time surfaces (Fig. 4.9), the passing of time marking how things change from being indefinite (and so not yet existing) to definite (and so having come into being), with the present marking the instant when we can act and change things.

In this case we can associate the passing of time uniquely with the preferred spatial sections of Newtonian space-times, representing Newton's absolute time. "The present" exists and is unique. However, *although the events in it are uncertain, the nature of the future to-be space-time is known and immutable*. We do not have to engage in uncertain prediction in order to know what it will be.

[13] In effect this diagram shows the multiple options of the Everett–Wheeler branching Universe view (Isham 1997), but with specific choices made as the wave function collapses time after time (Penrose 1989), resulting in an emerging unique outcome as time progresses.

[14] For quotes from Broad's book in this regard, see Ted Sider's notes at http://fas-philosophy.rutgers.edu/sider/teaching/415/HO_growing_block.pdf.

[15] Tooley (1997) also puts forward such a theory in an interesting way, but (unlike what is presented here) proposes modifying special relativity for this purpose. This is clearly unlikely to gain acceptance.

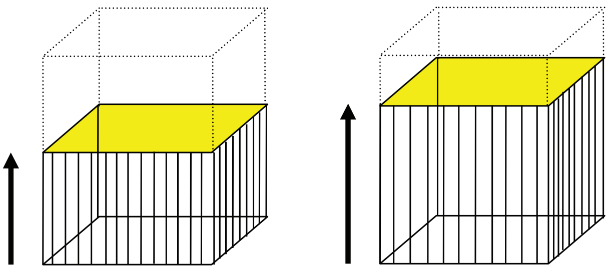

Fig. 4.9 Time represented as evolving in a Universe where events emerge as choices are made. The future is indeterminate, the past is fixed; thus they have a completely different status. The determinate part of space-time extends into the future as time evolves and potentiality becomes reality (through time irreversible macrophysics with underlying quantum uncertainty). However, the nature of the future space-time is already fixed before events in it evolve, in the cases of Newtonian theory and special relativity.

4.3.2 Special Relativity

This is like the Newtonian case: we represent the past of each event as fixed and immutable, but the future nature of events as still undetermined. The present is where uncertainty about events changes to certainty. And as in Newtonian theory, the *nature* of the future space-time is known and immutable (it is just Minkowski space-time), even though the events in it are unknown.

However, the time surfaces are no longer invariant under change of reference frame: they depend on the observer's motion relative to the coordinate system (see Fig. 4.3). So the usual objection to the idea of a special relativistic evolving Universe is, *How do we choose which surfaces are associated with the evolution of space-time*? This choice is arbitrary, and so the unfolding of time is indeterminate: it is not a well-defined unique physical process. We need to turn to a world line-based picture, which is natural in general relativity, to get an answer.

4.3.3 General Relativity

In this case, the present is again represented as where the indeterminate nature of potential physical events changes to a definite outcome, but now *even the nature of the future space-time is taken to be uncertain until it is determined at that time*, along with the physical events that occur in it. A further major feature is that because space-time is curved, unlike the special relativity case, *in particular solutions of the Einstein equations there are in general geometrically and physically preferred spacelike surfaces and timelike world lines*, related to the specific physics of the situation. These represent broken symmetries in the solutions to the Einstein field equations: the solutions have less symmetry than the equations of the theory.

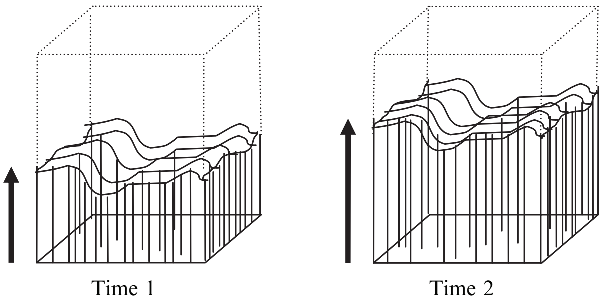

Fig. 4.10 An evolving curved space-time picture that takes macro-phenomena seriously. Time evolves along each world line, extending the determinate space-time as it does so (what might be changes into what has happened; indeterminate becomes determinate). The particular surfaces have no fundamental meaning and are just there for convenience (we need coordinates to describe what is happening). You cannot locally predict uniquely to either the future or the past from data on any "time" surface (even though the past is already determined). This is true both for physics, and (consequently) for the space-time itself: the developing nature of space-time is determined by the evolution (to the future) of the matter in it.

One can suggest that in this case, the transition from present to past does not take place on specific spacelike surfaces; rather *it takes place pointwise at each space-time event*. The implicit "now" of Figs. 4.9 and 4.10 (or of any "flowing time" concept) is replaced by a "here-now" (space-time point), and for both the "now" and the "here-now" the past is determined (exists relative to the [here]-now), the future is undetermined, and the [here]-now is a moment of passage from one state to the other.

However, the constraints on what future can emerge at a given here-now are not point-wise constraints but (in relation to any local coordinates) constraints involving spatial derivatives, or, roughly speaking, neighbouring points. So if evolution takes place pointwise, it still involves a degree of spatial coordination between neighbouring points, even though the neighbouring point might not "yet exist" relative to a different here-now until it lies in the past.

It is convenient to introduce local coordinates in order to determine how this works, involving a splitting of space-time into space and time as in the ADM formalism (Arnowitt, Deser, and Misner 1962; Misner et al. 1973: 520–528; Anninos 2001), and evolution along the coordinate lines introduced as determined by the shift function. But then it is physically sensible to focus on this feature: that is, while it may in a sense take place pointwise, it is more convenient to consider the *evolution as taking place along timelike world lines*,[16] rather than being determined by any universal time defined by spacelike surfaces. Indeed this is strongly suggested both by the way that time is determined in general relativity as a path integral along timelike world lines, and also by the nature of the examples discussed in the previous

[16] In most real-world contexts, even though it is in principle possible to exert causal influences along null lines (i.e. at the speed of light), this is in practice unimportant except for small-scale situations where lasers are important.

section, where physical effects that determine what happens are focused on changes that take place along histories of matter, represented by timelike worldlines.[17]

But then the question is, which world line should be chosen? The potential problem is the arbitrariness in the choice of the world lines. There seem to be two choices: either

(**ET1**) *we regard the evolution of time as being allowed to take place along any world lines whatever, none being preferred, or*

(**ET2**) *the evolution of time takes place along preferred world lines, associated with a symmetry breaking that leads to the emergence of time.*[18]

In the latter case, the question is, which world lines are the key ones that play this crucial physical role? In specific realistic physical situations, there will be preferred world lines associated with the average motion of matter present, as in the case of cosmology (Ellis 1971), and there will be preferred time surfaces associated with them if the matter flow is irrotational. This occurs in particular in the case of the idealised Robertson–Walker models of standard cosmology, where as long as matter is present, there exist uniquely preferred irrotational and shear-free world lines that are eigenvectors of the Ricci tensor. These then form a plausible best basis for description of physical events and the evolution of matter: there is a unique physical evolution determined along each such a family of world lines with its associated unique time surfaces, which are invariant under the space-time symmetries. Then one might propose that the evolution of time is associated with these preferred timelike world lines and perhaps associated spacelike surfaces, being an emergent property associated with the broken symmetries represented by these geometrical features in curved space-times.

But there may be several competing such choices of world lines in more realistic cases, for example, realistic perturbed cosmological models such as are needed for structure formation studies will have multiple-matter components present with differing 4-velocities (see e.g. Dunsby et al. 1992). For this reason, we might rather consider the evolution as taking place along *arbitrary families of world lines*, corresponding to the freedom of choice of the shift vector in the ADM formalism for general relativity (Arnowitt et al. 1962; Misner et al. 1973: 520–528; Anninos 2001).

A key result then is that no unique choice for these world lines needs to be made in the standard general relativity situation with simple equations of state; the ADM theory says *we locally get same result for the evolving space-time, whatever world lines are chosen.* You can choose any time lines you like to show how things will have evolved at different places (that is, on different observer's world lines) at different times (that is, at various proper times along those world lines). But this view has no foundationally preferred status: you could have chosen different world lines, corresponding to different shift vectors, and a different relation between times on the world lines, corresponding to different choices of the laps

[17] In those cases where radiation dominates, as in the early Universe before decoupling, the average motion of the radiation is represented by timelike world lines.

[18] And presumably to the arrow of time (Ellis and Sciama 1972; Davies 1974; Zeh 1992).

function; the resulting 4D space-time is the same. In any specific situation, some of those descriptions will be more natural and easier to use and understand than others; but this is just a convenience, and any other surfaces and world lines could have been chosen.

Thus, in the classical general relativity case, we get a consistent picture: things are as we experience them. Time rolls on along each world line; the past events on a world line are fixed and the future events on each world line are unknown. Space-time grows as in the Newtonian picture, but now even the space-time structure itself is to be determined as the evolution takes place (Fig. 4.10). The metric tensor determines the rate of change of time with respect to the coordinates, for this is the fundamental meaning of the metric (Hawking and Ellis 1973). A gauge condition determines how the coordinates are extended to the future. Conservation equations plus equations of state and associated evolution equations determine how matter and fields change to the future, including the behaviour of ideal clocks, which measure the passage of time. The field equations determine how the metric evolves with time, and hence, determine the future space-time curvature. The whole fits together in a consistent way, determining the evolution of both space-time and the matter and fields in it,[19] as is demonstrated for simple equations of state by the existence and uniqueness theorems of general relativity theory (Hawking and Ellis 1973: 226–255; Wald 1984: 252–267).

Thus, no unique choice needs to be made for the conventional ADM formalism, *which is deterministic*; for these standard theorems assume classical deterministic physics rules the micro-world. But the whole point of this paper is that most models are *not* deterministic, irreversible unpredictable processes, and emergent properties will take part in determining space-time curvature; on relatively small scales, even human activity does so (when we move massive objects around). A realistic extension of the above will take into account quantum uncertainty in the evolution of the matter and fields, giving a probability for the future evolution of particles, fields, and hence, for space-time, rather than a definite prediction.[20] Quantum evolution will determine the actual outcome that occurs in a probabilistic way (see the examples in section 2.4). But then the problem is, if two choices of world lines are made *in two different indeterministic futures*, then it is probable that the two evolutions will not agree. In such a scenario we would have something more like Wheeler's "many fingered time" – the different proper times along arbitrary world lines do not knit together to form a global concept of time that is meaningful in determining a unique evolution along all world lines. How then can an evolving block universe emerge from this situation?

In the following section, we consider how this might work out in the various indeterministic cases discussed above (in section 2).

[19] Smart's objections to an objective passage or flow of time (Smart 1967: 126) are comprehensively answered by Lockwood (2005: 13–18).

[20] This is what is done in the semi-classical calculations of inflationary Universe perturbations, for details see Kolb and Turner (1990); Dodelson (2003).

4.4 The Emergence of a Block Universe

The fundamental argument in this section is based on a simple observational fact:

> (**OF**) *In the real Universe domain we actually inhabit, a unique classical space-time structure does indeed emerge at macro-scales from the underlying physics.*

This is true for each of the cases discussed in section 2: in each case, we are indeed able to describe what happens via an evolving block universe. It is for this reason that we are able to regard special and general relativity as successful theories in their appropriate contexts within the local universe domain in which we live. Consequently we will take this as a fundamental observational fact about the real universe.[21]

The implication of (**OF**) is immediate:

> (**EB**) *Whatever conditions are needed to imply the existence of an emergent block Universe at macro-scales, whether related to a particular set of world lines as in* **(ET1)***, or the emergence of the same space-time whatever world lines are chosen as in* **(ET2)***, are satisfied in our real physical observable Universe domain.*

There could be other Universe domains or hypothetical Universes in which this is not true, in which for example, a classical space-time structure never emerges; but that is not the case on this Earth or indeed, as far as we can tell, within the visible Universe.

So even in those cases where we are at present unable to determine why this is the case or indeed what the required conditions are, it seems clear that *whatever integrability or consistency conditions are needed to guarantee the emergence of a growing block Universe are in fact satisfied in the real Universe we see around us*. That is the basic feature we assume in what follows.

4.4.1 Classical Cases

A main cause of indeterminism discussed above is coarse-graining (sections 2.1 and 2.2 above). Here there is an underlying deterministic theory where one can apply things like existence and uniqueness theorems to evolution of the underlying fields, but this is lost as a result of coarse graining at a macroscopic level.

This picture might lead to a description of an EBU that would be consistent with a (local) pointwise evolution as suggested in section 3. For example, a stochastic macroevolution could be produced for a statistical ensemble (e.g. canonical) of initial micro-states corresponding to a given coarse-grained state, which could lead to a stochastic ADM formulation.[22] However, one may also claim that any averaging

---

[21] The observable part of the expanding Universe, since the time of decoupling of matter and radiation.

[22] I am grateful to a referee for this suggestion.

scheme in fact involves choices of special world lines around which the averaging is defined. Thus this is plausibly consistent with the vision **ET2** of emergence of time (seen at the averaged scale) as taking place along preferred world lines that are intimately bound up with the averaging process.

A second possible cause of indeterminism is a "time-dependent equation of state" (section 2.3 above), perhaps associated with emergent complexity or human agency. If the equation is time-dependent, the question arises as to which time is the relevant one? In some situations it will be some local time defined along a particular set of world lines, which will presumably provide the answer to the question as to the world lines along which the evolution actually takes place. In other situations the time dependence might be given by some foliation, in which case this again would seem to give rise to a natural EBU in terms of the process **ET2** considered above.

4.4.2 Quantum Indeterminism

A different form of indeterminism is provided by quantum uncertainty (sections 2.3 and 2.4 above). The way that this fits in with an EBU will depend very much on the description one gives of quantum processes and measurements. Even at the level of quantum fields on a curved background there will be many technical difficulties in relating this to an EBU and these would be even worse if one attempted to say something about quantum gravity. The situation is also problematic because there is no agreed conceptual formulation of quantum theory in a cosmological context. In this context the "measurements" of section 2.4 cannot be clearly identified, and if one instead use a decoherent histories approach there remain problems of determining the families of supports and related operator algebras that yield decoherent histories. Even the attempt to pose the alternatives (**ET1**) and (**ET2**) may not make sense. On the other hand semi-classical approaches that sidestep these problems are in widespread use, particularly the extension of the standard theory that is used in inflationary theory. This takes quantum uncertainty into account, and is based on a description via preferred time lines and spacelike surfaces, and so is compatible with **ET2**.

However, from the viewpoint of this section, one does not need to solve these difficult technicalities. Rather one can take a pragmatic approach, as indicated in the examples discussed in sections 2.3 and 2.4, where well-known and proven properties of quantum theory provide the basis of the conclusions, without a need to discuss how they arise from the underlying quantum field theory. There are two alternatives: (1) that (for reasons as yet unknown) a unique classical space-time necessarily emerges from the underlying quantum theory, because of the nature of that theory and the relevant boundary conditions; (2) that there is no guarantee in general that a good classical space-time emerges, hence cases occur where space-time pictures simply cannot be drawn for the real Universe that eventuates; in this case, further integrability conditions must be imposed on quantum theory if one is to guarantee this emergence.

In view of the observation (**OF**) above and its consequence (**EB**), it does not matter for present purposes which is the case: *either quantum theory on a curved background does indeed imply unique outcomes for physics and space-time under the conditions that occur in the Universe domain in which we live*, and an evolving block universe is a natural outcome of the physics, or *there are extra integrability conditions that should be imposed on those theories to guarantee that this outcome occurs; and it is then an observational fact these conditions are satisfied in the real Universe in which we live, as an evolving block universe does indeed occur*.

Investigating what these conditions are, and how they are satisfied, is a separate issue – essentially the contentious question of the nature of emergence of both classical physics and a classical curved space-time. This will not be tackled here.

4.4.3 Global Issues

Where difficult dynamics issues may well arise is in terms of the global extension of the local results. There may indeed be no globally unique evolution, because this may genuinely depend on choice of families of world lines – different global extensions of a local region may result from different such choices, see for example, the case of the Taub-NUT Universe (Hawking and Ellis 1973: 170–78). There may be a need to make a definite choice, motivated by physical considerations, in order to attain a unique global extension. This is an issue that will need investigation.

4.4.4 The Far Future Universe

What is the final fate of this growing Universe? It is nothing other than the usual fixed unchanging block universe picture, but understood in a new way as the ultimate state of the evolving block universe (EBU) proposed here. It is *what will be in the far future, when all possible evolution has taken place along all world lines*. It is thus, when properly considered, the *final block universe* (FBU). In this picture of the far future universe, time is no longer an indicator of change that will take place in the future, because it has all already happened. At every event along every world line, the present has been and gone. The implications of the existence of this unchanging final state are very different from those usually imputed to the block universe. Things are unchanging and eternal here not because they are immutably implied by the past, but because they have already occurred. This view represents the final history of the universe when all choices have been made and all alternative histories chosen: when it has been determined that the Earth would come into existence, that specific continents would develop, and that particular types of dinosaurs would emerge and then die out. The moving finger has ceased writing; all that will ever be written has been written.

4.4.5 Summary: An Evolving Block Universe

Thus, the view proposed here is that space-time is extending to the future as events develop along each world line in a way determined by the complex of causal interactions; these shape the future, including the very structure of space-time itself, in a locally determined (pointwise) way. It is an evolving block universe that continues evolving along every world line until it reaches its final state as an unchanging final block universe. One might say that then time has changed into eternity. The future is uncertain and indeterminate until local determinations of what occurs have taken place at the space-time event "here and now", designating the present on a world line at a specific instant; thereafter this event is in the past, having become fixed and immutable, with a new event on the world line designating the present. There is no unique way to say how this happens relatively for different observers; analysis of the evolution is conveniently based on preferred (matter related) world lines rather than time surfaces. However, in order to describe it overall, it will be convenient to choose specific time surfaces for the analysis, but these are a choice of convenience rather than necessity.

This paper does not attempt an analysis of how this relates to the philosophy of time (see e.g. Markosian 2002; Fieser and Dowden 2006). Rather I just make one remark in this regard at this point[23]: in terms of the metaphysics of time, this view is that of *possibilism* (the tree model), described in Savitt (2001: section 2.1) and Hunter (2006: section 3).

4.5 Overall: A More Realistic View

The standard block Universe picture is based on reversible microphysics, not realistic irreversible macrophysics. It is described poetically by T S Eliot in his poem *Burnt Norton* as follows (Eliot 1936):

Time present and time past
Are both perhaps present in time future,
And time future contained in time past.
If all time is eternally present
All time is unredeemable.

However, when coarse-graining and emergent effects such as biology are taken into account, with internal variables leading in effect to highly non-linear time-dependent equations of state, time does roll on, indeed this is one of the most fundamental features of our lives: intention changes the future; the past is fixed forever and cannot be changed (Le Poidevin 2004), and Omar Khayyam's poem in section 2.5 is more apposite. Thus the block space-time picture does not represent a realistic view of the real universe. The existence and uniqueness theorems underlying the

[23] See also section 4.3 below.

usual block universe view (see Hawking and Ellis 1973), implying we can predict uniquely to both the future and past from any chosen time surface, do not apply to space-times including complex systems because the equations of state they assume are too simple–they do not include friction and dissipative effects, hierarchical structures, feedback effects, or the causal efficacy of information (Roederer 2005; Ellis 2006a), and they do not take quantum uncertainty into account. A better picture of the situation when realistic physics and biology is taken into account is provided by evolving block universe proposal outlined above, in which the space-time is seen as extending in a pointwise way as time evolves along all possible world lines.

The usual block view description is reasonably accurate for large-scale space-times with very simple matter content (e.g. vacuum or barotropic equations of state), and so is acceptable for cosmological purposes or astrophysical studies such as collapse to a black hole. It is not adequate for small-scale descriptions of space-time with complex matter or active agents such as living beings. It does represent the far future fate of the growing Universe adequately, but in so doing does not imply that data on any spacelike surface enables one to predict or retrodict uniquely to later and earlier times. Rather it represents the quirky contingent nature of what actually happened in the real Universe, which only became apparent as it unrolled.

4.5.1 Determinism and Becoming

How do issues about determinism bear on Becoming? Does determinism entail a static block universe? Does indeterminism entail a dynamical or growing block universe? It seems that the first question has a negative answer, but the second one does not. The growing block universe could grow in either a deterministic or indeterministic fashion; the idea is compatible with both standard general relativity as expressed in the ADM theory, and with more realistic situations where the outcome of events is only determined as time unrolls (see section 2). Thus, determinism does not necessarily imply a static block universe. On the other hand while a special relativity static block space-time could contain indeterministic events, the events themselves in it would emerge as time progresses, so an EBU description would be appropriate for those events. But general relativity implies the space-time structure itself would then be changing. Through their gravitational effects, indeterministic events will change space-time structure in a way determined by the outcome of those events, and imply a growing block universe picture is the appropriate one. In some circumstances these effects will be very small, nevertheless they will be there.

How do issues of the time reversibility of the laws of physics bear on Becoming? A static block universe would seem to be compatible with both time reversibility and time irreversibility of laws, if space-time is seen as the unchanging arena of physics; but as mentioned above, this is not the case when gravitational effects are taken into account. Then time irreversible laws imply a time-direction of “becoming” at each event, which would seem to fit better with a growing block universe, because they go some way to demonstrating that the future and past are of a different character – which is what is made explicit in the idea of a growing block universe.

A common view is that Broad's idea of a growing block universe is unacceptable if "Becoming" is relativised to either a foliation of spacelike hypersurfaces or a family of timelike world lines. For example, Gödel (1957) thought that such a relativised Becoming was not worth having: "The concept of existence … cannot be relativised without destroying its meaning completely." The viewpoint in this paper is that Becoming does indeed take place, and so physical theory had better recognise this feature. If the implication is a relativisation of time in relation to a foliation of timelike lines, so be it. This is something we will have to accept and live with.

4.5.2 The Block Universe and Free Will

The block Universe picture has sometimes been used as an argument against free will (Hoefer 2001; Brennan 2006)[24]; but the physics implied in that picture is not sufficient to even begin contemplating issues of free will. This space-time description does not provide a valid context for that discussion.

The more complex view presented here provides an adequate context for considering the issue, and does not give evidence against the existence of free will; on the contrary, it is broadly congruent with such existence.

4.5.3 The Chronology Protection Conjecture

Does the picture presented here have any implications for the possible existence of closed timelike lines and associated causality violations? It is known that general solutions of the Einstein Field Equations do indeed allow such lines (see, for example, Hawking and Ellis 1973), but Hawking has proposed the *chronology protection conjecture*: "*It seems that there is a Chronology Protection Agency which prevents the appearance of closed timelike curves and so makes the universe safe for historians*" (Hawking 1992; see also Visser 2002; Lockwood 2005; Hunter 2006).

An evolving block universe model, with potentiality transforming into existence as time progresses along each world line, may be a natural context for considering this question, providing a dynamic setting for considering constraints on the developing space-time. One can prescribe an active form of the chronology protection conjecture in this setting, namely that as space-time evolves along a set of physically chosen world lines, *it is forbidden that these world lines enter a space-time*

[24] For a survey of opinions related to this issue, see http://www.maths.nott.ac.uk/personal/gaj/final.htm

domain that already exists. This would provide the needed protection in a natural way.[25]

There might be a cost in terms of existence of space-time singularities in order that this can be accomplished in all cases (geodesic incompleteness might occur when this injunction is invoked); whether this is so needs investigation. But then the singularity theorems of general relativity theory (see Hawking and Ellis 1973) have often had this tension in them: in many cases, it is predicted that either a singularity occurs or there is a causality violation. There may additionally be a lack of uniqueness in the maximal causality-violation free extension; this needs investigation.

4.5.4 The Arrow of Time

If the EBU view is correct, the Wheeler–Feynman prescription for introducing the arrow of time by integration over the far future (Wheeler and Feynman 1945), and associated views comparing the far future with the distant past (Ellis and Sciama 1972; Penrose 1989), are not valid approaches to solving the arrow of time problem, for it is not possible to do integrations over future time domains if they do not yet exist. Indeed the use of half-advanced and half-retarded Feynman propagators in quantum field theory then becomes a calculational tool representing a local symmetry of the underlying physics that does not reflect the nature of emergent physical reality, in which that symmetry is broken.

The arrow of time problem needs to be revisited in this EBU context, with the collapse of the quantum wave function being a prime candidate for a location of a physical solution to the problem. We do not consider it further here.

4.5.5 Issues of Ontology

The hidden issue underlying all this discussion is the question of the ontological nature of space-time: *does space-time indeed exist as a real physical entity, or is it just a convenient way of describing relationships between physical objects, which in the end are all that really exist at a fundamental level*? Is it absolute or relational? Could it after all be an emergent property of interacting fields and forces (Laughlin 2005), or from deeper quantum or pre-quantum structure (Ashtekar 2005: Chapters 11–17)?

---

[25] It has been said to me that quantum field theory assumes there is travel into the past, so causality conditions do not hold. My view on this is that Feynman diagrams with past directed world lines are a possible description of what happens, but there are preferable descriptions in which causality is compatible with special relativity in that all particle paths are future directed. In this case the time reversed diagrams express a symmetry of the system rather than the way the physics actually occurs.

I will not pursue this contentious point here (for discussions, see e.g. Earman 1992; Hoeffer 1998; Huggett 2006). Rather I emphasise here that *the discussion in this paper is about models or representations of space-time,* rather than making any ontological claims about the nature of space-time itself. However I do believe that the kind of proposal made here could provide a useful starting point for a fresh look at the ontological issue, and from there a renewed discussion of the degree to which our representations of the nature of space-time are an adequate representation of its true existential nature.

Acknowledgement I thank Bill Stoeger, two referees, and an editor for comments that have considerably improved this paper.

References

P Anninos, *Computational Cosmology: From the Early Universe to the Large Scale Structure*. http://relativity.livingreviews.org/Articles/lrr-2001-2/articlesu15.html, Section 6.1, The Einstein Equations.

R Arnowitt, S Deser and C W Misner, The Dynamics of General Relativity. In *Gravitation: An Introduction to Current Research*, Ed L Witten (Wiley, 1962), pp. 227–265.

A Ashtekar, *100 Years of Relativity: Space time Structure: Einstein and Beyond*. (World Scientific, 2005).

E B Babcock and J L Collins, Does natural ionizing radiation control rate of mutation?. *Proc Nat Acad Sci* **15**: 623–628 (1929).

J Brennan, Free will in the block universe. Notes available at http://www.u.arizona.edu/∼brennan/freedomblockuniverse.pdf.

H Bondi, "Some special solutions of the Einstein equations". In *Lectures in General Relativity, Brandeis Summer Institute in Theoretical Physics*, Vol. I, Eds A Trautmann, F A E Pirani and H Bondi (Prentice Hall, 1965), 431–434.

C D Broad, *Scientific Thought* (New York: Harcourt, Brace, 1923). For Table of Contents and some chapters, see http://www.ditext.com/broad/st/st-con.html.

P C W Davies, *The Physics of Time Asymmetry*. (Surrey University Press, London, 1974).

P C W Davies, That Mysterious Flow. *Scientific American* **287**: 40 (September 2002).

S Dodelson, *Modern Cosmology* (Academic Press, 2003).

P K S Dunsby, M Bruni and G F R Ellis, Covariant perturbations in a multi-fluid medium. *Astrophys J.* **395**: 54–74 (1992).

J Earman, *World Enough & Space-Time: Absolute vs. Relational Theories of Space & Time.* (Bradford Book, 1992).

T S Eliot, Burnt Norton (No. 1 of 'Four Quartets'). In *T. S. Eliot: Collected Poems 1909–1962* (Faber and Faber, 1974).

G F R Ellis, Relativistic Cosmology. In *General Relativity and Cosmology*, Varenna Lectures, Course XLVII. Ed R K Sachs (Academic Press, 1971), 104–179.

G F R Ellis, Relativistic Cosmology: its Nature, Aims and Problems. In *General Relativity and Gravitation*, Ed B Bertotti et al. (Reidel, 1984), 215–288.

G F R Ellis, Physics, complexity, and causality. *Nature* **435**: 743 (2005).

G F R Ellis, Physics and the Real World. *Phys Today* (July 2005), 49–54. Lett Response, **59**: 12–14 (March 2006).

G F R Ellis, Physics and the real world. *Foundations Phys* (April 2006), 1–36 [http://www.mth.uct.ac.za/∼ellis/realworld.pdf].

G F R Ellis, On the Nature of Emergent Reality. In *The Re-emergence of Emergence*, Ed P Clayton and P C W Davies (Oxford University Press, 2006) [http://www.mth.uct.ac.za/∼ellis/emerge.doc].

G F R Ellis and T R Buchert, The universe seen at different scales. *Phys Lett* A **347**: 38–46 (2005) [http://za.arxiv.org/abs/gr-qc/0506106].

G F R Ellis and D W Sciama, Global and Non-global Problems in Cosmology. In *General Relativity*, Ed L O'Raifeartaigh (Oxford University Press, 1972), 35–59.

G F R Ellis and R M Williams, *Flat and Curved Space Times*. (Oxford University Press, 2000).

R Feynman, *QED: The Strange Theory of Light and Matter.* (Princeton, 1985).

E Fitgerald, *The Rubaiyat of Omar Khayyam*. (Penguin 1989), Stanza lxxi.

B Dowden, Time. In *The Internet Encyclopaedia of Philosophy*, Eds J Fieser and B Dowden (2006). http://www.iep.utm.edu/t/time.htm.

P W Glimcher, Indeterminacy in brain and behaviour. *Annu Rev Psychol*, **56**: 25 (2005).

K Gödel, A Remark About the Relationship between Relativity Theory and Idealistic Philosophy. In *Albert Einstein, Philosopher-Scientist, Vol. 2,* Ed P A Schilpp, (Tudor, 1957).

J Hadamard, *Lectures on Cauchy's Problem in Linear Partial Differential Equations.* (Yale University Press, New Haven, 1923).

S W Hawking, The chronology protection conjecture. *Phys Rev* **D46**: 603–611 (1992).

S W Hawking and G F R Ellis, *The Large Scale Structure of Space Time.* (Cambridge University Press, 1973).

G Hinshaw, WMAP data put cosmic inflation to the test. *Phys World* **19**: (5) 16–19, (May 2006).

C Hoefer, Absolute versus relational spacetime: for better or worse, the debate goes on. *Brit J Philos Sci* **49**(3): 451–467 (1998).

C Hoefer, Freedom from the inside out (2001). Available at http://www.lse.ac.uk/collections/CPNSS/pdf/DP_withCover_Measurement/Meas-DP%2016%2001.pdf.

N Huggett, The regularity account of relational spacetime. *Mind* **115** (457): 41–73 (2006).

J Hunter, Time Travel. In *The Internet Encyclopaedia of Philosophy* (2006), http://www.iep.utm.edu/t/timetrav.htm.

C J Isham, *Lectures on Quantum Theory, Mathematical and Structural Foundations.* (Imperial College Press, 1997).

R Kane, *A Contemporary Introduction to Free Will*. (Oxford University Press, 2005).

E W Kolb and M S Turner, *The Early Universe.* (Addison Wesley, 1990).

R B Laughlin, *A Different Universe: Reinventing Physics from the Bottom Down* (Perseus, 2005).

R Le Poidevin, The Experience and Perception of Time. *In The Stanford Encyclopedia of Philosophy (Winter 2004 Edition)*, Ed E N Zalta, http://plato.stanford.edu/archives/win2004/entries/time-experience/.

M Lockwood, *The Labyrinth of Time: Introducing the universe.* (Oxford University Press, 2005).

N Markosian, Time. *In The Stanford Encyclopedia of Philosophy (Winter 2002 Edition)*, Ed E N Zalta, http://plato.stanford.edu/archives/win2002/entries/time/.

K Martin, J Sugerman and J Thompson, *Psychology and the Question of Agency.* (SUNY, 2003).

D H Mellor, *Real Time II.* (Routledge, London, 1998).

C W Misner, K S Thorne and J A Wheeler, *Gravitation.* (Freeman, 1973).

Committee on the origins and evolution of life, National Research Council: *The Astrophysical Context of Life* (National Academy Press, Washington, 2005) [http://www.nap.edu/catalog/11316.html].

R Penrose, *The Emperor's New Mind.* (Oxford University Press, 1989).

I Percival, Schrödinger's quantum cat. *Nature* **351**: 357 (1991).

F R Prete (Ed), *Complex Worlds from Simpler Nervous Systems.* (MIT Press, 2004).

J Roederer, *Information and its Role in Nature.* (Springer, 2005).

L J Rothschild, Microbes and Radiation. *In Enigmatic Micro-organisms and Life in Extreme Environments*, Ed J Seckbach (Kluwer, 1999), p. 551.

S Savitt, Being and Becoming in Modern Physics. *The Stanford Encyclopedia of Philosophy (Spring 2002 Edition)*, Ed E N Zalta, http://plato.stanford.edu/archives/spr2002/entries/spacetime-bebecome/.

J Scalo, J Craig Wheeler and P Williams, Intermittent Jolts of Galactic UV Radiation: Mutagenetic Effects *In Frontiers of Life; 12th Rencontres de Blois*, Ed L M Celnikier (2001) [astro-ph/0104209].

J J C Smart, Time. *In The Encyclopedia of Philosophy*, Ed P Edwards (Collier-MacMillan, 1967), viii. 126–134.

M R Spiegel, *Theory and Problems of Theoretical Mechanics* (Schaum's Outline Series, McGraw-Hill, 1967), p. 65 and Ex. **3.19** on pp. 73–74.

J M T Thompson and H B Stewart, *Nonlinear Dynamics and Chaos*. (Wiley, 1987).

M Tooley, *Time, Tense, and Causation.* (Oxford University Press, Oxford, 2000).

M Visser, The Quantum Physics of Chronology Protection. *In The Future of Theoretical Physics and Cosmology: Celebrating Stephen Hawking's 60th Birthday*, Ed G W Gibbons, E P S Shellard and S J Rankin.

R M Wald, *General Relativity.* (University of Chicago Press, 1984).

J A Wheeler and R P Feynman, Interaction with the Absorber as the Mechanism of Radiation. *Rev Mod Phys* **17**: 157–181 (1945).

H D Zeh, *The Physical Basis of the Direction of Time*. (Springer, Berlin, 1992).

Chapter 5
The Real World and Space-Time

Hans C. Ohanian

5.1 Introduction

Einstein provided us with an apt description of the scientist's attitude toward philosophy and epistemology:

> The scientist ... accepts gratefully the epistemological conceptual analysis; but the external conditions, which are set for him by the facts of experience, do not permit him to let himself be too much restricted in his construction of his conceptual world by the adherence to an epistemological system. He therefore must appear to the systematic epistemologist as a type of unscrupulous opportunist: he appears as *realist* insofar as he seeks to describe a world independent of the acts of perception; as *idealist* insofar as he looks upon the concepts and theories as free inventions of the human spirit (not logically derivable from what is empirically given); as *positivist* insofar as he considers his concepts justified *only* to the extent to which they furnish a logical representation of relations among sensory experiences. He may even appear as *Platonist* or *Pythagorean* insofar as he considers the viewpoint of logical simplicity as an indispensable and effective tool of his research. [1]

I will interpret this statement as giving me *carte blanche* to attack the problem of the dimensionality of the real world without any preliminary declaration of my affiliation with any particular school of philosophy. This will save me from having to expose my ignorance of philosophy, although it might also mean that my remarks will be judged as naïve by cognoscenti. But sometimes truth comes from the mouth of babes.

As a physicist, I think that philosophers of science have overstated their case for a real 4D world. By mistake or by exaggeration, they have endowed the theory of relativity and the inertial coordinates commonly used in this theory with deep layers of meaning that are not justified by the physics. On the basis of physics, it can be asserted that space-time is a 4D manifold with a 4D geometry, but whether the real world – that is, the totality of all the material things that inhabit space-time – is 4D or 3D, cannot be decided by physics. Both 4D and 3D descriptions of material systems are possible in physics, and physicists use both of these modes of description interchangeably.

V. Petkov (ed.), *Relativity and the Dimensionality of the World*, 81–100.

I will argue that the only firm conclusion that can be drawn from physics is that the recent observable world, as perceived by an observer at some given location in space-time, is 3D. However, this recent observable world is not a flat 3D space. It consists of the past light cone of the observer, which is a curved hypersurface that extends into 4D space-time. Thus, in some sense, the recent observable world is both 3D and 4D: as a manifold it is 3D, but as an embedded curved hypersurface it is 4D.

I will avoid the term *reference frame* in my discussion. In the jargon of physics, a reference frame is a 3D lattice of measuring rods and an array of synchronized clocks. But *observer's reference frame* is also used loosely for *observer's point of view*, and I have even seen this term used as a synonym for *observer's world line*. For the sake of clarity, I have expunged all uses of *reference frame* from this article. Instead, I use the specific term *coordinate system* to mean a 4D coordinate system in space-time.

I will also avoid the term *observer* wherever possible. Some philosophers of science appear to labor under the bizarre misconception that coordinate systems (or reference systems) are associated with individual observers, so each observer has a lattice of measuring rods and synchronized clocks attached to his head, like a saintly halo. The introduction of such an association of coordinate systems and observers seems to have percolated into relativity from quantum theory, where observers indeed play an important role in the interpretation of the measurement process and the collapse of the wave function. None of the early papers and books on relativity (by Einstein, Minkowski, von Laue, Born, or Pauli [2]) introduced observers into the discussion.

A coordinate system is merely a bookkeeping device, to keep track of events in space-time and to assist in the manipulations and calculations with the laws of physics. There is no necessary association between observers and coordinate systems – any observer can use any coordinate system whatsoever, whether at rest with respect to the observer or not, whether inertial or not. The observer can even refuse to use all coordinate systems and identify relevant events in space-time by describing in full detail his apparatus and his observational procedures and results (as Galileo did when he recorded the occultations of the moons of Jupiter his notebook). What an observer perceives depends on what signals arrive at his eyes and ears – this is determined by physical processes, not by what coordinate system he is using. The time coordinate of the coordinate system has no necessary connection to any apperception of simultaneity by the observer. As Stein [3] stated clearly in his incisive critique: "'*a time coordinate*' *is not* '*time*'... what Einstein's arguments showed was that *a certain procedure of measurement singles out a time axis and gives numerical time differences dependent upon that distinguished axis*; not that an observer's state of motion imposes on him a special view of the world's structure. This illegitimate metaphysical interpretation of the time coordinate appears perhaps most plainly in Rietdijk's phrase describing [two observers], when at rest with respect to one another, as "experiencing the same 'present'"; there is of course no such "experience"; the fact that there is no experience of the presentness of remote events was one of Einstein's basic starting points."

The insertion of gratuitous observers and their spurious association with coordinate systems is perhaps the most serious source of confusion in the discussions of the 3D and 4D pictures of the world.

5.2 The Space-Time of Relativistic Physics

In both Newtonian physics and relativistic physics we can define space-time as the set of all *possible* locations of events.[1] Experience teaches us that, on a macroscopic scale, space-time is a 4D manifold with the topology of 4D Euclidean space.[2] The success of current theories of elementary particles and their application to the evolution of the early Universe suggests that space-time is also 4D on a microscopic scale, down to the Planck length, 10^{-33} cm. But we do not know what the structure of space-time is on shorter scales, where gravitational quantum fluctuations probably produce severe topological distortions ("quantum foam"). Devotees of string theory believe that on such short scales, space-time is 11D, but they have not yet found any observational evidence to support this belief.

Although Newtonian physics and relativistic physics use the same 4D space-time manifold, they endow this manifold with different geometries. Newtonian physics uses disjoint geometries for space and for time (technically, the geometry is the "product" of 3D and 1D Euclidean geometries, $E^3 \times E^1$), whereas relativistic physics uses a single, unified geometry (a "complexified" Euclidean 4D geometry, CE^4). The geometry of relativistic space-time is usually described by the Minkowksi space-time interval $s^2 = c^2(\Delta t)^2 - (\Delta x)^2 - (\Delta y)^2 - (\Delta z)^2$. But this expression hinges on the adoption of an inertial coordinate system and separate measurements of space and time intervals, whereas the space-time geometry is actually independent of coordinates. Thus, instead of extracting the space-time interval from separate measurements of space and time intervals, it is more appropriate to obtain it by a single, unified measurement procedure. This is achieved by the "radar-ranging" procedure illustrated in Fig. 5.1 Here A and B are two space-time points between which we want to measure the space-time interval, and AQ is the worldline of a clock in free motion (no external forces) whose world line passes through point A. The clock emits a light signal PBQ which starts at P is reflected at B and returns to the clock at Q. In terms of the proper time τ indicated by the clock, the space-time interval AB is then given by $s^2 = c^2(\tau_P - \tau_A)(\tau_Q - \tau_A)$ [7].

This measurement procedure may be regarded as the space-time analog of measuring spatial distances with a meter rod. Since the measurement procedure treats all intervals in space-time in the same way, regardless of their direction, it gives us operational evidence for the 4D character of the space-time geometry.

[1] I emphasize *possible* because some authors, e.g. Misner, Thorne and Wheeler [4] and Geroch [5] have described space-time as the set of actual events. This is false. The set of actual locations of events is a discrete subset of all possible locations of events; it is not a continuum. At most space-time points nothing has actually happened, is happening, or will happen.

[2] A charming discussion of how we can verify the 4D character of space-time at everyday, macroscopic distances was given by Synge [6].

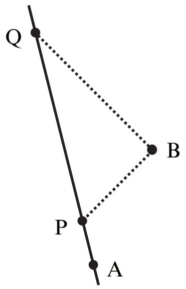

Fig. 5.1 Coordinate-independent measurement of the space-time interval AB.

The quantity $s^2 = c^2(\tau_\mathrm{P} - \tau_\mathrm{A})(\tau_\mathrm{Q} - \tau_A)$ is invariant with respect to a replacement of the clock by a new clock with a world line of different slope (a new clock with different speed). This "clock invariance" is equivalent to the invariance of the speed of light, as postulated by the second of Einstein's two principles for relativity. Note that the usual invariance of the space-time interval under coordinate transformations is a trivial consequence of the adoption of the radar-ranging procedure – the measured value of the space-time interval is coordinate independent, hence, necessarily invariant under coordinate transformations (not only under Lorentz transformations, but under any coordinate transformation whatsoever).

Space-time points sometimes can be identified anecdotally ("the space-time point where and when the apple fell off the tree at Woolsthorpe Manor while Isaac Newton was looking on"). But such anecdotal identification is feasible only for those points at which there is an actual event. Most space-time points do not correspond to any actual events, and we need space-time coordinates to identify these "empty" points. We can select these coordinates in any which way we please, but the most convenient coordinates are inertial coordinates, established by an unaccelerated lattice of measuring rods and an array of synchronized clocks. For the synchronization of the clocks in inertial coordinates, we can adopt either the usual Einstein procedure with two-way light signals, or slow-clock transport. According to basic tenets of the theory of relativity (constant speed of light *and* validity of Newton's laws for particles moving at low speed), these two synchronization procedures are equivalent.

Coordinate systems, their arrays of synchronized clocks, and their equal-time hypersurfaces are merely bookkeeping devices, of no fundamental significance. The actual physics is independent of the choice of these bookkeeping devices, and the only difference between one way of keeping the books and another is in the amount of computational labor required. Some coordinate systems are more convenient than others, but an observer can use any coordinate system that strikes her fancy – she does not need to adopt inertial coordinates, and she does not need to adopt a coordinate system in which she is at rest. Relativity, in the sense of inability to detect uniform motion through space, is a statement about the physical behavior of material systems, and it is valid regardless of what coordinates the observers choose to adopt. The laws of relativistic physics do not require inertial coordinate systems – they can also be formulated in general, noninertial coordinate systems, and they can even be formulated in a coordinate-independent manner, by Cartan's calculus of differential forms (which can be regarded as a generalization of the usual 3D vector calculus). Of course, the formulation of the relativistic laws in general coordinates is messy, and the Cartan calculus is rather abstruse. In practice, numerical calculations of experimental results are most conveniently done in inertial coordinates, where the laws take their simplest form. But in principle, the laws of physics can written in any general coordinates, and calculations of experimental results in such general coordinates will always agree with the (simpler) calculations in inertial coordinates.

The advantage of inertial coordinate systems over other, more general coordinate systems is that Newton's laws of motion are valid in their simple, standard form for particles moving at low speeds. Simplicity is the *only* distinction enjoyed by inertial coordinate systems. Einstein described these coordinate systems as "...Cartesian systems of co-ordinates, the so-called inertial systems, with reference to which the laws of mechanics (more generally the laws of physics) are expressed in their simplest forms" [8], and he added that in these coordinates "For infinitely slow motions, Newton's laws of motion are undoubtedly valid..." [9]. Thus, in inertial coordinates, Newton's second law has the simple form $F = ma$, which contains only the mass of the particle, its acceleration, and the force that acts on it. In more general, noninertial coordinates, we must insert extra terms into the second law, with extra parameters that characterize the coordinate system. For instance, in an accelerating or rotating coordinate system, we need extra terms (pseudoforces) depending on the acceleration or on the speed of rotation of the coordinates.

As I have argued elsewhere [10, 11], in inertial coordinates the synchronization is completely determined by the requirement that Newton's laws be valid in the limit of low speeds. Thus, in inertial coordinates, there is no freedom to alter the synchronization in the manner proposed by Reichenbach and, later, by Grünbaum and others. The Reichenbach mistake can be traced back to a mistake by Einstein, who thought that measurement of the one-way speed of light requires the use of clocks synchronized by light signals, which would reduce any attempt at such a measurement to a vicious circle, and would mean that the constant value of the one-way speed of light must treated as a definition, that is, as a convention. With his characteristic stubbornness, Einstein stuck to this view throughout his life; he declared that the constancy of the speed of light was "neither a *supposition nor a hypothesis* about the physical nature of light, but a *stipulation* which I can make at my free discretion to arrive at a definition of simultaneity" [12], and he agreed that Reichenbach's resynchronization was a viable alternative to his own.

But in this Einstein was mistaken. The one-way speed of light does not have to be measured by means of clocks synchronized by light signals; it can be measured by means of clocks synchronized by slow-clock transport. It can even by measured *without* any prior synchronization of clocks. Perhaps the most elegant method is the following, which relies on two clocks that are identical (same clock rate), but not necessarily synchronized. The two clocks are placed at the ends of an east–west track, and a light signal is sent from, say, west to east along this track. Instead of returning the light signal immediately (which would give us the round-trip speed), we hold the light signal for *exactly* half a sidereal day at the east end, and then send it back to the starting point at the west end. Because of the rotation of the Earth, the light signal will then travel in the same direction of space both during the first and the second portions of its trip. From the times registered by the clocks (discounting the half sidereal day for the "hold"), we can calculate the one-way speed of light for the given direction in space. Evidently, this method relies implicitly on slow-clock transport by the rotation of the Earth; it relies on the assumption that the transport by this rotation does not affect the clocks.[3] This method, in essence, was used recently for one-way measurements of light signals in an optical fiber [13]. A variant of this method was used for one-way measurements of radio signals between GPS satellites and ground stations [14]. The most precise available synchronization procedure currently in use is the TWSTFT procedure, which relies on a combination of clock transport and radio signals from geostationary satellites [15].

Thus, to the extent that it purports to apply to inertial coordinate systems, Reichenbach's conventionality of synchronization is a fallacy. In inertial coordinate systems, slow-clock transport provides a unique synchronization in accord with the simple, standard form of the laws of Newtonian mechanics. However, conventionality of synchronization is applicable to general, noninertial coordinate systems. Even more: in general, noninertial coordinate systems, we are not restricted to the Einstein synchronization nor to the Reichenbach synchronization. We can adopt *any* synchronization whatsoever, that is, we can select the equal-time hypersurfaces in any way whatsoever, subject only to some conditions of mathematical continuity and smoothness. Successive equal-time hypersurfaces do not have to be parallel in space-time, nor do they have to be flat surfaces. The only absolute requirement is that they must not intersect one another. In this broad sense, conventionality of synchronization infests all general, noninertial coordinate systems.[4]

If we use hypersurfaces other than the standard $t =$ constant hypersurfaces of an inertial coordinate system, the laws of physics in our coordinate system will not have their standard form. For instance, if we adopt the Reichenbach synchronization, Newton's second law will differ from the standard form $F = ma$ by extra terms, which are analogous to the familiar "pseudoforces" that arise in accelerated and rotating coordinates. In general coordinates, both Newton's first and second law will differ from their standard forms. The first law can be expressed as an equation for geodesic motion, and the second law can be expressed as an equation that gives the deviation from geodesic motion produced by the force. These nonstandard forms of the laws look complicated, but they work perfectly well for calculating the outcome of any experiment, although they demand much more of a calculational effort.[5]

[3] This can be tested experimentally by repeating the experiment at higher latitudes (where the speed of rotation is smaller), and then extrapolating to zero speed.

[4] In general relativity, the absence of global inertial coordinates implies that conventionality of synchronization infests *all* coordinate systems; the construction of the space-time coordinates used in general relativity always hinges on the adoption of conventions.

[5] As an illustration of how messy things get, see Ohanian [10] for the formulation of Newton's laws in coordinates with the Reichenbach synchronization and see Will [16] for an analysis of several relativity experiments in coordinates with arbitrary synchronization (akin to the Reichenbach synchronization).

The first of Einstein's two principles for relativity – that is, the invariance of all laws of physics with respect to changes of uniform translational motion of the coordinates – is usually expressed in inertial coordinates as an invariance of these laws under Lorentz transformations. But it, too, can be reformulated without any mention of inertial coordinates.

To express the principle of relativity and all laws of special-relativistic physics in general coordinates, we can borrow the language of general relativity and (formally) treat special relativity as a special case of general relativity. We can then say that in special relativity the laws of physics are covariant and that the Riemann curvature tensor of space-time vanishes. The requirement of covariance is somewhat technical, but, in essence, it demands that the laws are vector or tensor equations containing *only* the vectors or tensors that characterize the dynamical system under consideration (such as the electromagnetic field tensor, the energy–momentum vector, the energy–momentum tensor, etc.) and the "absolute" metric tensor of space-time, but no other "absolute" vectors or tensors whatsoever.[6] Such a covariant formulation of physical laws in special relativity has been examined in detail by Fock [18].[7]

Although the distinction between the space-times of Newtonian and relativistic physics is usually thought to reside in their different geometries, this distinction actually goes deeper, to a pregeometric, topological level of relationships involving particle signals or light signals.

For a given space-time point P we define the *absolute past* as the set of all events from which signals of finite speed can be sent toward P, and we define the *absolute future* as the set of all events toward which signals of finite speed can be sent from P.[8] Upon the adoption of a time coordinate, the nominal past and future can be defined as $t < 0$ and $t > 0$, respectively, *in the chosen coordinate system.* In Newtonian physics, the nominal past and future agree with the absolute past and future, if and only if the coordinate system is inertial. In relativistic physics, the nominal past and future do not agree with the absolute past and future; the nominal past and future contain the absolute past and future, but not vice versa.

The distinction between absolute past and future is purely topological; in principle, these regions can be identified without any metrical measurements, by simple experiments that test whether or not signals can be sent. The topology of the absolute past and future regions defines the causal structure of space-time; it determines which events can be causally related to which other events.[9]

[6] Anderson [17] calls tensors that do not depend on the dynamical system absolute. In special relativity, the metric tensor is absolute, but in general relativity it is not.

[7] Fock's objective was propaedeutic; he regarded the covariant formulation of special relativity as a good introduction to general relativity. But here I am doing the reverse: I use the language of general relativity to elucidate special relativity.

[8] The notion of finite speed can be considered as nonmetrical, because a finite-speed signal is simply a signal such that there exists another signal that can overtake it.

[9] In general relativity this causal structure is often called the conformal structure of space-time, because it corresponds to the structure of light cones (null cones), and these determine the metric of space-time to within an overall, "conformal" factor. See Ehlers, Pirani, and Schild [19] and Penrose [20].

In Newtonian physics, the *absolute present* can be defined topologically as the frontier of the absolute past (and also the frontier of the absolute future). The absolute present is a 3D hypersurface consisting of all possible world lines of signals with a speed $v \rightarrow \infty$ that pass through P. Thus, in the space-time of Newtonian physics, the absolute present can be defined by direct physical considerations, without introducing any coordinates – the absolute present is what an observer at P perceives by means of infinite-speed signals. The union of absolute past, present, and future spans the entire space-time.

In relativistic physics, the absolute past and the future and their frontiers do not span the entire space-time. Between the absolute past and the future lies a 4D "neutral," or "acausal" region containing events that cannot send signals to P nor receive signals from P. The frontier between the neutral region and the absolute past and future regions is a 3D hypersurface consisting of the world lines of all the light signals that pass through P, that is, the light cone. Thus, the topological structure of the regions that can and cannot be connected by signals is different in Newtonian and relativistic physics, and we do not need to perform geometrical measurements to recognize this difference.

Minkowski asserted that "Henceforth space by itself, and time by itself, are doomed to fade away into mere shadows, and only a kind of union of the two will preserve an independent reality" on the grounds that the space-time geometry is 4D [21]. But he could have justified his assertion equally well at a topological level by the 4D structure of light cones, that is, the conformal structure of space-time. In Newtonian physics, there is no such inherently 4D topological structure that characterizes signals.

Nehrlich [22] has argued that since the mathematical formalism of relativity permits the existence of tachyons, the light cone does not correspond to the causal structure of space-time. Tachyons signals would make it possible to connect all regions of space-time with all other regions, and our simple way of discriminating between the regions interior and the exterior to the light cone would be lost. This is a if-cows-could-fly argument. Just as there are solid physical reasons why cows cannot fly, there are solid physical reasons why tachyons cannot exist. If tachyons of positive energies exist, then tachyons of negative energies must also exist, and such negative-energy tachyons would lead to violent instabilities – the vacuum could spontaneously generate an unlimited number of tachyons. The fact that the mathematics of relativity (or, more precisely, the mathematics of representations of the Lorentz group) permits representations that correspond to tachyons does not mean that tachyons might exist – it merely means that mathematics, on its own, is sometimes dumb. The existence of "tachyon solutions" in the analysis of representations of the Lorentz group is analogous the existence of spurious or imaginary solutions of some quadratic or cubic equations in physics; mathematically such solutions are permitted, but they have to be rejected on physical grounds.

Even if, as an intellectual exercise, we admit tachyon signals, the light cone of relativistic physics would still play an essential role. Particles can be unambiguously (and permanently) classified as "normal" and as tachyons. Normal signals are confined to the regions inside the light cone, and tachyon signals to the region outside the light cone. Thus, the space-time of relativistic physics still has an inherently 4D topological structure, which is absent in Newtonian physics. Furthermore, tachyons would not permit us to identify an absolute present. If, in imitation of Newtonian physics, we attempt to identify such an absolute present by means of tachyons signals of infinite speed, we run afoul of the limitations that physics imposes on the production and detection of such signals. Tachyons of infinite speed have zero energy, and this makes them undetectable, because they are masked by thermal fluctuations in the emitter and the detector (in general, if the effective noise temperature of the emitter or detector is T, a detectable signal must have an energy of at least kT). Thus, an observer cannot perceive any tachyons signals from his nominal present and he cannot identify an absolute present observationally. I will ignore tachyon speculations hereafter.

Any of the hypersurfaces in the 4D acausal region between the absolute past and future regions can be regarded as an equal-time hypersurface for the point P, provided that the surface includes the point P and that every point on the surface is in the acausal region of every other point on the surface (that is, the surface is "spacelike"). Thus, in relativistic physics there is no unique, physically distinguished definition of simultaneity. To decide what is simultaneous, we need to adopt some prescription for the construction of equal-time hypersurfaces, or, equivalently, a procedure for clock synchronization. Simultaneity then becomes a matter of convention, without any direct physical significance. In contrast to Newtonian physics, the equal-time hypersurfaces of relativistic physics are not visually observable, and they have no direct correspondence to the perceptions of an observer located at P – the equal-time hypersurfaces in relativistic physics do not enjoy any topological distinction. By adoption of a set of equal-time hypersurfaces in conjunction with spatial coordinates, we construct a 4D coordinate system. But such equal-time hypersurfaces and coordinate systems are artifacts, they are merely bookkeeping devices.

Thus, in relativistic physics there is no absolute time, and no physical meaning can be attached to simultaneity at different locations. When we assert that two events are simultaneous, we are merely saying that the events have the same time coordinate in our adopted coordinate system. This is just as meaningless as the assertion that two points in 3D space have the same x coordinate, which tells us something about the orientation of the x-axis, but tells us nothing about the actual separation between the points. Any events that are outside of each others light cone can be made simultaneous, by adopting an equal-time hypersurface that contains these points. Time intervals between such events have no physical meaning; they merely reflect the choice of coordinates. Only the space-time interval s^2 between the events has an absolute, coordinate-independent meaning.

Contrary to a widespread misconception, the equal-time hypersurfaces of an inertial coordinate system are of no fundamental significance. The *only* distinction enjoyed by inertial coordinate systems in relativity is that in these coordinate systems the laws of physics take their simplest, most convenient form. When we assert that two events are simultaneous in an inertial coordinate system, we are merely saying that they are simultaneous in the most convenient coordinate system. Thus, the relativity of simultaneity associated with inertial coordinate systems is merely a relativity of convenience – what is convenient in one coordinate system is not convenient in a different coordinate system. The real lesson of relativity is not that simultaneity is relative, but that convenience is relative.

5.3 The 3D World vs. the 4D World

It is important to make a clear distinction between the dimension of space-time and the dimension of the physical "world," that is, the set of all material things that can be perceived by our sense organs or by extensions of our sense organs. There is no doubt about the dimension of space-time – the discussion in the preceding section establishes that the space-time of relativistic physics is 4D, as demonstrated by the geometry of the space-time interval (and the unified measurement procedure illustrated in Fig. 1) and also by the structure of the light cones. The physical world inhabits this 4D space-time, but the physical world does not occupy all of space-time – it occupies only a subspace of space-time. The dimension of the physical world is the dimension of this occupied subspace.[10]

It is obvious that this occupied subspace is at least 3D. In any given space-like hypersurface in space-time (that is, at any given nominal instant), there exist planets, and people, and atoms, and electromagnetic field, and other physical 3D objects. But philosophers have become embroiled in a lengthy debate on whether these objects should be regarded as "perduring," so they exist not only at one instant, but along the full extent of their worldtubes, which would mean they exist as 4D entities. Although the arguments are largely metaphysical (and sometimes semantical), some philosophers have attempted to use relativistic physics to support the 4D picture of the world. I do not feel qualified to comment on the metaphysics, and I will restrict my comments to the use – and misuse – of relativistic physics in these arguments. Some of my criticisms have been anticipated by Stein, [3, 23], with whose assessment of the basic aspects of relativistic space-time I fully agree. In substance, what Stein says about relativistic space-time and about the absence of absolute simultaneity corresponds with what I say in section 5.2, except that he fails to make a clear distinction between what is purely topological and what is metrical in relativistic space-time, and he gives an excessively abstract treatment of space-time geometry, via inner products. Maybe the abstract style of Stein's approach explains why his incisive arguments seem to have fallen on deaf ears.

The question of the dimensionality of the world is closely related to the question of defining the present. If we can construct a unique, preferential equal-time hypersurface that defines an absolute present, then we can assert that the world

[10] The distinction between the dimensions of space-time and the physical world becomes acute if we adopt a microscopic view and assume that the physical objects that inhabit space-time are ideal, pointlike particles. The physical world is then either 0D or 1D (the latter corresponds to "perduring" particles, see below); in either case, its dimension is smaller than that of space-time. But for the purposes of this paper, I will adopt a macroscopic view and assume that physical objects have three spatial dimensions.

that exists in this present hypersurface is 3D. Obviously, in Newtonian physics the hypersurface that forms the frontier between absolute past and absolute future plays a preferential role; it defines a unique, absolute present, without the prior introduction of any coordinate system or of any definition of clock synchronization. Physically, events are simultaneous if they can be connected by signals of infinite speed. Since this definition of simultaneity is attained without any convention about the choice of coordinates, simultaneity is absolute in Newtonian physics.[11]

Thus, in Newtonian physics, the absolute present is well-defined, and the material things that exist in this absolute present are 3D, which is in agreement with our everyday intuitive experience based on sensory data from our binocular vision and our sense of touch. However, it needs to be emphasized that although Newtonian physics permits a 3D picture of the world, its does not compel it. In Newtonian physics it is perhaps natural to characterize the continuing existence of an object as "endurance," with the object existing successively at different instants of absolute time. However, "perdurance" can also be accommodated, with the object existing at all times (or at all times within some finite range).[12] In fact, prerelativistic philosophers did not regard Newtonian physics as an obstacle to perdurance; for instance, Sider [24] lists Edwards, Hume, and Lotze as adherents of perdurance. If for some metaphysical reason, we choose to believe that the world is 4D, then we can think of the 3D world as a "snapshot" of this 4D world. Newtonian physics is consistent with a 3D picture of the world and also with a 4D picture of the world.

In relativistic physics, some philosophers have sought to interpret the space-time structure and the space-time geometry as evidence for the 4D world. The arguments that supposedly compel a 4D picture of the world fall into three categories: (i) arguments that claim that the relativistic length contraction and time dilation cannot be understood unless the world is 4D [25, 26, 29]; (ii) arguments that rely on the relativity of simultaneity in inertial coordinate systems to establish that all events on the world-line of any particle exist jointly [26–28]; and (iii) arguments that claim that the absence of an absolute present make a unique 3D world impossible [30, 31]. All these arguments are flawed; they all misconstrue diverse aspect of relativistic physics and draw inappropriate conclusions.

1. The claim that length contraction and time dilation cannot be understood unless the world is 4D is easily contradicted by showing that these relativistic effects can be derived from a purely 3D calculation. In fact, historically, the first derivation of the length contraction was obtained by such a 3D calculation by Lorentz in 1904, before Einstein's elegant and general derivation in 1905. In his calculation, Lorentz assumed that the length of a rigid body varies with speed in the same way as the length of a system of electric charges in electrostatic equilibrium. Maxwell's equations imply that the electric fields of a moving charge contract longitudinally,

---

[11] From this it is clear that, in Newtonian physics, Reichenbach's conventionality of synchronization is ruled out on purely topological grounds – there is actually no need to consider the dynamical consequences to rule out the Reichenbach scheme. Such an examination of dynamical consequences becomes necessary only if we regard Newton's laws as low-speed approximations to relativistic laws [10].

[12] I am following the terminology of Sider [24].

which had been known since 1880, from the Liénard–Wiechert potentials. From this contraction of the fields, Lorentz derived a corresponding contraction of the equilibrium configuration of a system of electric charges, and this gave him exactly the relativistic length contraction, by means of which he explained the null result of the Michelson–Morley experiment.

Admittedly, Lorentz's demonstration of the length contraction as a consequence of electromagnetic effects had serious flaws (of which Lorentz was well aware). His basic idea was sound, but in the days before quantum mechanics, physicists had no clear understanding of what determines the size of an atom and what determines the size of a lattice of atoms, such as found in a measuring rod. However, today we can repeat the Lorentz calculation with full quantum-mechanical details, and we can verify that, indeed, a moving rod and any other physical system contracts in consequence of dynamic effects that come into play during motion. The Lorentz approach to the length contraction was strongly advocated by Bell [32], who illustrated this approach with a simple "toy model" of an atom involving relativistic mechanics and electromagnetism. For pedagogical reasons, Bell did not use quantum mechanics, and he did not attempt an exact solution of the differential equation of motion, but only an approximate numerical solution.[13] It is actually quite easy to show that the length contraction is a rigorous consequence of the relativistic wave equations for electrons in quantum-mechanical steady states around nuclei.

An explicit and exact solution of the relativistic wave equation (Dirac equation) for the 10^{26} or so electrons and nuclei that make up a measuring rod is impossible, but for the purposes of extracting the length contraction an explicit solution is not needed. It suffices to compare the putative solutions for the case when the nuclei are at rest with the case when the nuclei are all moving with uniform speed in some direction, say, the x direction. Suppose that the solution of the many-electron wave equation for nuclei at rest gives an equilibrium configuration of length $\Delta x = L$. Then consider the many-electron wave equation for nuclei in motion with velocity V, and perform a *change of variables* in this wave equation, replacing the electron and nuclear coordinates x, t by new variables x', t' according to the substitutions $x = \gamma(x' + Vt')$ and $t = \gamma(t' + Vx'/c^2)$, with $\gamma = 1/(1 - V^2/c^2)^{1/2}$, and replacing the electron spinor wave function by its corresponding spinor transformation. Of course, a knowledgeable reader will recognize these substitutions as a Lorentz transformation, but for the present purposes, the physical interpretation of the transformation is irrelevant – it is to be regarded merely as a change of dependent and independent variables, a standard trick for the solution of partial differential equations. Because of the Lorentz invariance of the wave equation, this change of variables eliminates (or hides) the velocity V and reduces the wave equation to exactly the same form as the wave equation for nuclei at rest. The equilibrium length in terms of the variables x', t' is therefore necessarily $\Delta x' = L$, and, by inverting the change

[13] Bell was a brilliant physicist, and it is clear from his paper that he knew how to derive the length contraction exactly; but he felt that students would find his approximate treatment more transparent.

of variables, we immediately find that the equilibrium length in terms of the original variables x, t is $\Delta x = L/\gamma$, which is the expected length contraction.

Petkov [29] criticizes the Lorentz–Bell approach and claims that length contraction is a contraction of space. He concludes this from the example of cosmic-ray muons traveling downward at high speed from the top of the atmosphere, in whose coordinate system the distance from the top to the bottom of the atmosphere is contracted, which he views as a contraction of space. But to speak of a contraction of space is meaningless. A space contracted in one direction would have a metric tensor in which the, say, 11 component differs from the Minkowski tensor by some numerical factor. However, the space-time geometry for such a "modified" Minkowski tensor is identical to the original geometry – it is still a flat geometry, and it only seems to differ from the original geometry because we have introduced a modified coordinate, shortened by a numerical factor. If the length contraction represents anything physical, it must be a contraction of physical bodies, not a "contraction of space." The height of the atmosphere is defined by physical measuring rods in the coordinate system of the Earth (this would be true even if the atmosphere were an evacuated region except for a thin top layer in which muons are created), and the length contraction in the coordinate system of a moving muon is, effectively, a contraction of these measuring rods. (Note added in proof: I recently found that this argument was first given by Swann [36]).

Thus, transformations between inertial coordinates and Minkowski diagrams are *not* needed to understand the length contraction. We could, in principle, do physics in the manner of Lorentz, in one single coordinate system (what Lorentz called the ether frame). The Einstein–Minkowski derivation of the length contraction, which asks us to compare the positions of the measuring rod in two coordinate systems, one at rest with respect to the rod and one in motion, is merely a convenient alternative to the Lorentz derivation. The basic ingredient in both derivations is the same: the invariance of the equations of physics under Lorentz transformations. The only difference is how this invariance is exploited. Einstein uses this invariance implicitly, in assuming that a measuring rod retains its length of 1 m in the inertial coordinate system in which it is at rest, regardless of the speed that the rod has with respect to any other coordinate system; Lorentz uses the invariance explicitly in the comparison of solutions of the differential equations that determine the equilibrium length. Einstein's approach is elegant and general, but it makes the length contraction look like an accounting trick, and it has sometimes led to the misconception that the length contraction is merely a matter of appearances or a matter of perspective.[14] Lorentz's approach is more physical; it reveals how the length contraction arises from the dynamical properties of the electromagnetic fields that determine the equilibrium length.

[14] See, for instance, Born's misbegotten assertion that "the contraction is only a consequence of our way of regarding things and is not a change of a physical reality" and his claim that "it does not come within the scope of cause and effect"[34]. Born blindly accepts Einstein's Jesuitical approach to relativity: the end (constant speed of light) justifies the means (length contraction, time dilation, relativity of simultaneity). A deeper justification is achieved by cause–effect (dynamic) arguments.

Since it is possible to formulate a derivation of the length contraction without the use of any 4D notions or changes of coordinate system, the length contraction gives no evidence for (or against) the 4D picture of the world. The same applies to the time dilation, which can, again, be derived by examining the dynamics of a moving clock (more specifically, by examining the solution of the quantum-mechanical equations for an atom or a molecule) by methods similar to those described above.

2. The argument for the 4D world extracted from the relativity of simultaneity in inertial coordinate systems was first given by Rietdijk [27] and Putnam [28].[15] A very clear version of the argument was given by Petkov [26], who phrased it in terms of the progression of a clock along its world line. For the sake of variety, and for the sake of preventing any confusion between the progressing clock and the clocks of the coordinate system, I will rephrase the argument in terms of the progression of an egg along its world line, or more precisely, its worldtube. Suppose that a chicken lays an egg at some event, and that this egg hatches at some later event. In essence, the Rietdijk–Putnam argument asks us to consider these events from the point of view of two observers placed at some large distance, so both events are in the acausal region for these observers. Suppose that for one of these observers, the laying of the egg is on the equal-time hypersurface $t=0$ of his inertial coordinate system. Suppose that the other observer is (instantaneously) at the same location as the first, but in motion with respect to the first, and the speed is such that for this observer, the hatching of the egg is on the equal-time hypersurface $t' = 0$ of her inertial coordinate system. Then the first observer claims that the laying is real (it exist in his present world) and the second observer claims that the hatching is real (it exists in her present world). Thus, Rietdijk–Putnam claim that both the laying and the hatching exist jointly. By introducing more observers with different speeds, Rietdijk–Putnam can likewise claim that every event in the worldtube of the egg between laying and hatching exists jointly with every other event. According to this argument the egg exists as a 4D entity, that is, the world is 4D.

This argument hinges on the simplistic assumption that an inertial coordinate system is associated with each observer, and that what the observer perceives as his or her present necessarily corresponds to the equal-time hypersurfaces of his or her inertial coordinate system.[16] But there is no compelling reason for the adoption of such inertial coordinates systems at rest relative to the observers. The observers are not using these coordinates for the purpose for which they are intended, that is, the simplest formulation and application of the laws of mechanics. Since the argument makes no explicit use of the laws of mechanics or of any other laws of physics, the adoption of inertial coordinates is gratuitous and irrelevant. If any compelling conclusion is to be drawn from their argument, Rietdijk–Putnam must give compelling reasons for the adoption of inertial coordinates – which they fail to do.

[15] Putnam's own argument is rather muddled. But, like the Rietdijk argument, it evidently relies on the relativity of simultaneity in inertial coordinate systems, and that is the only issue that concerns me.

[16] Rietdijk takes this to an absurd extreme by supposing that the observer might briefly change speed and thereby cause his equal-time hypersurface to swing from one orientation to another.

Rietdijk vaguely seems to sense that the adoption of inertial coordinates with standard synchronization is a weak point of his argument, and he tries to preempt criticism by attacking the Reichenbach conventionality of synchronization, declaring that he "cannot accept Professor Grünbaum's interpretation" and the Reichenbach resynchronization of clocks. The only physical reason he advances for his nonacceptance of this resynchronization is that isotropy of space requires the constancy of the light-velocity. But this is nonsense: Rietdijk confuses isotropy of space with isotropy of the coordinate system. In special relativity, space is isotropic (technically, the homogeneity and isotropy of space are characterized by Killing vectors [33]); but whether the coordinate system and the synchronization in the coordinate system are isotropic hinges on what conventions we adopt in the construction of the coordinates. Anisotropy in the coordinates does not destroy the isotropy of space, although it makes it harder to recognize this isotropy.

What is more, the multiplicity of observers inserted into the Rietdijk–Putnam argument is also gratuitous and irrelevant. Observers are (perhaps) needed for the interpretation of quantum mechanics, but they are certainly not needed for the interpretation of non-quantum physics. Instead of a whole slew of observers, it suffices to rely on a single observer ("me"), and even this single observer is merely needed as a timing device, to establish an event in relation to which we consider the simultaneity of other events. The multiple observers are inserted into Rietdijk–Putnam argument under the misconception that each inertial coordinate system must belong to some observer, and that each observer is required to use a coordinate system in which he or she is at rest.

If we eliminate the gratuitous observers, the argument reduces to this: at some event, I can adopt an inertial coordinate system such that the laying of the egg is in my nominal present or I can adopt an inertial coordinate system such that the hatching is in my nominal present. Rephrased in this way, the argument is seen to be nothing more than a statement about the various inertial coordinate systems that are at my disposal. Attempts to extract broad philosophical conclusions from the adoption of such inertial coordinates are meaningless. The inertial coordinates have no special significance, except in that they provide the simplest, most convenient formulation of the laws of physics. Thus, the philosophical arguments based on inertial coordinates are delusional; in these arguments, simplicity and convenience masquerade as philosophy. Such games with coordinates systems tell us nothing about the physics of the real world, and – contrary to what Rietdijk would have us believe – they tell us nothing about determinism. They tell us only that inertial coordinate systems do not permit the identification of a unique present, so the nominal present is ambiguous. Whether any meaningful conclusion can be extracted from this ambiguity is an issue that will be addressed in the discussion of the next argument.

3. In relativity, the nominal present depends on the choice of coordinate system, so any spacelike hypersurface that includes "me" can be regarded as my present for some choice of coordinate system. This means that I cannot identify a unique present, and I cannot identify a unique 3D world as existing in my present – any of the 3D worlds in any spacelike hypersurface that includes "me" can be thought of as

my present world. This ambiguity indicates that the 3D world is ill-defined. Some philosophers of science have interpreted this ambiguity as evidence against the 3D picture of the world.

Weingard [30] has proposed to eliminate this ambiguity by redefining the *entire* acausal region as the present of the space-time point P. This means the present is not confined to a single spacelike hypersurface, but instead includes *all* spacelike hypersurfaces that pass through P. The present is then 4D, and the world that exists in this present is likewise 4D. Weingard claims that his redefinition of the present yields a close analogy to the Newtonian present. But to a physicist, his redefinition of the present makes no sense, because the equations of physics are supposed to permit the determination of the future state of a dynamical system from its present state, that is, the equations should provide a well-posed initial-value problem. With initial values specified on one given spacelike hypersurface, the equations of physics do indeed have a well-posed initial-value problem. But they do not have a well-posed initial-value problem if the initial data are given on *several* hypersurfaces; this overdetermines the initial values and results in inconsistencies. The only way to avoid such inconsistencies among the initial values is to select *one* hypersurface as "primary" and adjust the initial data on all the other hypersurfaces to be consistent with the primary data – but this effectively recognizes the primary surface as the primary present, and to speak of the other hypersurfaces as also present is then merely wordplay. The selection of the primary hypersurface is, of course, subject to the same ambiguity as the selection of a 3D world.

However, it is not at all clear how much importance should be attached to this ambiguity of the 3D world. There are no physical consequences of this ambiguity – all the conceivable 3D worlds that I might regard as my present are in the acausal region, where I can neither observe them or be affected by them in any way. Thus, the ambiguity merely concerns how I imagine the world, but it does not affect what I perceive and measure and predict. It is a quintessentially metaphysical ambiguity, of no direct relevance for the operational application of the laws of physics. As a physicist, I can safely ignore this ambiguity.

If for whatever metaphysical reason, I want to adhere to a 3D picture of the world, I can do so by selecting one single, preferential coordinate system with a specified standard synchronization. This makes my 3D world into a convention, based on a mode of cognition – an "*Erkenntnisform*," as Kant would have said, but not an "*apriori Erkenntnisform*" – that I have adopted to cognize the 3D world. Weingard opposes this way of selecting a 3D world ("...being real, I take it, cannot be merely a matter of convention" [30]) and so does Petkov ("...conventionality of simultaneity implies conventionality of what exist, which is clearly unacceptable" [26]). As a physicist, I don't share their certainty about what is or is not acceptable for an unobservable, purely imaginary and hypothetical 3D world. The 3D picture does not deal with anything observable, it cannot lead to observable inconsistencies, and it cannot be refuted by physics. And if the 3D world is unobservable, I am content to fix it by convention. Sider characterizes this approach as "scientifically revisionary," [24] but offers no other, substantive objections.

Although the adoption of a preferential coordinate system is contrary to the spirit of relativity, it does not lead to any observable violations of relativity. The essence of relativity is that the result of any experiment performed with any kind of apparatus, at rest or in uniform motion, is unaffected by the motion. For the analysis of an experiment, we usually like to adopt a coordinate system in which the apparatus is at rest, which permits a simple and convenient implementation of Einstein's principle of relativity. But the use of such a "rest" coordinate system is not necessary. What coordinates we use for the description of the experiment is ultimately irrelevant – only the result of the experiment matters.

The Michelson–Morley experiment illustrates this point. The null result of this experiment can be understood most easily in an inertial coordinate system in which the apparatus is at rest, where the absence of a fringe shift is an immediate consequence of the constant value of the speed of light along the two perpendicular arms. But the null result can also be understood in an inertial coordinate system with respect to which the apparatus is in motion (say, a coordinate system at rest with respect to the Sun). In such a coordinate system, the difference between the speed of light and the speed of the apparatus (the "closing" speed) is not the same for the two arms, but the experiment again gives a null result in consequence of the length contraction of one of the arms.

Lorentz's work on relativity provides a clear historical precedent for the adoption of a single, preferential coordinate system. Lorentz adopted an inertial coordinate system at rest in the "ether," which he imagined to permeate the entire universe. He believed that the time defined in this coordinate system was the only "true" time, that is, he believed in an absolute time. He called the time calculated from the Lorentz transformation the "local" time, and he regarded this time transformation merely as a transformation of the independent mathematical variable, to be used for the solution of differential equations by "change of variables." Nevertheless, he believed relativity to be valid, in the sense that experiments are unaffected by motion relative to the preferential coordinate system. This made the ether undetectable, but Lorentz thought that oscillating light waves needed some medium for their oscillations – for him the oscillations of a light wave were in themselves sufficient evidence for the ether. Lorentz originally took the ether to be at rest relative to the Sun. With what we now know about the cosmology and the cosmic microwave background radiation (CMB), it would make more sense to take the ether to be at rest relative to the CMB.

Thus, the various inertial coordinate systems used by physicists are not needed to implement the relativity of motion. These various inertial coordinate systems are nothing but accounting devices introduced for simplicity and convenience. A single, preferential coordinate system is actually enough for handling all of physics, and the adoption of such a coordinate system would permit us to maintain the 3D picture of the world. I hasten to add that I am not advocating this rather arbitrary way of resolving the ambiguities associated with different choices of coordinates and equal-time hypersurfaces. I am merely saying that this resolution is permitted by physics.

Alternatively, instead of a single, preferential coordinate system, we could use different coordinate systems with different uniform velocities, but always with the same time coordinate (that is, the same equal-time hypersurfaces). This approach to relativity has been advocated by Tangherlini [35], who formulated a pseudo-Lorentz transformation between such coordinate systems. Tangherlini failed to recognize that his coordinate systems are not inertial, that is, Newton's second and third laws are not valid in their standard forms. In this regard, the Tangherlini coordinates suffer from the same deficiency as the Reichenbach coordinates. But if we are willing to tolerate noninertial coordinates, then his deficiency is not a defect, merely an inconvenience.

Relativistic physics permits both a 3D and a 4D picture of the world. And, in practice, physicists use these two views of the world interchangeably. This is readily seen in the alternative formulations of relativistic mechanics. In one formulation (closely analogous to Newtonian mechanics), the equations motion for a particle are written as three differential equations with three dependent variables (x, y, z) and one independent variable (t). In an alternative formulation, the equations are written as four differential equations with four dependent variables (ct, x, y, z, or x^0, x^1, x^2, x^3) and one independent variable (τ, the proper time). The first of these formulations corresponds to a 3D picture of the world and gives us position as a function of time. The second corresponds to a 4D picture of the world and gives us the parametric equation for the worldline, with τ as parameter. Correspondingly, the increment in proper time can be written in two mathematically equivalent forms: $d\tau = (1 - V^2/c^2)^{1/2}\, dt$ and $d\tau = (c^2dt^2 - dx^2 - dy^2 - dz^2)^{1/2}$. The former expresses $d\tau$ in terms of the speed in 3D space, whereas the latter expresses $d\tau$ in terms of displacements along the worldline. Thus, the former corresponds to a 3D picture of phenomena, and the latter to a 4D picture. Furthermore, the former expresses the lapse of proper time as a rate $d\tau/dt = (1 - V^2/c^2)^{1/2}$, or $(1 - V^2/c^2)^{1/2}$ seconds of proper time per second of coordinate time, whereas the latter expresses the proper time as a quantity that evolves along the worldline. These two formulations correspond to what cognoscenti call B time and A time [25].

Thus, the overall situation regarding the 3D and the 4D pictures of the world is essentially the same in relativistic physics and in Newtonian physics. Both pictures are consistent with the physics, and the decision between these views of the world must be made on purely metaphysical grounds.

5.4 The Observable World

For "me," at given space-time point P, the acausal region and the "present" are not observable. It is impossible for me to know for certain what exists in my "present" (however defined) or what exists anywhere in my acausal region. Does the Sun exist now? I am sure it existed 8 min ago (reckoned in an inertial coordinate system at rest relative to me), when the sunlight that reaches me now was emitted by the Sun. But I cannot know what happened to the Sun after that. Maybe it was engulfed in a shock wave from the explosion of a nearby supernova (which will also engulf me 8 min later), and maybe the Sun has already disintegrated and ceased to exist. I can know for certain only what is on and within my past light cone. In principle, I cannot know for certain what is happening in the unobservable world in the acausal region, outside of my past light cone.

The world on and within my past light cone is the *observable world.* Although physics gives us no definite answer about the dimension of "the world," it does give us a definite answer about the dimension of the observable world. The observable world consists of physical objects and their histories (that is, their world lines or worldtubes) within my past light cone. Thus, the observable world is evidently 4D, and this is equally true in relativistic physics and in Newtonian physics.

On the frontier of the observable world lies the *recent observable world*, which is on the past light cone. Astronomers usually call this recent observable world the "observable universe," because it is what they see when they observe with their telescopes, radiotelescopes, or x-ray and gamma-ray detectors. The recent observable world consists of the intersection of physical objects with the past light cone. If the object is 3D, the intersection is also 3D; it comprises a time series of 2D "tomographic" sections through the object taken at times in the past proportional to the radial distance. For instance, the light cone slices through the Moon in a range of times around 1 s ago, the Sun around 8 min ago, the Andromeda galaxy around 2 million years ago, etc. When looking into the distance, the astronomer is quite literally looking into the past – to paraphrase Laplace, "*le passé serait présent à ses yeux*."

This recent observable world is the closest analogue to the Newtonian absolute present.[17] In fact, the Newtonian absolute present can be regarded formally as a limiting case of the relativistic recent observable world with $c \to \infty$; this "opens" the angle of the light cone and makes it into the flat hypersurface $t = 0$, that is, the absolute present of Newtonian physics.

As a manifold, the recent observable world, like the Newtonian absolute present, is 3D. But whereas the Newtonian absolute present has a flat geometry, both intrinsically and extrinsically, the relativistic recent observable world is intrinsically flat, but extrinsically curved – the 3D light-cone hypersurface has zero Riemannian or Gaussian curvature, but it curves into 4D, and one of its principal curvatures is nonzero. As a manifold the recent observable world is 3D, but as an embedded curved hypersurface it is 4D.

Thus, to the question of the dimension of the recent observable world, physics gives us a two-faced answer. The dimension of this recent observable world is 3D or 4D depending on whether we view the geometry of this world intrinsically or extrinsically.

[17] Weingard [30] has argued that the acausal region plays the role of the Newtonian present. This is a bad analogy on two counts: the Newtonian present is a 3D manifold and the Newtonian present is observable, whereas the acausal region is 4D and unobservable. The only thing that is right about Weingard's analogy is that in the limit $c \to \infty$, the acausal region shrinks toward the Newtonian present. But since the past and future light cones also approach the Newtonian present in this limit, this is inconclusive.

References

1. A. Einstein: Reply to Criticisms. In: *Albert Einstein, Philosopher-Scientist*, vol. 2, ed. P. A. Schilpp (Tudor Publishing, New York, 1949) p 684
2. W. Pauli: *Theory of Relativity* (Pergamon Press, London, 1958) p xiii
3. H. Stein: J. Philos. **65**, 289 (1968) p 16
4. C. W. Misner, K. S. Thorne, and J. A. Wheeler: *Gravitation* (W. H. Freeman, San Francisco, 1973) p 6
5. R. Geroch: *General Relativity from A to B* (University of Chicago Press, Chicago, 1978) p 8
6. J. L. Synge: *Talking About Relativity* (North-Holland, Amsterdam, 1970) pp 58–79
7. R. F. Marzke and J. A. Wheeler: Gravitation as Geometry–I. In: *Gravitation and Relativity*, ed. H.-Y. Chiu and W. F. Hoffmann (Benjamin, New York, 1964) pp 54, 55
8. A. Einstein: *The Meaning of Relativity* (Princeton University Press, Princeton, 1955) p 24, 25
9. A. Einstein: *Einstein's 1912 Manuscript on the Special Theory of Relativity* (G. Braziller and the E. J. Safra Foundation, Jerusalem, 1996) p 100
10. H. C. Ohanian: Am. J. Phys. **72**, 141 (2004)
11. H. C. Ohanian: Am. J. Phys. **73**, 456 (2005)
12. A. Einstein: *Relativity, the Special and the General Theory* (Three Rivers Press, New York, 1961) p 27
13. T. P. Krisher et al.: Phys. Rev. D **42**, 731 (1990)
14. P. Wolf and G. Petit: Phys. Rev. A **56**, 4405 (1997)
15. C. Audoin and B. Guinot: *The Measurement of Time* (Cambridge University Press, Cambridge, 2001) p 102
16. C. M. Will: Phys. Rev. D **45**, 403 (1992)
17. J. L. Anderson: *Principles of Relativity Physics* (Academic Press, New York, 1967) p 83
18. V. Fock: *The Theory of Space, Time, and Gravitation* (Pergamon Press, New York, 1964) pp 158–182
19. J. Ehlers, F. A. E. Pirani, and A. Schild: The Geometry of Free Fall and Light Propagation. In: *General Relativity*, *Papers in Honour of J. L. Synge* (Clarendon Press, Oxford 1972) pp 72–74
20. R. Penrose: *The Road to Reality* (Knopf, New York 2005) pp 403, 723
21. H. Minkowski in Lorentz et al.: *The Principle of Relativity* (Methuen, London, 1923) p 75
22. G. Nehrlich: *What space-time explains* (Cambridge University Press, Cambridge, 1994) pp 58–62
23. H. Stein: Philos. Sci. **58**, 147 (1991)
24. T. Sider: *Four-Dimensionalism*, *An Ontology of Persistence and Time* (Clarendon Press, Oxford, 2001) p 3
25. S. McCall and E. J. Lowe: Analysis **63**, 114 (2003)
26. V. Petkov: *Relativity and the Nature of Space-time* (Springer, Berlin, 2005)
27. C. W. Rietdijk: Philos. Sci., **33**, 341 (1966)
28. H. Putnam: J. Philos. **64**, 240 (1967)
29. V. Petkov: Is There an Alternative to the Block Universe View? In: *The Ontology of Space-time*, ed. D. Dieks and M. Redei (Elsevier, Amsterdam, 2006) http://philsci-archive.pitt.edu/archive/00002408
30. R. Weingard: Brit. J. Phil. Sci. **23**, 119 (1972)
31. V. Petkov: Brit. J. Phil. Sci. **40**, 69 (1989)
32. J. S. Bell: *Speakable and Unspeakable in Quantum Mechanics* (Cambridge University Press, Cambridge, 1987) pp 67–80
33. R. Penrose, op. cit., pp 320, 457
34. M. Born: *Einstein's Theory of Relativity* (Dover Publication, New York, 1962) p 254
35. F. R. Tangherlini: Suppl. Nuovo Cimento **20**, 1 (1961)
36. W. F. G. Swann: Rev. Mod. Phys. **13**, 197 (1941)

Chapter 6
Four-dimensional Reality and Determinism; an Answer to Stein

Wim Rietdijk

> *People like us, who believe in physics, know that the distinction between past, present, and future is only a stubbornly persistent illusion.*
>
> *Albert Einstein*

Abstract

1. We argue that the constancy of c implies a realistic now at a distance in all inertial systems, which in turn leads to determinism.
2. We elaborate that denying this violates physical consistency, that requires the concept "velocity".
3. Physics implies models making the world observationally and logically coherent, also as to the experiences of different observers. Velocity and now at a distance are cases in point.
4. Retroaction – which implies a realistic future – can be demonstrated to appear in some experiments.
5. We give a demonstration of determinism via reality at a distance that has an effect here and now too.
6. If the phenomenon of Einstein, Podolsky and Rosen (EPR) does not imply retroaction, it makes necessary nonlocal signals. This is highly relevant to reality and "now" at a distance.
7. The clock paradox shows concrete visible consequences of "merely metrical" relativistic coordination and time flow at a distance.

Keywords: determinism, special relativity, indeterminism, clock paradox, block universe

6.1 Introduction

Figure 6.1 sketches the inertial frames (x, O, t) and (x', A, t') of two observers P and Q, respectively. (The frames imply that Q moves towards P.)

In common-sense language, P in O considers point-event A to belong to his now-at-a-distance, to *reality* at a distance. In turn, Q at A "knows" B (in P's absolute future) to belong to his (determined) reality at a distance.

V. Petkov (ed.), *Relativity and the Dimensionality of the World*, 101–113.

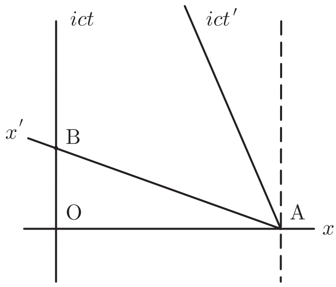

Fig. 6.1 "Now" at a distance: I at O can argue: "For observer Q at A, now for me, my future experience B is reality at a distance; hence it is determined and even realistically existing".

Now the gist of my argument in Ref. [1] was this: P at O (say, I myself) can correctly reason that Q at this very moment $t = 0$ experiences B in my absolute future as *real now* at a distance – that is, determined. Therefore, B is defined for him now at my $t = 0$ and, because being defined, or determined, is an intrinsic (ontological) state, B is determined for me too. Now all point-events in my absolute future can play the part of B for some well-chosen Q in my present at a distance.

Stein, in Ref. [2], virtually objects as follows:

1. A and Q being there "now" for me are merely so in a *metrical*, not a realistic physical sense; now at a distance is at all a problematic concept.
2. The transitivity used by me – A is realistically there and determined for me by being in my distant present, and B is realistically there in Q's present, hence B is defined for me too – is even more problematic.
3. Actually, I can only truly say that (point-)events in the past half of my light cone at O – of which I in principle can have observational knowledge – are definite, "observationally determined".

My answer to Stein will proceed via different paths:

1. I will consider analogues of the situation of Fig. 6.1 with which the argument about reality at a distance can be made extremely cogent and it can be shown that the world would no longer be physically coherent without relativistic now at a distance being realistic.
2. We consider four situations and experiments radically different from that of Fig. 6.1 which imply direct manifestations of future events (retroaction) or demonstrate that we should take relativistic metric at a distance – including that about time – deadly serious in a realistic sense, on pain of violating experience or logic.

6.2 Reality of a Mirror Reflecting My Light Signal Now at a Distance

Consider Fig. 6.2 with the inertial frames (x, O, t) and (x', O, t') of two observers passing each other at O: the first I am myself and my colleague is F. m is the world line of a mirror M, so that M is at rest with respect to me. I sent a light signal L_1

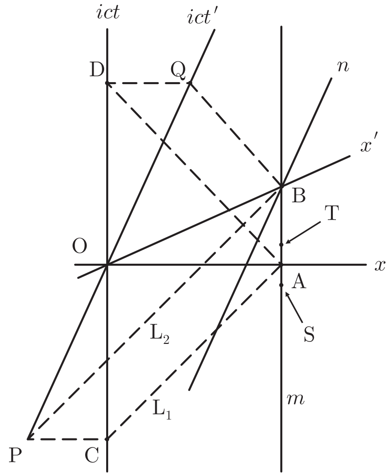

Fig. 6.2 My colleague F [inertial system (x', O, t')] passes me [inertial system (x, O, t)] at O. *For him* a light signal from P via B to Q is reflected at B at the same time as our passing at O. My signal from C via A to D is reflected at A, *for me* at the moment F passes me at O. Hence, both A and B should be realistic and determined events for us at O.

from C to M. It is reflected by M at A and recaptured by me at D. F sent a signal L_2 from P which he recaptures at Q after its reflection by M at B. An essential relativistic feature of both signals is that, for me and F, their velocities on the way there and on the way back are equal, *c*. This manifests itself, *inter alia*, in the equalities CO = OD and PO = OQ, which means that for F and me our signals take equal times for traveling before and after their reflections, in our respective inertial systems. We further consider an observer K with M, at rest with respect to it and to me, and with whom I synchronized clocks earlier.

It should be mentioned that there is some confusion in the literature as to the velocity of light in two opposite directions: some maintain that we only *by convention* assume that not merely the *average* velocity of the signals on their travels towards and from M is *c*, but that on both ways separately it is *c* too. In Ref. [3] we directly proved from an interference experiment that the "two-ways" velocities are indeed equal. (Apart from this, the idea of *un*equal relevant velocities is rather artificial.)

Simple logic now implies that, the traveling times being equal, I can legitimately say at being at O: "*Now* my signal L_1 is reflected by M (at A)". Precisely this causes CO = OD to hold true. If one rejects this one should either abandon the constancy of the velocity of light or the concept of now at a distance at all. Note that the reflection of L_1 on M as a common $t = 0$ "now" event for me and observer K with M *is not a convention but a logical consequence of the experimental fact that c is constant, and the same on L_1's way to M and its way back.*

Then, the crux of the matter is that F, passing me at O, can correctly say to me: "*Now* my signal L_2 is reflected by M at B, in your future, which is real now for me". (A non-essential variant of this thought experiment is that L_2 is not reflected by M but by a mirror N with worldline *n*, still at B. This makes my and F's situations more equivalent.) F at passing O cannot but recognize B and the reflection there to be real now for him on pain of his denying that L_2 takes equal times for its ways to M (or N) and back (PO=OQ). The rest is simple logic. Furthermore, since F would be irrational in assuming that his signal would reflect on M *before event B on m became real (and were still future and undetermined), or when event B would already be past*, he is entitled to physically, not merely metrically, consider his signal to reflect on a realistic mirror in his present but in my future (after A).

Note that the above logic about "now" for the reflection of a light signal could also be substituted by one about a *bullet* fired to some object. Ballistic experts then could calculate: "Now the bullet hits its target" (indeed: now *in our inertial system*). Could one maintain the hit to be so "undetermined" that the experts' calculation *would not make realistic sense but is merely "metrical"?* The hit *appears* at some moment; why not coordinate it in time according to physical law? Law about ballistics, about the velocity of light, or whatever. Physical laws, moreover, are all mutually coherent. Essentially: *if our above argument is incorrect, what then is the physical meaning of the constancy of the velocity of light at all?*

One more thought experiment in the spirit of the above proceeds as follows. Reconsider Fig. 6.2 with A representing the sun at my "metrical" time $t = 0$. An instrument on the sun sends a light signal L to me from such "point-event" A. Now the distance from me to the sun is 150 million km and c=300,000 km/s. Hence, L is 500 s on its way. Well, when I then receive L at D, knowing the numbers, what else can I conclude then except that L *cannot but have been emitted from the sun at the same 500 s ago as when I was at O and A was "metrically" now at a distance for me?* The emission event there should have been very realistic in order to get L at D in time. That is, calculating at D ($t = 500$) about when L was emitted, I cannot but conclude that it should have been at the point-event A that was "merely metrically" $t = 0$ at a distance for me at my being at O. Still, such "theoretical now at a distance" actively sent L to me on pain of our denying either the 150 million km or the 300,000 km/s. Hence A ($t = 0$) will have been physically realistically active at my experiencing it as "now" from my position at O. F can argue similarly, substituting A by B and D by Q.

Vital conclusion: things at O and at the distant A, *both at my $t = 0$, have to be equally realistic (possible conscious local experimenters included) because otherwise natural law about light would no longer tally as to different places and times.*

6.3 Further Verification of the Above Thought Experiment

We can vary the thought experiment of Fig. 6.2 as follows. Let K with M and I again synchronize our clocks before the experiment, so that mine and his clocks both indicate $t = 0$ at our being at O and A, respectively. We further agree that K only sets up M during a very small period before and after his witnessing A ($t = 0$) on *m*. Then I can verify at D that my signal will indeed be recaptured there. Subsequently (in a repeated experiment) I agree with K that he activates M only at S or T at our (K's and mine) common times $t = -1$ and $t = 1$, respectively. Then I will not see my signal L return at all. Conclusion: If I say at O that L is reflected at A *at time zero* in a distant part of my inertial system, I speak realistically physically: my now at a distance is physically active at my *and K's* $t = 0$ by reflecting L. For I can verify at D that the reflection (K's now experience with it included) occurred at point-event A because only there did M ever appear. Similarly, observer F can agree with K that he only installs M at B. Both observers verify that the reflection is realistically there where relativistic metric positions it in space and time.

Especially realize that if K during a short period sets up M, he does so consciously, experiencing "now" – which, in Stein's conception, is the local physical transition from the ontological state "indefinite" to the one "past and definite". That is: at my distant $t = 0$ at point-event A – where the reflection demonstrably occurs because only then and there was a mirror – we not merely see appear the short emergence of mirror M, but also a "now" experience of the local observer K, who legitimately concludes that at the reflection (at his and my "metrical" $t = 0$, or "now") all point-events on m below A are determined by being past. (Again the same thing, *mutatis mutandis*, holds for F with respect to point-event B.)

Note that in the indeterminist's view, "being past and determined" versus "being future and undetermined" are two ontologically, intrinsically, different states of a physical point-event, just as point-events (or "object-events") being either "blue" or "red". This, *inter alia*, holds *for me at the origin O of my inertial system U*. Hence, a plane or curved hyperplane H through O in Minkowski space Mi should separate the two relevant different sets of physical point-events for me now. H need not at all coincide with my system's now hyperplane. Still it would be paradoxical if, in Fig. 6.2, H as it divides Mi in moving in the t direction on all worldlines, *would pass – I still being at O of U and H therefore containing O too – through, say, S while local clocks indicate the "*merely metrical*" time* $t = 0$ *both at O and A*. For a contradiction would appear with the former paragraph in which it appeared that H passed through O and A rather than O and S at my experiencing O and K's experiencing $t = 0$ too. Hence, H cannot exist and all point-events cannot but have the same state: either definite or undetermined. As some are certainly in the first category, all will be.

F should conclude that some H passes simultaneously for him through O and B. Hence, again, H has an impossible property and does not exist.

Stein does not use H. But he does not assume it in spite of the fact that it should play an essential part in his concept of Mi and reality: it cannot *physically and ontologically* be so that, for instance, merely my past light-cone half at O (whose events could *causally* influence me verifiably at O) is determined and the rest of Mi is not, because this would imply my objectively privileged position in the universe. (In footnote 11 of Ref. [2] Stein says that he considers the distinction between past and future to be absolute. This cannot merely be so for one observer!) Stein, though "rather noncommittally", still assumes a particular role of such light cone half (Ref. [2], pp. 8 and 14), be it subjectively. In doing so he neglects an essential point of the relation between different observers. He refrains from integrating the situations, experiences and logic as to the "objective transition" future $\rightarrow$ past of the various observers at different locations and times. This is inconsistent with what physics does as to all other phenomena and situations. Those at different locations in Mi play essential parts, for instance, in Coulomb's and Ohm's laws, and in the equations of motion.

My above integration of mutually distant phenomena is consistent, but by not giving an alternative, and at most hinting at the past half of an arbitrary observer's light cone, Stein actually evades his model's leading to paradoxes. If he refrains from giving arguments about integrating (coherently coordinating) different observers' now experiences, that also refer to the intrinsic physical difference between relevant point-events as to their (un)determined state in Mi, he remains vague: *he does not show what consistent indeterministic model could substitute mine and Putnam's*, leaving the latter without competition as regards consistency. He makes his standpoint about an ontological difference between past and future *unfalsifiable* by not going into details about some hyperplane H and about the ontological state – determined or undetermined – of physical point-events in my "relative" past and future, beyond my light cone. Why not mutually integrate the experiences and conditions of the various observers as to the difference and relations between their objective past and future, in contrast with other differences and relations (such as "whether the local water is boiling or not", or as regards mutually distant interacting charges)?

6.4 As Influences from the Future Sometimes Appear (Retroaction), the Future Itself has to Exist

In Refs. [4–6] I demonstrated that in several experiments retro-causality (influences from events in our absolute future and only operative within Heisenberg's uncertainty margins) should exist on pain of the violation of conservation laws such as with respect to momentum and angular momentum. Now one could consider abandoning such conservation because it might not always be true in situations in which experimental verification is difficult. But the point is that conservation has also been demonstrated by Noether's theorem. Hence, we cannot abandon it.

Because I realize that retroaction is about the last thing for most physicists to be considered plausible, I will give an outline of the proof in Ref. [5].

Consider Fig. 6.3, which is a variant of Young's interference experiment. We vary two features. First, we imagine the experiment on a very large scale so that the momentum carriers (electrons, photons, etc.) need an hour to reach screen T from the slits A and B in S. Second, we arrange that if T is removed, the momentum carriers will hit one of a system of plates P_1, P_2, P_3, ... whose produced parts all pass through C on S, between A and B.

As we know, in the conventional performance of the experiment we will see interference fringes on T. But now remove T so that the plates become operative. Then, because points on the upper sides of the plates can only be reached by waves from A, and the lower sides only by waves from B, there will not be interference in such case. That is, no fringes appear but a more even distribution of impacts, which

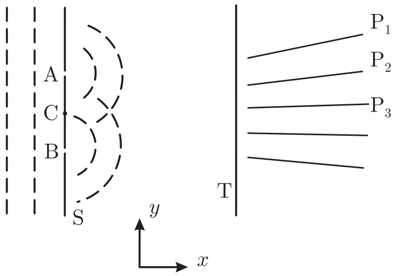

Fig. 6.3 A variant of Young's double-slit experiment. The momentum carriers cannot later than at their interaction with S adjust their y-momenta to whether they should be distributed in the T region either evenly or in a fringe-like way.

also implies a more even distribution of the momentum directions of the systems past A and B. We imagine that generally no more than one momentum carrier at a time passes S.

Now the experimenter with T decides 1 min before the (bulk of) momentum carriers arrive (59 min after their departure from S) whether he will leave T in place or not. If he does, interference fringes will appear. If he does not, the more even distribution will.

Then the crucial point is that conservation of momentum implies that, on their way through the vacuum from S to the T region, the carriers cannot change as to momentum. *Hence, from A and B on they had definite momenta that were already attuned to either an even or a fringelike distribution of impacts, for the momenta could not adjust between S and the T region. But then some influence on the carriers will have enforced such adjustment already at A and/or B: an adjustment to the decision 59 min later by the experimenter with T. That is, retroaction from what has been done with T.* Or rather, from the subsequent impact events.

As to our problem of an existing future the above shows that the future situation of T and the plates is indeed in a position of influencing momentum transfers from S to some momentum carriers passing A and/or B 59 or 60 min earlier. Hence such situation *actively exists* in some way at the carriers' passing S.

6.5 A Thought Experiment Showing Here and Now that the Future Elsewhere Exists

We can devise an experiment in which on the one side processes here and now, and on the other side those at a distance, are interrelated so directly that specific mutually neighbouring observers O and S *can conclude from local phenomena that O's future, which is also now at a distance for his nearby colleague S, is realistic.* (We discussed such thought experiment earlier in Refs. [7] and [8].)

Consider Fig. 6.4, which shows a rotating belt L moved by two shafts through P and Q, respectively. L contains M atoms. Observer O at C is at rest with respect to P and Q. He sees L move near him at velocity v. For him, $M/2$ atoms are in the part

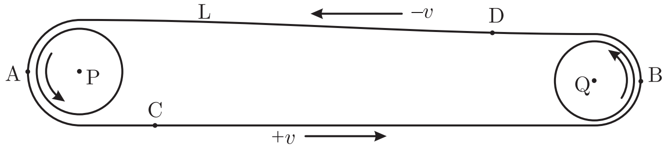

Fig. 6.4 The belt L rotates at velocity v. We find that observers O and S at C, who have different velocities, can logically conclude from the local density of L's atoms that a part of L that for O did not yet pass B, did so very realistically for S.

ACB of L. The present parts A and B of L (for O) are marked A_S and B_S, respectively. L's halves ACB and BDA will tend to contract for O by a factor $\sqrt{1-v^2/c^2}$ on account of the velocity v. Actually, they cannot do so because PQ does not contract for O. This causes L's material to be stretched for O by a compensating factor $1/\sqrt{1-v^2/c^2}$.

Now also consider a second observer S at C who moves with the local part of L at velocity v. Since S is at rest with respect to part ACB of L, such part is not Lorentz contracted for him. But it is still stretched (and hence, diluted) because stretching is an objective physical condition, irrespective of a local observer. Furthermore, $PQ \approx AB$ is Lorentz contracted for S. Therefore, the number N of L atoms on part ACB of L for S at C is

$$N = \frac{1}{2}M\sqrt{1-v^2/c^2}\,\sqrt{1-v^2/c^2} = \frac{1}{2}M\left(1-v^2/c^2\right). \tag{6.1}$$

Hence, for S there are

$$\frac{1}{2}M - N = \frac{1}{2}M - \tfrac{1}{2}M(1-v^2/c^2) = \frac{1}{2}Mv^2/c^2 \tag{6.2}$$

atoms of L which already passed B and the Q region, and for S at C are "already" in the part BD of the higher part of L. That is, the marked B_S is at D for him. We take AP and AC negligible as compared with PQ, so A_S is at A for both O and S at C.

The above makes clear that we cannot "dismiss" the relevant $\frac{1}{2}Mv^2/c^2$ distant atoms of L, which S locates "now" between B and D, as "merely metrically" being there. On the contrary, they should have passed B and the Q shaft *very realistically*. For at C *in his direct neighbourhood* S finds a dilution of L, which has to be combined with a Lorentz contraction of PQ. Therefore, it is *absolutely impossible for S to locate $M/2$ of L's atoms on ACB*, whereas O indeed does so. (For O, B_S is at B.)

The conclusion is that the presence of the relevant $\frac{1}{2}Mv^2/c^2$ atoms of L between B and D for S is equally realistic for him as the stretching and atomic density of L are in his immediate here and now at C. At the same time O cannot yet locate such atoms as already having passed B without his also being confronted with a *local* impossibility as to the density of L atoms on the non-contracted ACB. This means that the relevant $\frac{1}{2}Mv^2/c^2$ atoms that did not yet pass B for O at C, did so for S as realistically as is the atom density S establishes at C. We cannot dismiss as "a merely metrical phenomenon" the circumstance that, for S, many atoms passed B which did not yet do so for O, for if we ignore it as a physical phenomenon at a distance this results in a direct contradiction with logic and experiment about density here and now at C.

Our conclusion is that the above thought experiment shows that (our, or O's) denying the physical reality of what is relativistically "now-at-a-distance" for some other observer (say, S) next to us or to O can lead to a direct violation of local experience and logic. Hence, such distant now for him should be considered as realistic and determined, in spite of its still being future for us.

6.6 EPR is Highly Relevant to the Problems of Reality at a Distance and Determinism

Consider Fig. 6.5, which is a 4D picture of the experiment leading to the paradox of Einstein, Podolsky and Rosen. E denotes the "splitting-event" where two correlated orthogonal photons P and Q separate. Say, their polarizations are measured by crossed analysers at A and B, respectively. We take EA $<$ EB. It is well known that if the analysers at A and B are attuned to letting pass photons polarized in the y- and z-directions, respectively (they are "crossed"; also compare the x-direction in the figure), P and Q either both pass or none does. Then we can conclude that if P passes at A, Q will pass at B. We know for certain that, between B_1 and B, Q is polarized in the z-direction. Now relativity requires that physical laws equally hold in all inertial systems. If Q is z-polarized with certainty, P will be polarized in the y-direction, also in other inertial frames, i.e., if Q is z-polarized at B_2, P will be y-polarized at A_1. But then we can reason further that Q will be z-polarized at B_3, and so on. The conclusion is that P and Q were pre-polarized in the directions y and z, respectively, from E on. This means retroaction from the A and B measurements to splitting event E.

We can add that if, say, P and Q did arrive at A and B, respectively, while having mutually orthogonal polarization directions that, however, made angles of $\frac{1}{6}\pi$ with the y- and z-directions of the crossed analysers at A and B, whereas P would interact merely *locally* at A and Q did so at B, it is easily seen that a 100% correlation of the two measurement results would certainly not appear. Still it does. This excludes

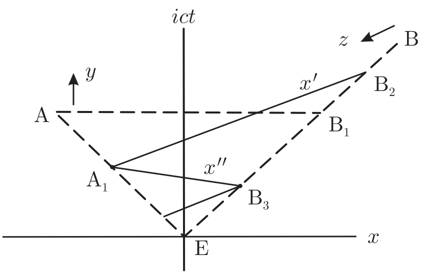

Fig. 6.5 An EPR setup. The correlated photons may either influence each other by retroaction (feedback) via E or do so possibly via instantaneous signals. In both cases determinism can be inferred.

a merely local explanation of any hidden variable at stake here, in agreement with Bell's result in Ref. [9]. (Mind: the hidden variable decides on either passing or not.)

Of course, the above specimen of retroaction again demonstrates that the future "is", because it appears to cast its shadows before. That is, such future sometimes is physically active.

Some might consider the above retroaction from A and B to E, or feedback contacts A → E → B and B → E → A, unnecessary because they feel *direct* nonlocal influences from A to B and/or conversely to be at stake. Apart from our not being in a position to establish them equally arguedly as retroaction above, we should realize that direct (instantaneous) physical influences, say, from B or B_1 to me at A *would again irrefutably prove the reality of my distant "now"* (i.e. *B or B_1) in any inertial system*. For mind that we can take A and B so that AB plays the part of the x-axis in my now-hyperplane, and vary them according to how I may choose my inertial system. And reality of my distant present in various inertial systems leads to determinism because of our earlier argument.

6.7 The Clock Paradox too Demonstrates that Physical Processes at a Distance Realistically Behave According to "merely metrical" Relativistic Relations

Consider Fig. 6.6, which sketches the clock paradox. OP is the world line of the earth, on which is observer A. His twin brother B follows world line OB_1 after having been accelerated to a velocity v near O. In and near B_1 he decelerates and returns, eventually arriving at P. The essence of the so-called clock paradox is now

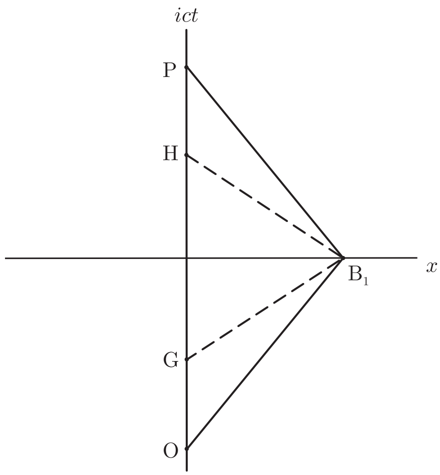

Fig. 6.6 The clock paradox shows that relativistic time retardation at a distance is realistic. Since it has to be integrated with (and more than compensated by) a change of now-at-a-distance for one of the two relevant observers, in order to explain the paradox, such "now" should be equally real as the retardation.

the following. Arguing exclusively special-relativistically, we seem to be confronted with a paradox. For B moves with respect to A and his clock and ageing process will be retarded by a factor $\sqrt{1-v^2/c^2}$ with respect to A's clock, on B's way to B_1 and on his way back to the earth. Indeed, A will find in P that B, his clock and his age will be in accordance with the above retardation factor. But now the paradox is that, for B, the earth and A were also moving with respect to him (B) at velocity v, which at first sight would suggest A's clock and ageing also to stay behind of B's at their meeting at P.

One can solve the paradox by considering B's inertial system during his journey. In particular note the change, rotation, of the x-axis of B's inertial system at his de- and acceleration near his turn-around point-event B_1: such axis rotates from position B_1G to position B_1H. Reasoning from the standpoint of B now leads to his following conclusion about the course of time on the earth during his absence. First, during his covering OB_1, a time corresponding to OG passed on the earth; second, during his mere de- and acceleration near B_1, a time corresponding to GH passed on the earth according to B's account. Finally, during his (B's) way back to the earth, a period corresponding to HP elapsed there. Note that A, on the other hand, never experienced something like the rotation of B's x-axis at B_1, which corresponds to "many years on the earth". For A and the earth do not change their inertial system during B's journey because they never *accelerate* correspondingly.

The crucial point with respect to our determinism problem can now be discussed. Stein actually reproaches Putnam and me that we take relativistic time and metric at a distance seriously in a physical and realistic sense. In essence, we have a similar point here as that with the belt L in section 6.5: how physically realistic should things be or proceed at a distance, as they are coordinated by relativity, in order to avoid paradoxes here and now? In section 6.5 the answer was very direct and precise: they should be completely realistic at a distance too, and in conformity with relativity and its spacetime metric.

In our present thought experiment we see something analogous: a phenomenon so realistic as clock retardation (that has been experimentally verified and in the present experiment *can be seen at the return at P*), also the one of A with respect to B during A's covering (in time) OG and HP, *can only be compensated by B's changing his now-at-a-distance (on the earth) and x-axis by the latter's rotating from position B_1G to B_1H*. Hence we can consider the corresponding "jump in time" GH *to refer to equally realistic situations as the clock retardation itself, which it essentially complements*. Why should G and H refer to less realistic situations in B's coherent model of the world than the phenomenon of the clock retardation with A for B that corresponds to OG and HP? *Mind that, conversely, B's clock retardation for A is so realistic that it can be verified at their meeting at P! Why should the times of A for B (e.g., as to G, H and P) be taken less seriously than conversely?*

In particular realize that it is not the "saltus" G $\rightarrow$ H which is realistic on the earth but G and H as situations, because they are as real for B as A's clock retardations corresponding to OG and HP.

6.8 Comments on Ohanian's Criticism

In his contribution to this book, Ohanian criticizes my and others' block-universe (4D) position. He gives various arguments, which can best be answered point by point:

1. In Ref. [3] it has been demonstrated from interference experiments that the velocity of light is the same in opposite directions, and that there is nothing conventional about this.

   In coherence with our argument going with Fig. 6.2 this (for any observer) produces an objective physical criterion as to *what simultaneity at a distance and its physical reality actually mean*, with all their consequences with respect to 4D.

   Such simultaneity has to be realistic in an arbitrary inertial system on pain of our ignoring the very, equally realistic, constancy of c in it.
2. Ohanian is of the opinion that, *inter alia*, coordinate systems that order physical phenomena simply and coherently, such as inertial systems going with an observer, do not correspond to a more realistic model of the world than, say, "contorted" systems with arbitrarily curved axes and $t = 0$ hyperplanes.

   Now my question is: "Do we have *in the first place* any criteria as regards (preliminary) "reality" and (preliminary) correct models about how the world works, *apart from* what variants among alternative ordering models or schemes (coordinate systems included) are most simple and coherent? The answer is no.

   That is, inertial systems with rectilinear axes such as going with observers in the special-relativistic way are superior in coordinating reality by their very contributing to the simplest and most coherent model of the world (reality), also as to the relativity principle, the constancy of c, an isotropic world, and the simplest and most coherent equations of motion and dynamics. Should we subordinate all of this to, say, some "freedom of choice", e.g., of philosophers preferring a 3D model of the world, and indeterminism, to a block universe? Do "contorted" coordinate systems and corresponding equations of motion, or noncommittal "conventions" about simultaneity – or whatever models or schemes at all –, derive any plausibility of being "equivalent" (as to any coherent reality) to far more simple and coherent relativistic ones, merely because some "choose" them without contradictions with experiment appear?
3. A concrete case in point is: If Lorentz' and Einstein's models of relativistic length contraction, of Michelson–Morley, etc. both explain the experiments but Einstein's one does so simpler and more coherently, we certainly have no longer a "free choice" as to corresponding models of reality.
4. In trying to refute the block-universe position Ohanian actually shows the weakness of the 3D one by his actually having to revert to pre-relativistic conceptions. *That is, he has to abandon the very physical model of reality that most simply and coherently goes with Special relativity.*
5. Ohanian is less than enthusiastic about my considering more than one observers (with inertial systems), and is even more sceptical about my also considering

an observer to change his velocity in the course of a thought experiment. (His note 16.) But why should we not be allowed in thought experiments to imagine various alternative situations?

6. Generally, there is in physics a well-known positivistic-formalistic tendency towards our abandoning understandable models of reality. This causes various concrete choices in order to make things imaginable to be evaded. Now it is far from being widely realized that the concomitant "poly-interpretability" and *our acceptance of what is not coherently imaginable* can actually amount to our *no longer being confronted with paradoxes*. For we abandon the requirement that things are plausible and tally precisely! Not meeting with paradoxes, we no longer are forced to critically reconsider our presuppositions or changing our conceptions. Hence, positivism is also a weapon of conventional thinking.

All of this also refers to our possible ignoring what among the coordinate systems Ohanian considers actually correspond to the simplest and most coherent model of reality.

In all, common sense can be indispensable for our abandoning prejudices and finding flaws in conventional thought.

References

1. C.W. Rietdijk, *Phil. Sci.* 33 (1966) 341–344.
2. H. Stein, *J. Phil.* 65 (1968) 5–23.
3. C.W. Rietdijk in M.C. Duffy (Ed.), *Phys. Interpret. of Relat. Theory*, 1996 Conference Proceedings, Late papers (London, 1996) 243–246.
4. C.W. Rietdijk, *Found. Phys.* 8 (1978) 615–628.
5. C.W. Rietdijk, *Found. Phys.* 11 (1981) 783–790.
6. C.W. Rietdijk, *Ann. Fond. L. de Broglie* 16 (1991) 91–96.
7. C.W. Rietdijk in M.C. Duffy (Ed.), *Phys. Interpret. of Relat. Theory*, 1990 Conference Proceedings (London, 1990) 246–255.
8. C.W. Rietdijk in C. Cormier-Delanoue a.o. (Eds.), *Courants, Amers, Écueils en Microphysique* (Paris, 1993) 361–378.
9. J.S. Bell, *Physics* 1 (1964) 195–200.

Chapter 7
Relativity, Dimensionality, and Existence

Vesselin Petkov

Since the [relativity] postulate comes to mean that only the four-dimensional world in space and time is given by phenomena, but that the projection in space and in time may still be undertaken with a certain degree of freedom, I prefer to call it the postulate of the absolute world.

H. Minkowski [1, p. 83]

The basic idea is to present the essentials of relativity from the Minkowskian point of view, that is, in terms of the geometry of space-time ... because it is to me (and I think to many others) the key which unlocks many mysteries. My ambition has been to make space-time a real workshop for physicists, and not a museum visited occasionally with a feeling of awe.

J. L. Synge [2, p. vii]

7.1 Introduction

A 100 years have passed since the advent of special relativity and 2008 will mark another important to all relativists anniversary – 100 years since Minkowski gave his talk "Space and Time" on September 21, 1908 in which he proposed the unification of space and time into an inseparable entity – space-time. Although special relativity has been an enormously successful physical theory no progress has been made in clarifying the question of existence of the objects represented by two of its basic concepts – space-time and world lines (or worldtubes in the case of extended bodies). The major reason for this failure appears to be the physicists' tradition to call such questions of existence philosophical. This tradition, however, is not quite consistent. In Newtonian mechanics physicists believe that they describe real objects whenever they talk about particles – one of the basic concepts of Newtonian physics. The situation is the same in quantum physics – no one questions the existence of electrons, protons, etc. Then why should the question of existence of worldtubes (representing particles in relativity) be regarded as a philosophical question?

The most probable answer a physicist would give is that the concepts of space-time and worldtubes belong to the 4D representation of special relativity, whereas in its 3D formulation these concepts are not used. Since both representations of special relativity are equivalent it appears that one should not worry about the existence of space-time and worldtubes. Most physicists and especially relativists appear to believe that by emphasizing the equivalence of the 3- and 4D descriptions of the

V. Petkov (ed.), *Relativity and the Dimensionality of the World*, 115–135.

world provided by special relativity the issue of the existence of space-time has been shown to be a nonissue. However, as we will see that would be quite a premature attempt to close such a fundamental issue.

The 3- and 4D representations of special relativity are equivalent, but I think it is a valid question to ask which of them is the adequate description of the world. One is naturally tempted to immediately question the validity of such a question by pointing out that these are just two descriptions and to ask which is the right one is meaningless. For instance, one could explain that such a question is as meaningless as to ask whether the Newtonian or the Lagrangian formulation of classical mechanics is more adequate. Once again this situation clearly demonstrates that each case in science (and not only in science) should be dealt with separately. The equivalence of the two representations of special relativity are drastically different from the equivalence of the Newtonian and Lagrangian formulations of classical mechanics. The 3D formulation of special relativity represents reality as a 3D world which evolves in time, whereas according to the 4D formulation the world is 4D with time entirely given as the forth dimension. As the world is *either* 3D *or* 4D, it is clear that either the 3D or the 4D description of the world is the correct one in a sense that only one of them correctly reflects the dimensionality of the world.[1] Also, physical bodies are *either* 3D *or* 4D objects. Therefore, not only is the question of the dimensionality of the world and the physical objects not a nonissue, but it is one of the most fundamental issues of the twenty-first century physics. Moreover, it is natural to address the question of the dimensionality of the world on the macroscopic scale according to relativity first, before dealing with the reality of extra dimensions introduced by more recent physical theories.

The main purpose of this paper is to demonstrate that the analysis of the kinematical effects of special relativity holds the key to answering the question of the dimensionality of the world. It is shown that these effects and the experiments which confirmed them would be impossible if the world were 3D. Section 7.2 shows that relativity of simultaneity, conventionality of simultaneity, and the existence of accelerated observers in special relativity would be impossible if the world were 3D. Section 7.3 deals with the dimensionality of physical objects and demonstrates that the relativistic length contraction and the twin paradox would be impossible if the physical bodies involved in these relativistic effects were 3D objects.

7.2 Special Relativity is Impossible in a Three-dimensional World

Perhaps most physicists would disagree with the statement in the section title. They would probably point out the well-known fact that special relativity was initially formulated in a 3D language. However, the only thing this fact says is that the

[1] Both the 3- and 4D descriptions of the world are correct in terms of their adequate descriptions of the physical phenomena, but in terms of the dimensionality of the world they cannot be both correct since the world is not both 3- and 4D.

relativistic effects can be given a 3D *description*[2]; it says nothing about whether or not these effects would be possible in a 3D world.

The best way to prove the statement in the section title is to analyze the kinematical effects of special relativity in order to reveal their true *physical meaning*. Such an analysis is especially needed since all physics books on relativity merely *describe* the relativistic effects without addressing the questions of the dimensionality and existence of the physical objects involved in these effects. Once those questions have been explicitly dealt with it becomes evident, as we will see below, that the physical meaning of the kinematical relativistic effects is profound – these effects turn out to be *manifestations* of the 4D of the physical world. That is why they would be impossible in a 3D world.

7.2.1 Relativity of Simultaneity is Impossible in a Three-dimensional World

Let us start the analysis of relativity of simultaneity with the question "What is the dimensionality of the world according to relativity?" As the observable world is 3D it is understandable why according to the widely accepted view, called presentism, reality is also a 3D world, which exists at the constantly changing present moment. But the dimensionality of the *observable* world tells us nothing about the dimensionality of the world *itself* since the observable world is not what is real according to the presentist view. It consists only of past events[3] – the past light cone. Therefore,

---

[2] Obviously, a description itself does not tell anything about the dimensions of what is described. For instance, the $x-y$ plane can be described either in 2D or 1D language (regarding the y dimension as a parameter).

[3] "The key-word in relativity is event" [3, p. 105]. "An event marks a location in space-time" [4, p. 10]. In most relativity books "event" is used as a synonym of "worldpoint" or "space-time point" [2, pp. 5–6], [2, p. 6], [6, p. 2–1, p. GL-4], [7, p. 427], [8, p. 53], [9, p. 4], [10], [11, pp. 1, 9], [12, p. 25], [13], [14, p. 23], [15], [16, p. 66]. This appears to be quite natural since an event, defined as a worldpoint, is determined only by its space and time coordinates and is not associated only with physical bodies or phenomena. The definition of a concept is of course a matter of convention and, for example, one can define "event" to mean "a worldpoint at which a physical object is located". However, such a definition invites a number of misconceptions. For instance, one is tempted to talk about *different* events that happen with the *same* body. But since a body is implicitly regarded as a 3D object it would mean that different events happen with *the same 3D object*, which makes no sense in relativity where a physical body is represented by a *4D* worldtube. A worldtube, or simpler, a world line consists of different world points, which represent *different 3D objects*. Therefore, to talk about different events happening with the same 3D body is clearly wrong in the framework of relativity (but it is fine in the pre-relativistic physics). The same physical body in relativity means the same 4D object – the worldtube of the body. The different events which constitute the worldtube involve different 3D objects – the physical body at different moments of its proper time. In this paper I will use the widely accepted definition of event as a world point, but will make that definition more explicit – an event is defined as a 3D object, a field point, or a space point at a given moment of time [17, p. 56]. I would like to stress specifically that all events of space-time (like all points of any dimensional space) have the *same status of existence*; otherwise, if the events of just one 3D hypersurface were existent, one could not talk about space-time at all.

the observable 3D world does not constitute even a 3D world since such a world consists of all events that are *simultaneous* at a given moment, whereas the observable world is a set of events belonging to *different* moments of time (the more distant an event is the more in the past it belongs).

If we assume that reality itself (not only the observable part of it) is also a 3D world, it can be easily shown that relativity of simultaneity is impossible. As a 3D world is defined as everything that exists *simultaneously* at the present moment, it follows that the three dimensional world (as the only existing entity) will be common to all observers[4] in relative motion; therefore they will have a *common* set of simultaneous events, which means that relativity of simultaneity is indeed impossible. The conclusion that relativity of simultaneity is possible only in a 4D world seems unavoidable provided that existence is regarded as absolute (frame- or observer-independent) – the observers in relative motion can have different sets of simultaneous events only in a 4- (or higher-) dimensional world; these sets constitute different 3D "cross-sections" of such a world.[5] In this sense relativity of simultaneity is a manifestation of the 4D of the world.

The conclusion that relativity of simultaneity implies a 4D world looked unavoidable to Minkowski [1], Rietdijk [24], Putnam [25], Maxwell [26], and to all who agree with the above argument.[6] However, one can formulate three objections to the relativity of simultaneity argument.

The first objection was raised by Stein [30, 31]. He pointed out that the relativity of simultaneity argument was flawed since it employed the pre-relativistic concept of distant present events, whereas according to the relativistic division of events into past, present, and future one can attribute presentness only to a single event (Fig. 7.1). The relativity of simultaneity argument is clearly based on the pre-relativistic view on reality that only the present, defined as the 3D world at the present moment, is real. However, the present is the set of *distant present events* – everything that exists simultaneously at the moment "now". Therefore the argument

---

[4] "**The word 'observer' is a shorthand way of speaking about the whole collection of recording clocks associated with one free-float frame.** No one real observer could easily do what we ask of the 'ideal observer' in our analysis of relativity. So it is best to think of the observer as a person who goes around reading out the memories of all recording clocks under his control. This is the sophisticated sense in which we hereafter use the phrase 'the observer measures such-and such'." [4, p. 39] (see also [6, p. 1–18]). The concept "observer" (or "experimenter" [7, p. 428]) is widely used in relativity [8, p. 95], [12, p. 48], [14, p. 23], [16, p. 67] [18] in many cases as a synonym of "reference frame" [11, p. 2], [20–23]. If during an experiment the observer is located at the origin of an inertial reference frame, which has been constructed in the standard way (see for instance [9, p. 6]), then there is no difference between "observer" and "reference frame".

[5] As sometimes physicists and philosophers are inclined to think that the 3D view can somehow be made consistent with the relativistic picture of the world they should obviously explain how relativity of simultaneity is possible in a 3D world, which by definition contains just one set of simultaneous events, whereas two observers in relative motion have different sets of simultaneous events.

[6] Einstein himself [27], Weyl [28], Weingard [29], and a number of physicists and philosophers of science view relativity as clearly supporting the 4D view. In 1952 Einstein wrote: "It appears … more natural to think of physical reality as a 4D existence, instead of, as hitherto, the *evolution* of a 3D existence" [27].

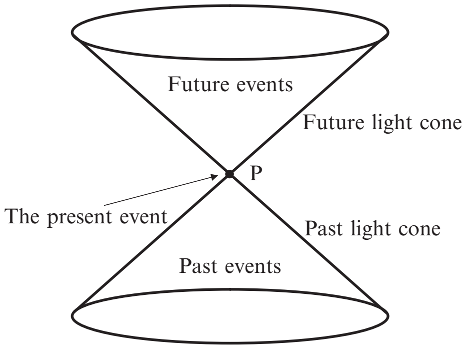

Fig. 7.1 In relativity the division of events into past and future is with respect to *one* event, which is considered present.

that due to relativity of simultaneity two observers in relative motion have different presents, which is possible only in a 4D world, does employ the concept of distant present events.

Due to the fact that the 3D world is the set of all distant present events Stein's objection is in fact directed against the view that reality is a 3D world [17, p. 123], [32]. Thus, by demonstrating that the concept of distant present events (and therefore the view that the world is 3D) contradicts special relativity, Stein de facto sharply raised the question of the dimensionality of the world according to relativity but did not address it.

The same objection is also raised by Ohanian [33]: "[I]n relativistic physics there is no absolute time, and no physical meaning can be attached to simultaneity at different locations." That is precisely the case, since "in special relativity, the causal structure of space-time defines a notion of a "light cone" of an event, but does not define a notion of simultaneity" [34]. But the crucial question is "What is the physical meaning of the relativistic fact that 'no physical meaning can be attached to simultaneity at different locations'?" As we have seen the immediate implication is obvious – the world cannot be 3D since it is defined as everything (at different locations) that exists *simultaneously* at a given moment. If reality were a 3D world, then a clear physical meaning had to be attached to simultaneity at different locations – a class of events (at different locations) would be simultaneous because it is only these events that are real at the present moment. Ohanian did not explicitly rule out the 3D of the world. Instead he took a careful position: "The physical world inhabits this 4-D space-time, but the physical world does not occupy all of space-time – it occupies only a subspace of space-time. The dimension of the physical world is the dimension of this occupied subspace. It is obvious that this occupied subspace is at least 3-D." It turns out, however, that this view is inconsistent with relativity for the following two reasons:

- If the physical world is represented by a 3D subspace of space-time, the contradiction with relativity, as we have seen, is inescapable – the set of simultaneous events, which constitute the 3D world, would be common to all observers in relative motion and therefore no relativity of simultaneity would be possible.

Even if the 3D physical world is regarded as somehow “more existent” than the remaining part of space-time, relativity of simultaneity would again be impossible – what is “more existent” for one observer must be also “more existent” for all other observers in relative motion with respect to the first one which means that simultaneity would be absolute. An assumption that every observer has his own set of “more existent” simultaneous events amounts to a relativization of existence which, as we will see below, contradicts relativistic facts which are not based on relativity of simultaneity.

- If the physical world occupies not all of space-time, but a 4D subspace of it – the acausal space-time region outside of the light cone at an event (Fig. 7.1) – it immediately follows that the events in the past and future light cone are as real as the events lying outside of the light cone [29]. To see this assume that the physical world is represented by all events lying outside of the light cone at event P (Fig. 7.2), i.e. all events that are not causally related to P. On this view the events in the past and the future light cone are not real. But if we consider a second light cone at event P', most of the past and future events associated with the first light cone lie, as seen in Fig. 7.2, in the space-time region outside of the second light cone and are therefore real. As for any event P the division of events into past and future is invariant (since it is an intrinsic, absolute feature of space-time), it follows that what is real for an observer at P' should be real for an observer at P as well. Therefore, if one starts with the assumption that the physical world is the space-time region outside of the light cone at an event, it does follow that the physical world is represented by all events of space-time.

The second objection against the relativity of simultaneity argument is based on the conventionality of simultaneity, which itself follows from the (invariant) relativistic division of events into past, present, and future (Fig. 7.1). Weingard [29] was the first to raise this objection[7] against the relativity of simultaneity argument used by Putnam [25]. Ohanian’s formulation of this objection is [33]:

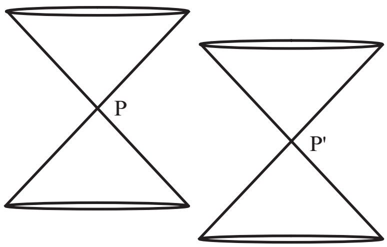

Fig. 7.2 If one assumes that only the events lying outside of the light cone at P are real, by the same argument the events lying in the area outside of the light cone at P' (which contains most of the events of the past and future events of the light cone at P) are also real and therefore, all events of space-time are real.

[7] Despite criticizing Putnam for his conclusion that relativity of simultaneity implies a 4D world Weingard arrived at the same conclusion on the basis of conventionality of simultaneity.

> Any of the hypersurfaces in the 4-D acausal region between the absolute past and future regions can be regarded as an equal-time hypersurface for the point P, provided that the surface includes the point P and that every point on the surface is in the acausal region of every other point on the surface (that is, the surface is "spacelike"). Thus, in relativistic physics there is no unique, physically distinguished definition of simultaneity. To decide what is simultaneous, we need to adopt some prescription for the construction of equal-time hypersurfaces, or, equivalently, a procedure for clock synchronization. Simultaneity then becomes a matter of convention, without any direct physical significance.

As seen from the quote this objection is an elaboration of the first one since it emphasizes that no 3D hypersurface in the space-time region outside of the light cone at a given event is in any way physically distinguishable from the other hypersurfaces in the same space-time region. This means that no hypersurface is privileged on account of its being associated with a single 3D world, which is a set of distant present events; if reality were a 3D world, then the hypersurface associated with it would be naturally privileged. That is what relativity tells us, but again the crucial question is "What is the physical meaning of the relativistic fact that 'simultaneity becomes a matter of convention, without any direct physical significance'?"

Like in the case of the first objection an immediate implication from that relativistic fact is that the physical world cannot be 3D. If the world were 3D it would be represented by one hypersurface (a set of simultaneous events) that is physically distinguishable[8] from the other hypersurfaces in the space-time region outside of the light cone at event P, which is not the case. A second argument is the following – since one is free to choose which hypersurface to regard as a set of simultaneous events at P, which means that simultaneity is a matter of convention, it follows that what exists would be also a matter of convention if reality were a 3D world (defined as the set of simultaneous events at a given moment) [29], [35]. Although this argument against the 3D of the world looks quite convincing (since what exists does not depend on our free choice) it seems not everyone agrees with it. For instance, Ohanian [33] offers the following counter-argument: "[I]f the 3-D world is unobservable, I am content to fix it by convention." If what exists is a 3D world it is indeed unobservable at the moment it exists, but does this mean that its existence depends in any way on our choice?

Although viewing reality as a 3D world directly contradicts relativity (since such a view implies that a hypersurface in the space-time region lying outside of the light cone at an event is physically distinguished), let us assume, for the sake of the argument, that reality is indeed a single 3D world represented by a hypersurface that is orthogonal to the worldtube of an inertial observer. This means that the Einsteinian synchronization of distant clocks has been used (i.e. the back and forth velocities of light have been taken to be equal). In this case we will have the ordinary relativistic effects (relativity of simultaneity, length contraction, time dilation, etc.). The experiments designed to test these effects have confirmed them. Let us now choose a non-Einsteinian synchronization of distant clocks, which would mean that the hypersurface representing the 3D world would not be orthogonal to the

[8] It would be physically distinguishable merely because there would be just one hypersurface that represents the single 3D world that exists.

worldtube of the inertial observer and the back and forth velocities of light would not be equal. All relativistic effects would be again derived, but the expressions for length contraction and time dilation, for instance, would be more complicated [36]. If we decide to perform an experiment to test length contraction, we should expect that it would confirm either the expression derived on the basis of the Einsteinian synchronization of distant clocks or the expression employing the non-Einsteinian synchronization. The reason for this expectation is clear – the hypersurface that represents the existing 3D world would either intersect the worldtube of the measured meter stick at a given angle ("cutting off" a 3D meter stick of a given length) or at a *different* angle ("cutting off" another 3D meter stick of a *different* length). Thus, the fundamental belief that what exists does not depend on our choice has observational consequences if the world is 3D. It should be stressed that if reality were a 3D world the experiment would confirm only one of the expressions for length contraction.

However, if we perform that experiment it would confirm both expressions for length contraction. Does this mean that it is we who decide whether or not the hypersurface representing a 3D world would be orthogonal to the worldtube of the inertial observer? Yes, it does, but not because we can fix the 3D world by convention. The reason is that the physical meaning of conventionality of simultaneity also turns out to be profound. Simultaneity can be a matter of convention "without any direct physical significance" only if the space-time region outside of a light cone at an event exists as a whole. In this case conventionality of simultaneity is trivial (with no physical significance) simply because the *whole 4D region exists* and we are free to choose any hypersurface in it (because we have from where to choose!). But as we have seen, the existence of the space-time region outside of the light cone implies the existence of the whole space-time. Thus, conventionality of simultaneity is possible only if space-time is a real 4D entity. If reality were a 3D world, simultaneity would be absolute and no conventionality would be possible – in that case there would exist just one 3D world (a single set of simultaneous events) and one would not have from where to choose his 3D world.

The third objection against the relativity of simultaneity argument is more philosophical in nature. It calls for the relativization of existence. As mentioned above, the relativity of simultaneity argument implicitly regards existence as absolute – having different sets of simultaneous events two observers in relative motion have different 3D worlds which means that both 3D worlds must exists for every observer (existence is absolute!). But this is only possible in a 4D world. The objection is: "Why should existence be absolute?" Relativity relativized motion, simultaneity, and now it is the turn of existence to be relativized. If this is done, each of the observers in relative motion would claim that only his 3D world would exist and would deny the existence of the 3D worlds of the other observers. This relativized version of the 3D view preserves the 3D of the world but is undefendable for the following reasons:

- It is in obvious contradiction with relativity since the 3D world of an observer would be represented by a hypersurface (a set of distant present event) lying outside of the light cone, which would be physically distinguished from the other hypersurfaces in the same space-time region.

- The relativization of existence is based on relativity of simultaneity and has no justification in relativistic situations where no relativity of simultaneity is involved. Moreover, as we will see below, the relativized version of the 3D view, which regards the world and the physical objects as 3D, is in direct contradiction with conventionality of simultaneity, the existence of accelerated observers in special relativity, and the twin paradox.

7.2.2 Conventionality of Simultaneity is Impossible in a Three-dimensional World

We have seen that conventionality of simultaneity is a consequence of the relativistic division of events into past, present, and future (Fig. 7.1) – any observer at event P is free to choose which hypersurface lying outside of the light cone at P can be regarded by him as the set of events that are simultaneous for him at P. As discussed above the profound physical meaning of this freedom is that the *whole* 4D space-time region outside of the light cone at P must exist. That is why – because it exists – we can choose different hypersurfaces from it. Otherwise, if reality were a 3D world no such freedom in choosing different hypersurfaces, representing different 3D worlds, would be possible.

One may be left with the impression that it is the 4D representation of special relativity and particularly the concept of a light cone, which demonstrated that conventionality of simultaneity is impossible in a 3D world. Moreover, it is generally believed that the 4D picture of the world itself became possible only after Minkowski's talk on space and time [1] in 1908. In fact, that picture is logically contained in the original formulation of special relativity by Einstein in 1905. Minkowski was the first to realize that the relativistic effects are manifestations of a world of higher dimensions – a 4D world with time being the extra dimension. That is why he pointed out that the essence of special relativity is not relativity of space, time, and other physical quantities as the principle of relativity had been interpreted to mean in the early years of special relativity. According to Minkowski that principle should be replaced by 'the *postulate of the absolute world*' [1, p. 83] since a rigorous analysis of the physical meaning of the relativity principle reveals that reality is an absolute (frame-independent) 4D world.

And indeed if we explicitly address the question of the physical meaning of the relativistic effects (assuming that we do it, say, in 1906) it does become evident that these effects are manifestations of the 4D of the world. We have seen this in the case of relativity of simultaneity – once the key role of simultaneity in the definition of a 3D world has been taken into account the conclusion that relativity of simultaneity is impossible in a 3D world is inescapable. In the next section we will see that length contraction and the twin paradox are manifestations of the reality of the 4D worldtubes of the objects involved in these effects. But now we will first show how the conclusion that conventionality of simultaneity is impossible in a 3D world could have been reached in 1906.

In the section "Definition of Simultaneity" of his 1905 paper Einstein discussed the introduction of a common time at two distant points *A* and *B*: "We have not defined a common 'time' for *A* and *B*, for the latter cannot be defined at all unless we establish *by definition* that the 'time' required by light to travel from *A* to *B* equals the 'time' it requires to travel from *B* to *A*" [37, p. 40]. This conclusion is a result of a deep analysis and shows that Einstein had rediscovered, after Poincaré,[9] the unavoidable conventionality in determining the one-way velocity of light and the simultaneity of distant events. However, that conclusion raises the obvious question "How can the one-way velocity of light be a matter of definition (convention), whereas it appears to be self-evident that *in reality* the back and forth velocities of light are either the same or not the same?" Had Einstein pursued further his analysis he would have most probably arrived at the conclusion that the impossibility to determine the one-way velocity of light had a profound reason – reality is a 4D world in which light (and anything else) does not travel at all since the whole history of a light signal is entirely realized in the (forever given) light signal's world line.

Let us outline the way that analysis could have been performed. Undoubtedly, Einstein had realized the vicious circle when one tries to establish the simultaneity of distant events, that is, to synchronize distant clocks, with the help of light signals – to synchronize two clocks at different locations the one-way velocity of light between them should be known, but to determine the one-way velocity of light the two clocks should be synchronized beforehand. A synchronization by a slow transport of a third clock also leads to the same vicious circle. To see the vicious circle here, one should keep in mind that the third clock's velocity cannot be assumed to be zero; otherwise, it would obviously not be able to reach the clock it is trying to synchronize. No matter how small, the velocity of the third clock is different from zero, which means that the time dilation effect (no matter how small) should be taken into account. To calculate that effect the one-way velocity of the third clock should be known, but in order to determine it the two clocks, which the third clock tries to synchronize, should be synchronized in advance.

Thus, the vicious circle in determining the one-way velocity of light and therefore the simultaneity of distant events is unavoidable. Then the natural question is what message that vicious circle conveys. The deep meaning of the message becomes evident when the impossibility to determine objectively which distant events are simultaneous is analyzed in terms of what exists. Assume that reality is a 3D world (everything that exists *simultaneously* at the present moment). As simultaneity of distant events is a matter of convention it follows that the 3D world is also a matter of convention. But what exists, no matter that it is unobservable at the moment it exists, cannot be a matter of convention. Thus, the only conclusion which one can draw from here is that conventionality of simultaneity is impossible in a 3D world[10]

---

[9] In 1898 Poincaré first realized that it was a postulate that the light "velocity is the same in all directions" and that "[t]his postulate could never be verified directly by experiment" [38, p. 220]. Poincaré also arrived at the conclusion that whether two events are simultaneous is a matter of convention [38, p. 222].

[10] Of course, the first reaction of anyone who arrives at this conclusion would most probably be to question the validity of the conventionality thesis. However, no matter how many times the

(and that conclusion could have been reached before the 4D formulation of special relativity was given in 1908).

Had Einstein arrived at the conclusion that simultaneity of distant events would not be established *"by definition"* if reality were a 3D world he would have faced quite a challenge. It is impossible to guess how much time it would have taken him to decode the message conveyed by the vicious circle and to realize that conventionality of simultaneity implies a 4D world in which space and time are united into an inseparable entity. After the publication of his 1905 papers his intellectual power had been concentrated on making the description of gravity consistent with special relativity. Thus, it appears that the development of general relativity had prevented him from dealing with the question of the physical meaning of our freedom to define which events in the world are simultaneous. What confirms that Einstein had not pursued his analysis of the conventionality of simultaneity further is his initial negative reaction to Minkowski's 4D formulation of special relativity. However, the man who was not afraid to ask fundamental questions and to seek their answers, the man who discovered the unimaginable at that time link between gravity and geometry, would have definitely been able to decode the message hidden in the vicious circle involved in any attempt to determine which events are simultaneous, if he had had the time. Most probably, Einstein would have realized that our freedom to choose our 3D world (the set of simultaneous events at the moment "now") implies that we have *from where to choose*. Then arriving at the idea that there is a link between conventionality of simultaneity and dimensionality of the world would not have been so unthinkable.

Once it is realized that reality is a 4D world conventionality of simultaneity turns out to be trivial – as all events of space-time are equally existent it is really our choice which events constituting a 3D hypersurface (lying outside of the light cone at a given event) will be regarded as simultaneous. In such a 4D world in which there are only world lines of particles and light signals the velocity of light is just a *description* of light world lines in terms of our 3D language since in reality (in space-time) light does not travel at all. If the world were 3D and light were really propagating, its one-way velocity could not be conventional because Nature would "know" what is the magnitude of that velocity.

The realization that conventionality of simultaneity is another manifestation of the 4D of the world is in fact an argument against any attempt to relativize existence since such a view would preserve the 3D of the world.

7.2.3 The Existence of Accelerated Observers in Special Relativity is Impossible in a Three-dimensional World

In the early years of special and general relativity there had been some confusion about accelerated motion – for some time, due to Einstein's equivalence principle, there had been a tendency to think that it is general relativity which deals with

---

analysis leading to the vicious circle would be repeated its existence would be confirmed and the conventionality thesis would follow.

accelerated motion. Gradually, however, the issue has been settled and in the early seventies Misner, Thorne, and Wheeler even entitled one of the sections of their *Gravitation* "Accelerated observers can be analyzed using special relativity" [2, p. 163]. Now the situation is completely clear – general relativity describes curved space-times, whereas special relativity is concerned with a flat space-time. Absolutely accelerated observers [17, Ch. 8], which are represented by non-geodesic world lines, exist in both curved and flat space-times.

So, accelerated observers are described by special relativity, but that would not be possible if reality were a 3D world. To see why this is so assume that what exists is only the present, i.e. the 3D world at the moment "now," and consider an accelerated observer whose worldtube is depicted in Fig. 7.3. Due to the fact that the worldtube of the accelerated observer is curved the presents that correspond to the different moments of his proper time are not parallel and intersect one another as shown in Fig. 7.3. Consider the accelerated observer's presents that correspond to the events M and N. At event M the space-time region between the two presents denoted by the question mark "?" is past for the accelerated observer (including part of his present at the later event N). However, the same space-time region lies in the future of the accelerated observer at the later event N (including part of his present at the earlier event M). Obviously, this would be impossible if the world were 3D. Thus, the existence of accelerated observers in special relativity is another manifestation of the 4D of the world and is therefore an argument against the view that existence should be relativized because this view regards the world as 3D.

Almost certainly a physicist would object to the above argument by pointing out that there are constrains on the size of an accelerating frame in relativity [2, p. 168]. These constrains result from the fact that a coordinate system associated with the accelerated observer cannot be extended to the left in Fig. 7.3 beyond event H. The first objection against using a global coordinate system in the case of an accelerated observer was mentioned above – coordinate time makes no sense beyond event H because what is past time at event M (in the region "?") is future time at the later event N. A second objection is the fact that the coordinates of events located in the space-time region between the past and future light cones, which contains the

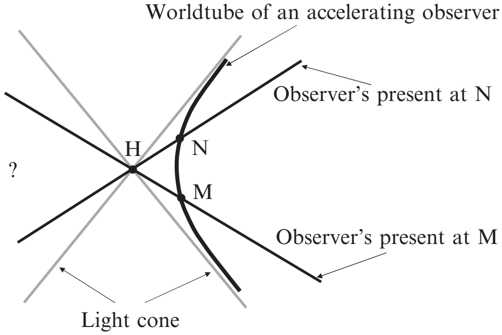

Fig. 7.3 At any moment of his proper time an accelerated observer has a set of simultaneous events (his present at that moment). Unlike an inertial observer's presents the presents of the accelerated observer intersect one another.

region "?", cannot be determined by sending and receiving light signals [2, p. 168]. The event H (in fact, a 2D surface) acts as a horizon for the accelerated observer – no signals can be received from and sent to the space-time region lying to the left of that event.

All this is correct, but the above two objections are concerned *only* with the *description* of the events of space-time and have nothing to do with the question of whether or not they exist. That the impossibility for an accelerated observer to communicate with events located to the left of event H (between the past and future light cone at H) has no effect on the existence of these events is best demonstrated by the fact that all comoving inertial observers at the different events of the worldtube of the accelerated observer can communicate with those events. And if they exist for the comoving inertial observers they should exist for the accelerated observer as well.

It should be noted here that the relativistic division of events into past, present, and future (Fig. 7.1), which defines the notion of a light cone, reveals only the causal structure of space-time and has no relation to the existence of the space-time events.[11] As we have seen above if one assumes that the events located outside of the light cone at an event exist it is easily shown that all space-time events exist.

7.3 Physical Objects are Four-dimensional Worldtubes

In this section we will be concerned with the dimensionality of the objects involved in the relativistic effects. This issue was dealt with in more detail in [17, Ch. 5]. Here I will discuss the relativistic length contraction by addressing Ohanian's objections [33] to the account given in [17] and the twin paradox by examining another version of it that rules out the acceleration as a cause for the time difference observed by the twins when they meet.

7.3.1 Length Contraction Would be Impossible if the Contracting Meter Stick were a Three-dimensional Object

It is unfortunate that the physics papers and books on relativity do not discuss the issue of the dimensionality of the objects subjected to relativistic effects. As discussed in the Introduction the most probable explanation might be that that issue is regarded by physicists as philosophical. However, I wonder how many physicists, if directly asked, would subscribe to the view that the dimensionality of physical objects (and the world) is a philosophical question.

If the dimensionality issue is explicitly addressed, not only the description, but a full explanation of these effects can be achieved. Take as an example length con-

[11] One can think of a 4D world with no causal structure.

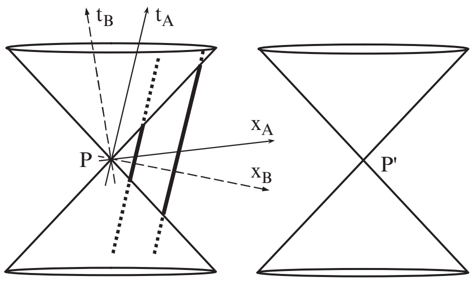

Fig. 7.4 A meter stick, whose end points' world lines are represented by the thick lines, is at rest with respect to observer A. The instantaneous space of observer B at event P intersects the worldtube of the meter stick in a 3D cross section, whose length is shorter than the cross section "cut off" by the instantaneous space of observer A. If one assumes that only the part of the meter stick's worldtube which lies in the space-time region outside of the light cone at P is real, then by the same argument the worldtube's parts that are in the past and future light cone at P are also real since they lie outside of a second light cone at P'.

traction. Consider a meter stick at rest with respect to an observer A (Fig. 7.4). A second observer B moving relative to A finds that the length of the meter stick is relativistically contracted. The natural question is "What is the physical meaning of the relativistic length contraction?" Physicists usually avoid addressing it by pointing out that what a physicist should be concerned with is its derivation through the Lorentz transformations. Although many physicists believe that the ultimate goal of physics is *understanding* of the real world (and do not share this purely descriptive role of physics), it is worth noting here that the very derivation of this effect from the Lorentz transformations involves an overlooked subtlety[12] [17, pp. 94–97] which clearly raises the question of the physical meaning of length contraction. This subtlety implies that the meter stick is not a 3D object, but a 4D worldtube – the 3D meter stick equally existing at all moments of its history. As a result, while measur-

[12] The subtlety involves an unusual use of the Lorentz transformations in the derivation of length contraction. Instead of performing what appears to be the correct transformation $A \Rightarrow B$ (which expresses the unknown in frame B coordinates of the end points of the meter stick as a function of their *known* in frame A coordinates) this effect is derived through the transformation $B \Rightarrow A$, i.e. by expressing the known coordinates of the meter stick's end points through their unknown coordinates. The requirement that the end points of the meter stick should be measured simultaneously in B cannot be used as a justification for the transformation $B \Rightarrow A$ since the transformation $A \Rightarrow B$ also ensures the simultaneity of the events of the measurement in B [17, p. 95]. (For comparison consider a process that takes place at a point in A and has a duration t_A in A. In order to determine its relativistically dilated duration t_B in B, observer B performs the "correct" Lorentz transformation $A \Rightarrow B$, not $B \Rightarrow A$.) The Lorentz transformation $B \Rightarrow A$ gives the correct length contraction because it is only this transformation that relates the end points of two 3D cross section of the worldtube of the meter stick; the transformation $A \Rightarrow B$ does not link pairs of events that constitute cross sections of the meter stick's worldtube. This fact implies that the worldtube of the meter stick is a real 4D object; if this were not the case than the transformation $A \Rightarrow B$ would be used which would lead to internal inconsistencies in special relativity and ultimately to contradictions with experiments that confirmed the relativistic length contraction (e.g. the muon experiment, which tested both time dilation and length contraction [39]).

ing the *same* meter stick the two observers measure *two* different 3D objects, which are different 3D cross sections of the meter stick's worldtube. This is shown in Fig. 7.4. The instantaneous spaces of A and B "cut off" *different* 3D cross sections from the worldtube of the meter stick and A's cross section, representing the proper length of the meter stick, turns out to be the longest of all possible cross sections. Thus, the realization that what we perceive as a 3D meter stick is in fact a 4D worldtube provides a 100% explanation[13] of the relativistic length contraction – the instantaneous space of B intersects the meter stick's worldtube in a 3D cross section which is different from and shorter than the cross section of A.

A possible objection to this "a 100% explanation" of the relativistic length contraction is that one should not take the space-time time diagram shown in Fig. 7.4 too seriously. First, it is the analysis of the derivation of this effect which demonstrates that what is depicted in Fig. 7.4 adequately represents the dimensionality of the meter stick. Second, the conclusion that A and B measure two different 3D objects follows directly from relativity of simultaneity when it is taken into account that the meter stick as an extended body is defined as the set of its "parts" which exist *simultaneously* at a given moment. Since the observers A and B are in relative motion they have different sets of simultaneous events and therefore *different* 3D meter sticks. It is evident from here that the worldtube of the meter stick must be real in order that A and B consider different 3D cross sections of it as their 3D meter sticks. Otherwise, if the meter stick were what everyone is tempted to assume as self-evident – a single 3D object – (just A's meter stick), no length contraction would be possible because that single 3D object would constitute a single set of simultaneous events, which would be common to all observers in relative motion in contradiction with special relativity.

Here, one could again raise the objection that the conclusion of the reality of the meter stick's worldtube is based on relativity of simultaneity, whereas one is free to choose any hypersurface in the space-time region outside of the light cone at event P (Fig. 7.4). However, like in the general case of the dimensionality of the world discussed in section 7.2 this objection is, in fact, directed against the view that the meter stick is a 3D object since a 3D object is defined in terms of simultaneity.[14] The freedom to choose any hypersurface which lies outside of the light cone, i.e. the freedom to choose any cross section of the meter stick's worldtube outside of the light cone, means that the part of the worldtube (depicted with thick lines) located outside of the light cone must be real in order to have *from where to choose*. The rest of the meter stick's worldtube (represented by the dashed lines) which lies in

[13] I call it "a 100% explanation" since it is not based on any other explicit or implicit assumptions. If the worldtube of the meter stick is real, that is all – it *completely* explains the physical meaning of length contraction. The "if" is convincingly removed as we will see below.

[14] The fact that no intrinsic (therefore frame-independent) feature of space-time can be associated with the notion "simultaneity in space-time," i.e. with the notion "distant present events," demonstrates that the concept of a 3D object (which is defined in terms of simultaneity) has no place in relativity. This conclusion appears to be in such an obvious contradiction with our common sense view on physical objects that the status of 3D objects in relativity has been avoided so far.

the past and the future light cone at P must also be real since it is outside of another light cone associated with event P' (Fig. 7.4).

Thus, the conclusion that the worldtube of the meter stick is a real 4D object is inescapable. To realize this even better let us ask the question "What is the dimensionality of the meter stick *itself* (not what we see or measure[15])?" Obviously, there are only three possibilities for the dimensionality of the meter stick: (i) a 3D object, (ii) part of the worldtube of the meter stick which lies outside of the light cone at event P (Fig. 7.4), and (iii) the entire worldtube of the meter stick. As we have seen above (i) contradicts relativity, whereas (ii) leads to (iii):

- If the meter stick is a 3D object (at rest with respect to observer A) relativity of simultaneity is impossible, since it follows from relativity of simultaneity that A and B measure two different 3D objects, which is only possible if the meter stick's worldtube exists.
- If the meter stick is the part of its worldtube which lies outside of the light cone at P then it follows that the entire worldtube of the meter stick is real, since the parts of the worldtube that lie in the past and future light cone at P (and therefore are regarded as nonexistent for an observer at P) lie outside of another light cone at P' and therefore are regarded as existent for an observer at P' (Fig. 7.4).

As we have seen not only does the assumption that the worldtube of the meter stick is real provide a natural and complete explanation of the relativistic length contraction, but also that explanation is the only one that is consistent with relativity itself. A striking feature of this explanation is that it is not dynamical since it turns out to be a manifestation of the 4D of the world and in particular a manifestation of the fact that the meter stick itself is a 4D worldtube, not a 3D object. However, the assumption of the reality of the worldtube of the meter stick is so counter-intuitive that after the original Lorentz–FitzGerald explanation of the length contraction effect, which involved a deformation of the meter stick caused by forces acting on its atoms, there have always been attempts to provide a dynamical explanation of the relativistic length contraction which should account for the deformation of the meter stick (see, for example [33, 40, 41]). Such an explanation, however, is in an insurmountable contradiction with relativity of simultaneity since it presupposes that two observers in relative motion have a common set of simultaneous events – the observers measure the length of the *same* 3D object, i.e. the same set of simultaneous events, which constitute the meter stick.

Another argument which demonstrates the failure of any dynamical explanation of the relativistic length contraction is the fact that such an explanation cannot ac-

---

[15] What the observers A and B *see* when they meet momentarily at P (Fig. 7.4) is the *same* 3D cross section of the meter stick's worldtube of the *same* length, but it does not constitute a 3D object since that image is the intersection of the past light cone with the worldtube of the meter stick, whereas a 3D object is the collection of all "parts" of the object that exist *simultaneously* at a given moment of an observer's time [32]. A thought experiment [17, p. 137] involving instantaneous measurements demonstrates that two observers in relative motion do measure two different 3D objects. Thus, not only the theoretical prediction of relativistic length contraction would not be possible but any experiments that test this effect would be also impossible if the meter stick's worldtube were not real.

count for the contraction of space itself where there are no atoms and forces acting between them [17, pp. 135–136]. However, Ohanian [33] disagrees with this argument[16]: "If the length contraction represents anything physical, it must be a contraction of physical bodies, not a 'contraction of space.' " I do not think this objection can be defended in the framework of relativity. Nowhere in special relativity there is any requirement that the Lorentz transformations must be applied to physical bodies only. The distance between two points of space contracts[17] relativistically as well. Otherwise, if only physical bodies contracted, one would arrive at the following paradoxical result (which can be found in some explanations of the geometry on a rotating disk). Consider the distance L_{MN} between two points M and N in space, say the distance between two objects at rest with respect to an observer A. Using a meter stick A determines that $L_{MN} = 10$ m. A second observer B moving relative to A applies the Lorentz transformations only to the physical meter stick and concludes that due to its relativistic contraction more meter sticks will fit between the points M and N and as a result the distance L_{MN} will be greater that 10 m, i.e. it will be relativistically dilated for B, not contracted. One can arrive at more paradoxical results if the assumption that only physical bodies contract is carefully analyzed.

7.3.2 The Twin Paradox Would be Impossible if the Twins Were Three-dimensional Bodies

The best way to prove the statement in the section title is to accept for a moment the common view on the dimensionality of physical bodies and assume that each of the twins A and B, represented by their worldtubes in Fig. 7.5, is a 3D body that exists only at the moment "now" of the proper time of the twin. On this view the twins' worldtubes are not real 4D objects; these are nothing more than graphical representations of the twins' histories, which do not have counterparts in the external world. On the 3D view time flows objectively which means that any discrepancy in the readings of the twins' clocks results from a different "rate" of the

[16] Ohanian also writes: "We could, in principle, do physics in the manner of Lorentz, in one single coordinate system (what Lorentz called the ether frame)" [33]. This could not be done since relativity of simultaneity and reciprocity of length contraction and time dilation require two inertial frames. Otherwise one could not talk about proper length (in the rest frame) and contracted length (in a frame moving relative to the rest frame) and about proper time and dilated time (which again require two frames). Lorentz himself admitted the failure of his attempt to use just one frame (the ether frame) in which the coordinates x, y, z, and t were the true coordinates, whereas the quantities x', y', z', and t' were nothing more than mathematical quantities [42]:

> The chief cause of my failure was my clinging to the idea that the variable t only can be considered as the true time and that my local time t' must be regarded as no more than an auxiliary mathematical quantity. In Einstein's theory, on the contrary, t' plays the same part as t; if we want to describe phenomena in terms of x', y', z', t' we must work with these variables exactly as we could do with x, y, z, t.

[17] And that distance can be measured by light signals, not only by physical bodies (meter sticks).

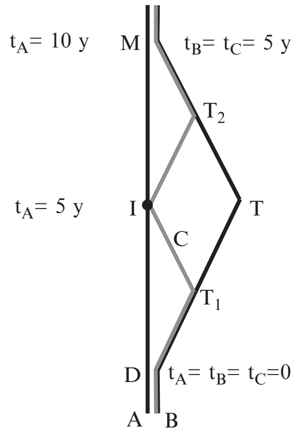

Fig. 7.5 Twins *A* and *B*, whose worldtubes are represented by the black lines, separate at event *D* and meet at event *M*. A third twin *C*, whose worldtube is depicted by the grey line, also departs at *D* and initially moves with *B* but at event T_1 turns back and after reaching *A* at event *I* departs again, intercepts *B* at event T_2, turns back and together with *B* meets *A* at event *M*.

twins' times. When *A* and *B* meet at event *M* they will determine that twin *B* is, say, 5 years younger than his brother. As the twins exist as 3D bodies only at *M* the only explanation of the 5-year difference in their times is an objective slowing down of *B*'s time. The space-time explanation that *B*'s worldtube is shorter[18] than *A*'s worldtube cannot be used since we started with the assumption that the twins exist as 3D bodies, not worldtubes.

The only cause for the slowing down of twin *B*'s time could be the acceleration to which he is subjected four times during his journey – acceleration at *D* when he departs, deceleration at *T* when he stops to turn back, acceleration at *T* on his way back to twin *A*, and final deceleration when he stops to meet his brother at *M*. However, it has been determined that acceleration does not cause the time difference in *A*'s and *B*'s clocks' readings at *M* by both (i) the experiments (see [14, p. 83]) which tested the "clock hypothesis" [2, p. 164], [11, p. 52], [20, p. 33] according to which the rate of an ideal clock is not affected by its acceleration, and (ii) the three-clock (or three-twin) version of the twin paradox (see, for instance, [43]). A third argument also involves a third twin *C* (Fig. 7.5) who, however, is not inertial since he accelerates eight times – at events *D* and *M* twin *C* has the same regime of acceleration as twin *B* at these events, at events T_1 and T_2 his acceleration regime is as the acceleration of *B* at *T*, and at event *I* twin *C*'s acceleration is a mirror image of *B*'s acceleration at *T*. Despite that *C* experiences more instances of acceleration, there will be again a 5-year difference in *A*'s and *C*'s times when he arrives at *M*. This shows that the acceleration does not cause any slowing down of time. Therefore, in the case of the standard version of the twin paradox *B*'s acceleration cannot be the cause for the slowing down of his time, which means that his time flows at the same "rate" as twin *A*'s time. The conclusion is that at event *M* the twins will be of the same age. Thus, the twin paradox would not be possible if the twins were 3D bodies which existed only at the present moments of the twins.

[18] In Fig. 7.5 twin *B*'s worldtube is longer but this is due to the fact that a situation in the pseudo-Euclidean Minkowski space-time is represented on the Euclidean surface of the paper.

The same conclusion that A and B will be of equal age at M, if they are 3D bodies existing only at M, also follows from the fact that A and B measure *proper* times. When the twins meet at M they compare their proper times [17, pp. 144–145], but the proper time does not change relativistically. Therefore, if the twins existed only at the event M as 3D bodies the objective flow of A's and B's times (possible only in a 3D world) would be the same which means that the twins would be of the same age at M. Thus, the twin paradox is impossible as a theoretical and an experimental result if it is assumed that the twins' worldtubes are not real and the twins are the observable 3D bodies that exist only at the constantly changing present moments of the twins' times. This results also rules out any attempt to relativize the existence of physical bodies since such a relativization of existence preserves the 3D of the world and the physical bodies.

The only noncontradictory *explanation*[19] of the twin paradox can be given by acknowledging the existence of the twins' worldtubes. Then this effect is simply the triangle inequality in the pseudo-Euclidean Minkowski space-time. Twin B is younger than A at M since his worldtube between the events D and M is shorter than A's worldtube between the same events. This is also a 100% explanation. It makes it evident why the acceleration is not the cause of the time difference between A's and B's clock readings at M – in space-time the acceleration of a body is represented by a curvature of the body's worldtube, but a curvature does not change the length of a worldtube (i.e. a body's proper time). As seen in Fig. 7.5 the lengths of the worldtubes of twins B and C between events D and M are the same (the segment T_1I of C's worldtube is equal TT_2 of B's worldtube and IT_2 of C's worldtube is equal to T_1T of B's worldtube). That is why B and C are of the same age at M.

Conclusions

The main aim of the paper is to demonstrate that the kinematical relativistic effects are manifestations of the 4D of the world. As such these effects would be impossible if the world were 3D. This was shown in the cases of relativity of simultaneity, conventionality of simultaneity, accelerated observers in special relativity, length contraction, and the twin paradox. Therefore, the concept of a 3D world contradicts not only special relativity (as a theory) but more importantly the experimental evidence which supports it.

I would like specifically to stress that no appeal to quantum mechanics or any future theories (e.g. quantum gravity) can change the fact of the contradiction of the 3D view with the *experiments* which confirmed the kinematical consequences of special relativity.

[19] This relativistic effect was initially *described* in 3D language. But when the question of its physical meaning is raised it becomes evident that the twin paradox is a manifestation of the reality of the twins' worldtubes and the 4D of the world.

We are approaching the 100th anniversary of Minkowski's talk on space-time, but the essence of the new world view he advocated has turned out to be difficult to accept. That is why let me conclude with a quote from Eddington [44]:

> However successful the theory of a four dimensional world may be, it is difficult to ignore a voice inside us which whispers: "At the back of your mind, you know that a fourth dimension is all nonsense." I fancy that that voice must often have had a busy time in the past history of physics. What nonsense to say that this solid table on which I am writing is a collection of electrons moving with prodigious speeds in empty spaces, which relatively to electronic dimensions are as wide as the spaces between the planets in the solar system! What nonsense to say that the thin air is trying to crush my body with a load of 14 lbs. to the square inch! What nonsense that the star cluster which I see through the telescope obviously there now, is a glimpse into a past age 50,000 years ago! Let us not be beguiled by this voice. It is discredited.

Acknowledgments Most of the ideas in this paper were presented at a 2-day workshop of the Canadian Strategic Knowledge Cluster TaU, Time and Universe (sponsored by the Social Sciences and Humanities Research Council of Canada) and its international associates, which was held at Concordia University, Montreal on June 12–13, 2006. I would like to thank the colleagues who attended the workshop and took part in the discussion after my talk.

References

1. H. Minkowski: Space and Time. In: H.A. Lorentz, A. Einstein, H. Minkowski, H. Weyl (eds): *The Principle of Relativity: A Collection of Original Memoirs on the Special and General Theory of Relativity* (Dover, New York 1952) pp 75–91
2. J.L. Synge: *Relativity: The Special Theory*, 2nd edn (North-Holland, Amsterdam 1965)
3. J.L. Synge: *Relativity: The General Theory* (North-Holland, Amsterdam 1960)
4. E.F. Taylor, J.A. Wheeler: *Spacetime Physics*, 2nd edn (W.H. Freeman, New York 1992)
2. C.W. Misner, K.S. Thorne, J.A. Wheeler: *Gravitation* (W.H. Freeman, New York 1973)
6. E.F. Taylor, J.A. Wheeler: *Exploring Black Holes: Introduction to General Relativity* (Addison-Wesley Longman, New York 2000)
7. B. Schutz: *Gravity from the Ground Up: An Introductory Guide to Gravity and General Relativity* (Cambridge University Press, Cambridge 2003)
8. J.B. Hartle: *Gravity: An Introduction to Einstein's General Relativity* (Addison-Wesley, San Francisco 2003)
9. S.M. Carroll: *Spacetime and Geometry: An Introduction to General Relativity* (Addison-Wesley, San Francisco 2004)
10. W. Rindler: *Relativity: Special, General and Cosmological* (Oxford University Press, Oxford New York 2001) p 91
11. G.L. Naber: *The Geometry of Minkowski Spacetime: An Introduction to the Mathematics of the Special Theory of Relativity* (Springer, Berlin, Heidelberg, New York 1992)
12. A. Das: *The Special Theory of Relativity: A Mathematical Exposition* (Springer, Berlin, Heidelberg, New York 1993)
13. E.G. Peter Rowe: *Geometrical Physics in Minkowski Spacetime* (Springer, Berlin, Heidelberg, New York 2001) p 3
14. R.A. Mould: *Basic Relativity* (Springer, Berlin, Heidelberg, New York 1994)
15. D.-E. Liebscher: *The Geometry of Time* (Wiley-VCH, Weinheim 2005) p 5
16. W.S.C. Williams: *Introducing Special Relativity* (Taylor & Francis, London 2002)

17. V. Petkov: *Relativity and the Nature of Spacetime* (Springer, Berlin, Heidelberg, New York 2005)
18. N. Straumann: *General Relativity With Applications to Astrophysics* (Springer, Berlin, Heidelberg, New York 2004) p 46
19. M.P. Hobson, G. Efstathiou, A.N. Lasenby: *General Relativity: An Introduction for Physicists* (Cambridge University Press, Cambridge 2006) pp 125, 357
20. R. d'Inverno: *Introducing Einstein's Relativity* (Clarendon Press, Oxford 1992) p 17
21. J.J. Callahan: *The Geometry of Spacetime: An Introduction to Special and General Relativity* (Springer, Berlin, Heidelberg, New York 2000) p 10
22. G. Barton: *Introduction to the Relativity Principle* (Wiley, New York 1999) p 10
23. T.-P. Cheng: *Relativity, Gravitation and Cosmology* (Oxford University Press, Oxford, New York 2005) p 5
24. C. W. Rietdijk: Philosophy of Science **33**, 341 (1966)
25. H. Putnam: Journal of Philosophy **64**, 240 (1967)
26. N. Maxwell: Philosophy of Science **52**, 23 (1985)
27. A. Einstein: *Relativity: The Special and General Theory* (Crown Publishers, New York 1961) p 150
28. H. Weyl: *Philosophy of Mathematics and Natural Science* (Princeton University Press, Princeton 1949) p 116
29. R. Weingard: British Journal for the Philosophy of Science **23**, 119 (1972)
30. H. Stein: Journal of Philosophy **65**, 5 (1968)
31. H. Stein: Philosophy of Science **58**, 147 (1991)
32. V. Petkov: Is There an Alternative to the Block Universe View? In: D. Dieks (ed.): *The Ontology of Spacetime* (Elsevier, Amsterdam 2006) pp 207–228
33. H.C. Ohanian: The Real World and Spacetime. In: this volume
34. R.M. Wald: American Journal of Physics **74**, 471 (2006)
35. V. Petkov: British Journal for the Philosophy of Science **40**, 69 (1989)
36. J. Winnie: Philosophy of Science **37**, 81 (1970); Philosophy of Science **37**, 223 (1970)
37. A. Einstein: On the electrodynamics of moving bodies. In: H.A. Lorentz, A. Einstein, H. Minkowski, H. Weyl (eds): *The Principle of Relativity: A Collection of Original Memoirs on the Special and General Theory of Relativity* (Dover, New York 1952) pp 37–65
38. H. Poincaré: *The Value of Science* (The Modern Library, New York 2001) pp 210–222
39. G.F.R. Ellis, R.M. Williams: *Flat and Curved Space Times* (Oxford University Press, Oxford 1988) p 104
40. J.S. Bell: How to Teach Special Relativity. In: J.S. Bell (ed.): *Speakable and Unspeakable in Quantum Mechanics* (Cambridge University Press, Cambridge 1987) pp 67–80
41. H.R. Brown: *Physical Relativity: Spacetime Structure from a Dynamical Perspective* (Clarendon Press, Oxford 2005)
42. H.A. Lorentz: *The Theory of Electrons and Its Applications to the Phenomena of Light and Radiant Heat*, 2nd edn (Dover, Mineola, New York 2003) p 321
43. P. Kroes: Philosophy of Science **50**, 159 (1983)
44. A.S. Eddington: *Space, Time and Gravitation: An Outline of the General Relativity Theory* (Cambridge University Press, Cambridge 1920) p 56

Chapter 8
Canonical Relativity and the Dimensionality of the World

Martin Bojowald

Different aspects of relativity, mainly in a canonical formulation, relevant for the question "Is spacetime nothing more than a mathematical space (which describes the evolution in time of the ordinary three-dimensional world) or is it a mathematical model of a real 4D world with time entirely given as the fourth dimension?" are presented. The availability as well as clarity of the arguments depends on which framework is being used, for which currently special relativity, general relativity and some schemes of quantum gravity are available. Canonical gravity provides means to analyze the field equations as well as observable quantities, the latter even in coordinate independent form. This allows a unique perspective on the question of dimensionality since the space-time manifold does not play a prominent role. After reintroducing a Minkowski background into the formalism, one can see how distinguished coordinates of special relativity arise, where also the nature of time is different from that in the general perspective. Just as it is of advantage to extend special to general relativity, general relativity itself has to be extended to some theory of quantum gravity. This suggests that a final answer has to await a thorough formulation and understanding of a fundamental theory of space-time. Nevertheless, we argue that current insights into quantum gravity do not change the picture of the role of time obtained from general relativity.

8.1 Introduction

When faced by the question of whether the world is 3- or 4D, the quick answer by a modern-day physicist will most likely be "four." This is indeed what relativity tells us formally where space and time are essentially interchangeable: Lorentz transformations, or their physical manifestations of Lorentz contraction and time dilatation, show that space and time not only play similar roles but can even be transformed into each other. Just as we can rotate a body in three dimensions to observe all its extensions, thereby transforming, e.g. its height in width, we can boost an

V. Petkov (ed.), *Relativity and the Dimensionality of the World*, 137–152.

object[1] so as to, at least to a certain extent, replace spacelike by timelike extension and vice versa. The qualification "to a certain extent" is necessary because even in relativity space and time are not quite the same but distinguished by the signature of the space-time metric. By itself, this difference in signature is not sufficient reason to deny time the same ontological status as space.

There are, however, differences between the usual treatment of space and time in physics going beyond relativity, although they are usually presupposed in relativistic discussions. In order to answer the question of the dimensionality of the world from the viewpoint of relativity, such hidden assumptions have to be uncovered and analyzed, or avoided altogether. Some of these issues lie at the forefront of current physics and still await explanation. For instance, while we can, and have to, limit objects to finite spatial extensions, we have no means to limit their time extensions safe for transformations such as particle decay or other reactions. Even though objects may change in time, they never cease to exist completely. There seems to be a simple reason for that: conservation laws. We simply cannot limit an object's extension in time because, e.g. its energy must be conserved. Thus, the object could be transformed into something else of the same energy but not removed completely. Such laws are derived as consequences of symmetries which first give local conservation laws in terms of currents. Going from local to global conservation laws, as they are required for an explanation of the persistence of objects in time, is a further step and requires additional assumptions. As the following more detailed discussion shows, conservation laws cannot be used to explain the difference of spatial and time-like extensions, for the derivation itself distinguishes space from time.

One starts with a local equation such as[2] $\nabla^a T_{ab} = 0$ for the energy-momentum tensor T_{ab}. If the space-time metric is sufficiently symmetric and allows a Killing vector field ξ^a satisfying $\mathcal{L}_\xi g_{ab} = \nabla_a \xi_b + \nabla_b \xi_a = 0$, the current $j_a = T_{ab}\xi^b$ is conserved: $\nabla^a j_a = 0$. At this stage, the only difference between space and time enters through the signature of the metric. A global conservation law is then derived by integrating the local conservation equation over a space-time region bounded by two spatial surfaces Σ_1 and Σ_2 and some boundary B which could be at infinity (Fig. 8.1). Stokes theorem then shows that the quantity $\int j_a \mathrm{d}S^a$ is the same on Σ_1 and Σ_2 and thus conserved, *provided that all fields fall off sufficiently rapidly toward the boundary B*. Thus, one already has to assume physical objects to be of finite spatial extent before obtaining a global conservation law, while there is no

---

[1] By "object" we will mean a physical system defined by a set of observable properties such that it can be recognized at different occurrences in space and in time. Objects will not be idealized to be point-like or event-like in order to remain unbiased toward the question of dimensionality. Thus, nonvanishing extensions of objects in space, as well as time are allowed in order to take into account the necessary unsharpness of measurements needed to verify the defining properties.

[2] We follow the abstract index notation common in general relativity (see, e.g. [1]). Indices $a, b, \ldots$ refer to space-time while indices $i, j, \ldots$ used later refer only to space. Repeated indices occurring once raised and once lowered are summed over the corresponding range $0,1,2,3$ for space-time indices and $1,2,3$ for space indices. The covariant derivative compatible with a given space-time metric g_{ab} is denoted by ∇_a which for Minkowski space reduces to the partial derivatives ∂_a in Cartesian coordinates.

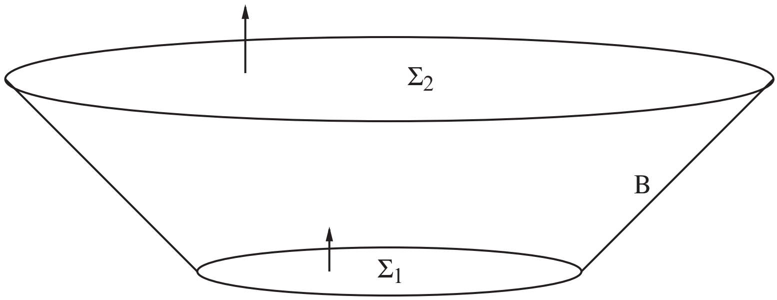

Fig. 8.1 Causal diagram of space-time region integrated over to derive global conservation laws.

such restriction for the timelike extension.[3] If fields do not vanish at B, one interprets $\int_B j_a \mathrm{d}S^a$ as the flux into or out of the spatial region evolving from Σ_1 to Σ_2. However, this different interpretation of $\int j_a \mathrm{d}S^a$ as conserved quantity on Σ_1 and Σ_2 and as flux on B treats space and time differently, corresponding to the nonrelativistic decomposition of energy–momentum in energy and momentum. This different treatment is not implied by the theory but put in by interpreting its objects. The issue of a limited spatial extent versus unconstrainable duration of objects thus remains and has to be faced even before coming to conservation laws.

For this reason, it seems to be potentially misleading to consider objects in space-time such as point particles or their world lines to address the dimensionality of the world, for there are already presuppositions about space and time involved. Indeed, from this perspective a world line, or the world-region of an extended object, seems inadequate for a relativistic treatment. It would be more appropriate to use only either space-time events or bounded 4D world-regions of extended objects. This already indicates a possible answer to the question of dimensionality: events[4] are 0D and to be considered as idealizations just as their analogs of point particles. The only option then is to consider bounded world-regions[5] as physical objects, which are 4D.

It is difficult to follow these lines toward a clear-cut argument for the 4D of the world due to our limited understanding of the nature of time. An alternative

[3] In this discussion we understood, as usually, that space-time is Minkowski as in special relativity. The energy conservation argument in our context works better if one considers instead a universe model with compact spatial slices, such as an isotropic model with positive spatial curvature, or a compactification of isotropic models of non-positive curvature. This requires one to go beyond special relativity, as we will do later on for other reasons, and to allow nonzero curvature or nontrivial topology. From our perspective, closed universe models are conceptually preferred because space is already finite without boundary such that non-spacelike boundaries are not needed to derive global conservation laws from local ones.

[4] Space-time events are the quantities for which the philosophical idea of space and time as *principia individuationis* — entities which are themselves not physical but required to individualize physical objects — most clearly arises.

[5] Indeed, to analyze an object by whatever means we not only need to capture it at one time — which is virtually impossible, anyway — but also hold and observe it for some time. Observations thus always refer to some finite extension in time during which we must be able to recognize the

approach is to disregard objects in space-time and rather consider the relativistic physics of space-time itself. For this, we need general relativity which, compared to special relativity, has the added advantage of removing the background structure given by assuming Minkowski space-time. As backgrounds can be misleading, if possible one should consider the more general situation and then see how special situations can be reobtained.

The following sections collect possible ingredients which can be helpful in the context of dimensionality. There are different formulations of general relativity, covariant, and canonical ones, which apparently reflect the possible interpretations of a 4D versus a 3D world: while covariant field equations are given on a space-time manifold, the canonical formulation starts with a slicing of space-time in a family of spatial slices. Canonical fields are then defined on space and evolve in time, suggesting a 3D world with an external time parameter [2]. Nonetheless, the question of dimensionality cannot be answered easily because, for one thing, the formulations are mathematically equivalent. In what follows, we will mainly use canonical relativity so as to see if it indeed points to a 3D world or, as the covariant formulation, a 4D one. Covariant formulations are also best suited to understand the relativistic kinematics and dynamics. After this general exposition we will specialize the formalism to Minkowski space in order to see which freedom is eliminated in special relativity compared to general relativity and how this can change the picture of time. We end with a brief discussion of dynamical consequences of general relativity as well as some comments on quantum aspects.

8.2 Canonical Relativity

The signature of the metric also has implications for the form of relativistic field equations on a given space-time which are hyperbolic rather than elliptic.[6] This means that a reasonable setup for solving these equations is by an initial value problem: for given initial values on space at an initial time one obtains a unique solution. For our purposes, this aspect is not decisive because we could interpret initial values as corresponding to objects placed in space before starting an observation, thus corresponding to a 3D world, or simply as labels to distinguish solutions which

system. A good example can be found in particle physics where too short decay times imply that particles appear rather as resonances without sharp values for all their properties. This is a consequence of uncertainty and thus quantum theory which we will come back to later. Even though common terminology often assigns the object status to an isolated system at a given time, evolving and possibly changing, observations always consider world-regions which could be assigned the object status as well. This nontraditional use of the term "object" is probably discouraged because it is too observer-dependent: it is the observer who decides when to end the experiment and thus determines the time-like extension of the space-time region. However, while the classical world allows us to draw sharp spatial boundaries and thus seems to imply individualized spatial objects, this is no longer possible in a quantum theory. Drawing the line around spatial objects is then a matter of choice, too, comparable to limiting the duration of an observation.

[6] See also the contribution by Carlos Barceló in this volume [3].

themselves play the role of objects of a 4D world.[7] The choice is then just a matter of convenience.

8.2.1 ADM Formulation

For the field equations of the metric itself the situation is more complicated and crucially different (see [4] for the general relativistic initial value problem). Einstein's equations correspond to ten field equations for the ten components of the space-time metric g_{ab}, a symmetric tensor. However, there are only six evolution equations containing time derivatives only of some components while the remaining equations are elliptic and do not contain time derivatives. Although there is no fixed coordinate system, it is meaningful to distinguish between time and space derivatives because, due to the signature of the metric, they are related to vector fields of negative and positive norm squared, respectively. Time evolution is described by an arbitrary timelike vector field t^a while spatial slices of space-time are introduced as level surfaces $\Sigma_t : t = \text{const}$ of a time function t such that $t^a \partial_a t = 1$. The space-time metric g_{ab} induces a spatial metric $h_{ab}(t)$ on each slice Σ_t as well as covariant spatial derivatives. The spatial metric is most easily expressed as

$$h_{ab} = g_{ab} + n_a n_b \tag{8.1}$$

where n_a is the unit future-pointing timelike co-normal to a slice. These are only six independent components because h_{ab} is degenerate from the space-time point of view: $n^a h_{ab} = 0$. These are also precisely the components of the space-time metric whose time derivatives[8] appear in Einstein's equations.

At this point, we may view the equations as describing the evolution of a 3D quantity h_{ab} in an external time parameter t. The remaining four space-time metric components encode the freedom in choosing the time evolution vector field, which can be parameterized as $t^a = Nn^a + N^a$ with components usually called lapse function N and shift vector N^a such that $n_a N^a = 0$. (Fig. 8.2) They are indeed metric components since Eq. (14.15) implies

$$\begin{aligned} g^{ab} &= -n^a n^b + h^{ab} = -\frac{1}{N^2}\left(t^a - N^a\right)\left(t^b - N^b\right) + h^{ab} \\ &= -\frac{1}{N^2} t^a t^b + \frac{1}{N^2}\left(t^a N^b + N^a t^b\right) + h^{ab} - \frac{1}{N^2} N^a N^b . \end{aligned} \tag{8.2}$$

---

[7] In fact, this is clearly brought forward by the canonical formulation where one can either specify states by a phase space given by initial data on the initial surface, or by the so-called covariant phase space consisting of entire solutions to the field equations. In both cases, the phase space is endowed with a symplectic structure, and the formulations are equivalent.

[8] Time derivatives are understood as Lie derivatives with respect to the time evolution vector field t^a.

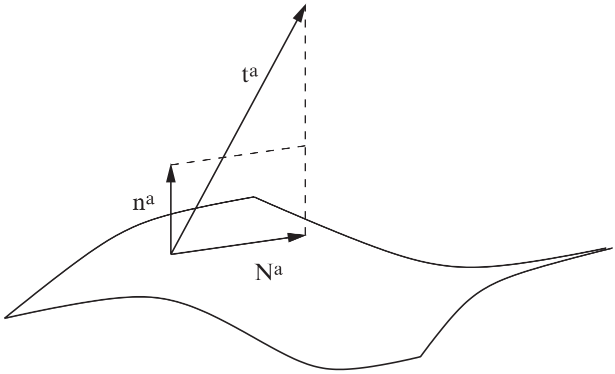

Fig. 8.2 Decomposition of the time evolution vector field t^a into the shift vector N^a and a normal contribution Nn^a.

The time–time component of the inverse space-time metric is thus $-N^{-2}$ while time–space components are $N^{-2}N^a$. These components enter the field equations, too, but they are not dynamical in the sense that they would have evolution equations determining their time derivatives. In addition to the six evolution equations for h_{ab},

$$\dot{\pi}^{ab} = f\left[h_{ab}, \pi^{ab}, N, N^a\right] \tag{8.3}$$

for the momenta $\pi^{ab}[\dot{h}_{cd}, N, N^c]$ conjugate to h_{ab}, there are then four constraint equations

$$C\left[h_{ab}, \pi^{ab}\right] = 0 \quad \text{and} \quad C^a\left[h_{bc}, \pi^{bc}\right] = 0 \tag{8.4}$$

which are of elliptic nature and restrict the values of the dynamical fields at any spatial slice (independently of N and N^a).

A possible interpretation is that there are six fields h_{ab} on space which change in time as governed by the evolution equations, depending on four prescribed but arbitrary auxiliary fields N and N^a. This would be a mixed viewpoint as far as dimensionality is concerned because h_{ab} would look like spatial objects while N and N^a would have to be *prescribed* as functions of space and time but are not evolving in time. The system is thus rather 4D since N and N^a have to be functions on a 4D space and determine the evolution of h_{ab} for which only initial values on space are needed. One can save the 3D interpretation by considering N and N^a as external functions for the evolution system of h_{ab}, but this has the drawback that there would be no predictivity and no uniqueness of solutions in terms of initial values for the dynamical fields, as solutions also depend on choices of N and N^a.

Here, the completely 4D view is much more attractive: we not only have to choose the functions N and N^a but can also supplement the dynamical equations for h_{ab} by evolution equations for space-time *coordinates*. This is indeed possible, for if we choose four functions N and N^a and ask that they play the role of space-time metric components as they enter the canonical Eq. (8.2) we have the transformation laws

$$-\frac{1}{N\left(t,x^i\right)^2} = q^{bc}\left(x'\right)\partial'_b t\partial'_c t \tag{8.5}$$

$$N(t,x^i)^{-2}N^j\left(t,x^i\right) = q^{cd}\left(x'\right)\partial'_c t\partial'_d x^j \tag{8.6}$$

from an arbitrary (inverse) space-time metric q^{ab} to the new functions. These transformations can be interpreted as evolution equations for the space-time coordinates which are then fixed by the choice of N and N^a in terms of coordinates on an initial spatial surface.[9] With this interpretation we obtain, for given initial values, unique solutions to our evolution equations up to changes of coordinates, corresponding to a change in the free functions N and N^a.

The functions N and N^a which must be defined on a 4D manifold thus determine coordinates x^a such that canonical field equations for h_{ab} result. In general, there is no way to split this globally into a time coordinate t and space-coordinates x^i which one would need for a 3D evolution picture of h_{ab}. Thus, a 4D interpretation results.[10] This attractive viewpoint is available only if we take as physical object on which field equations are imposed the entire space-time and not just the metric on spatial slices. We clearly have to consider general relativity which gives field equations for the metric and allows us to perform arbitrary coordinate changes, not just Lorentz transformations. Here, getting rid of the Minkowski background of special relativity, corresponding to a synchronization of rigid clocks and rulers throughout space-time under the assumption of the absence of a gravitational field, is required.

8.2.2 Relational Observables

Since we obtain a unique solution to the Einstein equations only up to arbitrary changes of space-time coordinates, predictivity requires observable quantities to be coordinate independent, too. While coordinates are usually used in explicit calculations, values of observable quantities must not change when transforming to different coordinates. Abstractly, one can also formulate the concept of an observable in an explicitly coordinate-free manner leading to relational observables: evolution is then measured not with respect to coordinate time but with respect to other geometrical or matter quantities. While this is appealing conceptually, it can be hard to do explicitly.[11] For instance, in a cosmological situation one can measure how the value of a matter field changes with respect to a change in the total spatial scale or

[9] Solutions $x(x')$ depend on the auxiliary metric q^{ab} as well, as they should since no change in coordinates is necessary if q^{ab} had already been of canonical form with the chosen N and N^a.

[10] See also section 3.3 of George Ellis' contribution [5] for a related evaluation of the ADM formulation in this context.

[11] "Such a question can, we are assured, always be answered from a sufficient set of initial data, though the performance of this task may call for considerable mathematical agility." [6]

volume. Since in such a picture coordinates are eliminated, an alternative view on the question of dimensionality is possible. It also allows us to show, as we will see, how Minkowski space is recovered and what is special about special relativity.

Coordinate changes on a manifold imply transformations for fields such as g_{ab} on that manifold. Observable quantities then must be expressions formed by the fields of a theory being invariant under any change of coordinates. Simple examples are integrals of densities[12] over the whole space-time manifold, but they are too special and do not give one access to local properties. A more general, abstract way of constructing observables is as follows[13]: We use the group of transformations of our basic fields corresponding to coordinate changes $x^a \mapsto x'^a(x^b)$, which in our case is the group of space-time diffeomorphisms.[14] In an explicit realization, group elements would have infinitely many labels corresponding to four functions on space-time, or a space-time vector field $\xi^a(x)$. A relational observable requires one to choose a quantity f to be measured with respect to as many other functionals Φ^a_x of the basic fields as there are parameters of the group.[15] These functionals Φ^a_x will be called internal variables,[16] for gravity labeled by the space-time index a and a point x in space-time. This corresponds to the freedom in labels of the diffeomorphism group. From f and Φ^a_x we construct an observable[17] $F[\Phi^a_x]_{\phi^a_x}$ as a functional of the basic fields parameterized by time values ϕ^a_x as real numbers: to compute the evaluation $F[\Phi^a_x]_{\phi^a_x}(g_{ab})$ of the observable on a given set of basic fields g_{ab} we first find a coordinate transformation for which g'_{ab} becomes such that $\Phi^a_x(g'_{ab}) = \phi^a_x$ equal the chosen time parameters. The value of the observable is then defined to be the original function $f(g'_{ab})$ evaluated in this transformed set of basic fields. For

[12] A density is a mathematical object transforming in the same manner as $\sqrt{|\det g|}$ under changes of coordinates such that its coordinate integration is well-defined.

[13] This idea goes back to Komar and Bergmann [6, 7] and has more recently been elaborated and used in, e.g., [8–14].

[14] Space-time diffeomorphisms are in general not in one-to-one correspondence with coordinate changes. For our purposes, local considerations are sufficient where this identification can be made. A local coordinate change is then infinitesimally given by $x^a \mapsto x^a + \xi^a(x^b)$ where the vector field ξ^a is of compact support, and the same vector field generates a diffeomorphism.

[15] For a specific example, f could be a matter field and Φ^a_x the spatial volume $\det h_{ab}$ in an isotropic cosmological model. Here, the infinite number of variables Φ^a_x is reduced to only one by the high degree of symmetry. This corresponds to the fact that only spatially constant time reparameterizations respect the symmetry. Thus, the label a disappears because spatial coordinates cannot be changed in a relevant manner (they can be rescaled in some cases, without affecting the basic fields), and x disappears due to spatially constant reparameterizations. We will come back to possible reductions in the number of independent variables from the counterintuitive infinite size in the following subsection.

[16] Often, "internal time" or "clock variables" are used in this context, as these quantities are commonly employed to discuss the problem of time. However, since they do not only refer to time and it is even unclear in which sense time is involved, we prefer a neutral term.

[17] The notation, similar to that in [12], is quite loaded and indicates that $F[\Phi^a_x]$ is a relational object telling us how f changes under changes of the internal variables Φ^a_x. The answer depends parametrically on infinitely many real numbers ϕ^a_x: for each fixed set of these parameters, $F[\Phi^a_x]_{\phi^a_x}$ gives coordinate independent information on the relational behavior as a functional of the basic field g_{ab}.

any set of time variables ϕ_x^a one obtains a functional of the basic fields g_{ab}. This clearly results in an observable independent of the system of coordinates, and is well-defined at least for certain ranges of the fields and parameters involved.[18]

The observable, interpreted as measuring the change of f relationally with respect to the internal variables rather than with respect to coordinates does, however, depend on the parameters ϕ_x^a which crucially enter the construction. One is rather dealing with a family of observables labeled by these parameters. While one obtains an observable for each fixed set of parameters, its interpretation would be complicated and loose any dynamical information of change. This is probably one of the clearest indications for the dimensionality of the world from a mathematical point of view: What we are constructing directly are relational observables depending on parameters ϕ_x^a, roughly corresponding to a set of world lines. While this can be restricted to fixed parameters,[19] it would be a secondary step. Moreover, if all parameters are fixed, also spatial dependence is eliminated; in such a case we end up only with nonlocal observables. The primary observable quantities are thus not spatial at all but rather give, in an intricate, relational manner, a 4D world.

On second thought, there seems to be a problem because we have infinitely many parameters. From special relativity, or any kind of non-relativistic physics, we would expect only one time parameter in addition to three space parameters as independent variables.[20] On the other hand, special relativity is obtained from general relativity by introducing a background given by Minkowski space-time. Physically, this corresponds to synchronizing all clocks to measure time (and using a fixed set of rulers to measure lengths). When all clocks are synchronized, there is only one time parameter, and so it is not surprising after all that general relativity, lacking a synchronization procedure, requires infinitely many parameters ϕ_x^a for its observable quantities. The mathematical situation is thus in agreement with our physical expectations. We will now make this more explicit by showing how a Minkowski background can be reintroduced.

8.2.3 Recovering the Minkowski Background

The synchronization procedure can be implemented directly for general relational observables, clearly showing the reduction from infinitely many parameters to only one time coordinate. This brings us to the promised recovery of special relativity

[18] Global issues, as always in general considerations for general relativity, are much more difficult to handle.

[19] In fact, even though the ϕ_x^a are sometimes called "time parameters," only for one value of a does it really correspond to an infinity of times while the remaining parameters are space parameters. Having also space-parameters is actually an advantage in light of our earlier discussion where entire world-regions rather than any kind of lower-dimensional object were preferred. Such a world-region is then spanned by suitable ranges of all the parameters ϕ_x^a.

[20] This refers strictly only to one observer. In special relativity one considers time and space coordinates between boosted observers. For a given observer, the synchronization conditions of special relativity imply that there is only one time and three space parameters.

by re-introducing the Minkowski background and illustrates the relation between the infinitely many parameters of relational observables and the finite number of coordinates in Minkowski space. We make use of expressions derived recently for general relativity [11, 12]. We assume that four internal field variables Φ^a_x have been chosen, having conjugate momenta Π_a in a canonical formulation, which in a space-time region we are interested in are monotonic functions of x^b. For simplicity, we assume that these variables are four scalar fields which are already present in the theory, rather than more complicated functionals of basic fields such as curvature scalars used in [6, 7]. Moreover, we ignore their dynamics, i.e. assume that there are no potentials, since our aim here is to reconstruct the non-dynamical Minkowski space-time. Geometrically, the momenta are given by the (density weighted) derivatives of the internal variables along the unit normal to spatial slices,

$$\Pi^a = \sqrt{\det h} n^b \partial_b \Phi^a \,. \tag{8.7}$$

This determines the rate by which the fields change from slice to slice. In a region of monotonic fields, we can thus view $x^a \mapsto \Phi^a_x$ as a coordinate transformation and transform our metric accordingly, observing Eqs. (14.15) and (8.7):

$$g'^{ab} = \partial_c \Phi^a \partial_d \Phi^b \left(h^{cd} - n^c n^d \right) = \partial_c \Phi^a \partial_d \Phi^b h^{cd} - \Pi^a \Pi^b / \det\, h \tag{8.8}$$

or, splitting into time and space components,

$$g'^{00} = \partial_i \Phi^0 \partial_j \Phi^0 h^{ij} - \Pi^0 \Pi^0 / \det\, h \tag{8.9}$$

$$g'^{i0} = \partial_j \Phi^0 \partial_k \Phi^i h^{jk} - \Pi^i \Pi^0 / \det\, h \tag{8.10}$$

$$g'^{ij} = \partial_k \Phi^i \partial_l \Phi^j h^{kl} - \Pi^i \Pi^j / \det\, h \,. \tag{8.11}$$

First, we suppress the components Φ^i to bring out the role of time which will be played by Φ^0. We thus assume that the spatial metric h^{ij} is already given by δ^{ij} as in Minkowski space in its standard coordinate representation. Under the remaining transformation corresponding to Φ^0, the original spatial metric h^{ij} is transformed to the new spatial metric g'^{ij} which to preserve Minkowski space should also equal δ^{ij}. Since we suppressed the spatial parameters Φ^i, we need to require $\Pi^i = 0$ such that the spatial coordinate system is fixed in time.[21] Time synchronization then implies that Φ^0 does not depend on spatial coordinates, so also $g'^{i0} = 0$ is of Minkowski form. For the final component of the metric g'^{ab} we obtain $g'^{00} = -(\Pi^0)^2$ which is of Minkowski form for $\Pi^0 = 1$. These conditions can be summarized by saying that spatial coordinates do not change in time ($\Pi^i = 0$), and time progresses at the same constant pace everywhere ($\Pi^0 = 1$).

For instance from the construction of relational observables in [11] it follows that with such a choice of internal variables a relational observable takes the form

$$F\left[\Phi^a_x\right]_{\phi^0} = \sum_{k=0}^{\infty} \frac{1}{k!} \dot{f}(\Phi^i) \left(\phi^0 - \Phi^0 \right)^k = f\left(\Phi^i, \phi^0 - \Phi^0 \right) \tag{8.12}$$

[21] Thus, our set of rulers does not change in time.

where the dot refers to the change in f under a change of the time field Φ^0. If we identify space-time coordinates with Φ^a, any function will be observable since the background is completely fixed for a given observer.

This refers to one observer who has performed a full synchronization. If we change the observer, we obtain the usual Lorentz transformations and between two different observers time does certainly not proceed at the same pace. To see this, we now allow all four functions Φ^a to be nontrivial. We want to describe a situation where one synchronized observer is given by a system of the form just derived, such that $h^{ij} = \delta^{ij}$, whose internal variables we now call Ψ^a. From there, we transform to a new system of internal variables $\Phi^a(\Psi^b)$ such that also the metric g'^{ab} is Minkowski for the new synchronized observer. Thus, the right hand sides of Eqs. (8.9–8.11) must be Ψ^a-independent and only linear functions Φ^a are allowed:

$$\Phi^i = \omega^i_j \Psi^j + \alpha^i \Psi^0 \tag{8.13}$$

$$\Phi^0 = \beta_i \Psi^i + \gamma \Psi^0 . \tag{8.14}$$

Derivatives in Eqs. (8.9–8.11) are now taken with respect to Ψ^i, and $\Pi^a = \partial \Phi^a / \partial \Psi^0$. From Eq. (8.9) we then obtain $g'^{00} = -1 = \beta_i \beta_j \delta^{ij} - \gamma^2$ such that

$$\gamma = \sqrt{1 + |\beta|^2} \tag{8.15}$$

where $|\cdot|$ denotes the norm of vectors in $h_{ij} = \delta_{ij}$. From Eq. (8.10) in the form

$$g'^{i0} = \partial_j \Phi^0 \partial_k \Phi^i \delta^{jk} - \Pi^i \Pi^0$$

we have $0 = \beta_j \omega^i_k \delta^{jk} - \alpha^i \gamma$ such that

$$\alpha^i = \frac{\beta_j \omega^i_k \delta^{jk}}{\gamma} . \tag{8.16}$$

Finally, Eq. (8.11) in the form

$$g'^{ij} = \partial_k \Phi^i \partial_l \Phi^j \delta^{kl} - \Pi^i \Pi^j$$

implies

$$\delta^{ij} = \omega^i_k \omega^j_l \delta^{kl} - \alpha^i \alpha^j . \tag{8.17}$$

Defining

$$\rho^i_j := \omega^i_j - \frac{\alpha^i \beta_j}{1+\gamma} , \tag{8.18}$$

for which we have

$$\begin{aligned} \rho^i_k \rho^j_l \delta^{kl} &= \omega^i_k \omega^j_l \delta^{kl} - \frac{\omega^i_k \alpha^j \beta^k}{1+\gamma} - \frac{\omega^j_k \alpha^i \beta^k}{1+\gamma} + \frac{\alpha^i \beta_k \alpha^j \beta^k}{(1+\gamma)^2} \\ &= \omega^i_k \omega^j_l \delta^{kl} - \alpha^i \alpha^j \end{aligned}$$

using Eqs. (8.15) and (8.16), shows that the freedom in ω^i_j is given by an orthogonal matrix ρ^i_j. Thus, only a vector β^i and a rotation ρ^i_j can be chosen freely to specify a transformation. The remaining coefficients α^i and γ are then fixed by Eqs. (8.16) and (8.15). This is easily recognized as the usual coefficients of Lorentz transformations if we only identify $\beta^i = v^i/\sqrt{1-v^2/c^2}$ and use ρ^i_j as the rotational part of the transformation.

Allowing different synchronized observers, observable functions as in Eq. (8.12) have to be Lorentz invariant and are not arbitrary. Completely arbitrary, non-synchronized observers then require the general relativistic situation with complicated relational expressions for observables.

From our perspective, this shows that the usual space and time parameters one has in special relativity are what is left after fixing all but finitely many of the infinitely many parameters ϕ^a_x. These infinitely many parameters occur automatically when one attempts to write observables in a relational manner. In general, none of these parameters is distinguished as a possible time parameter to describe the evolution of a 3D world. In the relational picture, thus, only the 4D option is available.

8.3 Challenges and Resolutions

The canonical structure of relativity and an analysis of what is observable thus gives good reasons for the 4D of the world. Some difficulties certainly remain because, for one thing, we considered only local regions and had to assume that we can find functions Φ^a_x which are monotonic there. In order to describe the whole space-time in this manner we would need globally monotonic functions which may be difficult to find in general. For strictly physical purposes such a global description is also an over-idealization because all observations we can ever make are restricted to some bounded region of space-time, however big this region may be in cosmological observations. There are more severe potential challenges to this picture, one resulting from properties of general relativity not considered so far, and the other resulting from quantum theory.

8.3.1 Singularities

Locally, solutions to Einstein's field equations always exist and determine the space-time metric as well as manifold. This played a crucial role in our arguments given so far where we wanted to eliminate backgrounds and consider dynamical space-time. These equations are, however, nonlinear and so global aspects are more difficult to control. One consequence is that most solutions which we think are relevant for what we observe are singular when extrapolated in general relativity. They allow one to describe space-time only for a finite amount of proper time for some, and in some cases all, observers after which the classical theory breaks down [15]. This is usually accompanied by a divergence of curvature, but in any case represents a finite boundary to space-time.

If the theory does not allow us, even in principle, to extend solutions arbitrarily far in one direction, it may be difficult to view this direction as a dimension of the world. Here, the 3D viewpoint seems more suitable because we would simply have to deal with space and objects in time, described by the theory for some finite range of time. To be sure, there are also solutions where space is finite, but even if there are such boundaries space-time can usually be extended and they are thus artificial.[22] This is not the case with singularities. If we are interested in a 4D interpretation, then, we will have to deal with fundamental limitations to the extension of 4D objects, including space-time itself.

8.3.2 Quantum Aspects

Just as it was helpful to embed special relativity into general relativity for a wider viewpoint, the classical description is itself incomplete not the least because it leads to space-time singularities. This requires a corresponding extension of general relativity to a quantum theory of gravity. But even before this stage is reached, quantum properties do have a bearing on some of the arguments that can be used to decide on the dimensionality of the world. For instance, the 4D interpretation is advantageous because it embodies the fact that we have to recognize an object in order to denote it as such, showing that the time extension plays a central role in assigning object status. Such a recognition is not possible in quantum mechanics where identical particles are indistinguishable. We can then never be sure that a particle we recognize is the same one we saw before, and so assigning object status to worldlines or world-regions would not make sense unless all identical particles are subsumed in one and only one object.

8.3.3 Resolutions

These puzzles are resolved easily if one just considers suitable combinations of quantum theory with special and general relativity, respectively. Combining quantum theory with special relativity leads to quantum field theory where indeed the particle concept is weakened compared to the classical or quantum mechanical picture. There is not a collection of individual but indistinguishable particles, but a field

[22] There can also be boundaries to space arising from singularities where space-time cannot be extended in spatial directions. Such timelike singularities, however, do not generically arise in relevant cosmological or black hole solutions and thus can be ignored here. In homogeneous cosmological models, from which most of the cosmological intuition is derived, such timelike singularities are ruled out by the assumption of homogeneity (be it a precise or approximate symmetry) while for black holes timelike singularities arise for negative mass where the singular behavior is even welcome to rule out negative mass and argue for the stability of Minkowski space [16]. Other black hole solutions where timelike singularities arise, such as the Reissner–Nordstrom solution for electrically charged black holes in vacuum, are unstable to the addition of matter. Generic singularities are then spacelike or null [17, 18].

whose excitations may in some cases be interpreted as particles. Thus indeed, one is treating all identical particles as one single object, the corresponding field, and any problem with recognizability is removed automatically. The field is a function on space-time or a world-region, a 4D object.[23]

Singularities of general relativity pose a more complicated problem, but there are indications that they, too, are automatically dealt with when the underlying classical theory, this time general relativity, is combined with quantum theory. While the classical space-time picture breaks down at a singularity, several recent investigations have shown that quantum geometry continues to be well defined, albeit in a discrete manner [20–25]. One can then extend the classical space-time through a quantum region, or view space-time as fundamentally described by a quantum theory of gravity which reduces to general relativity in certain limits when curvature is not too large.

Indeed, background independent versions of quantum gravity are not formulated on a space-time manifold such that the question of whether the 3- or 4D viewpoint should be taken does not really arise at all. One is either dealing with space-time objects directly, such as in discrete path integral approaches, or employs a canonical quantization where the central object is a wave function on the space of geometries and observables are relational as discussed before. In quantum gravity, the 4D, relational viewpoint is thus even more natural than in classical gravity. It is also crucial for the results on non-singular behavior which are based on the relational behavior of wave functions or other quantities which now play the role of basic objects. Extensions beyond classical singularities can then be provided by considering the range of suitable internal variables and their quantizations: The relational dependence can, and in all cases studied so far will, continue through stages where one would classically encounter a singularity. This is much more robust than looking at possible modifications of field equations and corresponding extensions of space-times in coordinate form which have turned out to be non-generic if available at all.

8.4 Conclusions

The question of whether a theoretical object is just a mathematical construct or empirical is always difficult to address in physics. Often, the answer depends on what theory is used, which itself depends on current available knowledge. Not just the theoretical structure needs to be understood well but also its ontological underpinning. This is notoriously difficult if space and time are involved, and often hidden assumptions already enter constructions.

In such a situation, it is best to make use of as flexible a framework as possible and to eliminate any background structure. Thus, we focused on general relativistic dynamics rather than special relativistic kinematics. We have highlighted some relevant consequences using a canonical formulation. Canonical formulations are often

[23] In algebraic quantum field theory [19] one considers algebras associated with world-regions of "diamond" shape as the basic objects, so also here it is bounded regions in space-time determining what objects there are in the theory.

perceived as not being preferable because they break manifest covariance. However, they also offer well structured mathematical formulations and can be particularly illuminating for the dynamical behavior.

In particular, canonical techniques allow, even require one to discuss observables in a coordinate free manner. This leads to a relational description where no coordinates are used but instead field values are related to values of other fields to retrieve observable information. Usually, these quantities take the form of families of functionals parameterized by real numbers (most generally, infinitely many ones). In contrast to coordinates, these parameters do not distinguish between space and time and even the signature of a space-time metric is irrelevant. This formulation is then of the most democratic form and removes the danger of being misled by the different forms of space and time coordinates. A difference between those parameters arises in special situations such as when a Minkowski background is reintroduced. This illustrates, again, that background structures are to be eliminated as far as possible.

In addition to this extension from special to general relativity it is believed that a further one is necessary to combine it with quantum theory. A theory of quantum gravity in a reliable and completely convincing form is not yet available, but from what we know it does not seem to change much of the arguments presented here. It can even eliminate potential problems such as that of singularities. At a kinematical level, one can still imagine different possibilities concerning the dimensionality[24] but one still has the full parameter families corresponding to a 4D world when it comes to observables.

There may also be conceptual advantages of a 4D understanding. If the world and its objects are 4D, they are simply there and do not need to become. There is then no need to explain their origin, eliminating a difficult physical and philosophical question.[25]

References

1. R. M. Wald: *General Relativity* (The University of Chicago Press 1984)
2. R. Arnowitt, S. Deser, and C. W. Misner: *The Dynamics of General Relativity* (Wiley, New York 1962)
3. C. Barceló: *Lorentzian Spacetimes from Parabolic and Elliptic Systems of PDEs* (Springer, Fundamental Theories of Physics, Berlin 2006)
4. H. Friedrich and A. D. Rendall: The Cauchy Problem for the Einstein Equations, *Lect. Notes Phys.* **540**, 127–224 (2000), gr-qc/0002074
5. G. F. R. Ellis: *Physics in the Real Universe: Time and Spacetime* (Springer, Fundamental Theories of Physics, Berlin 2006), gr-qc/0605049

[24] Background-independent formulations seem to agree on a lower dimensional kinematical nature on microscopic scales [26–30].

[25] Also for this question, quantum gravity or cosmology seems to be affirmative: Initial conditions for quantum cosmological solutions, which have traditionally been imposed by intuitively motivated choices [31, 32], can arise directly from the dynamical laws [33, 34]. Thus, although completely unique scenarios are difficult to construct, 4D dynamics can automatically select solutions and to some degree eliminate additional physical input to formulate an origin.

6. P. G. Bergmann: Observables in General Relativity, *Rev. Mod. Phys.* **33**, 510–514 (1961)
7. A. Komar: Construction of a Complete Set of Independent Observables in the General Theory of Relativity, *Phys. Rev.* **111**, 1182–1187 (1958)
8. C. Rovelli: What is Observable in Classical and Quantum Gravity?, *Class. Quantum Grav.* **8**, 297–316 (1991)
9. C. Rovelli: Quantum Reference Systems, *Class. Quantum Grav.* **8**, 317–332 (1991)
10. A. Ashtekar, R. Tate, and C. Uggla: Minisuperspaces: Observables and Quantization, *Int. J. Mod. Phys. D* **2**, 15–50 (1993), gr-qc/9302027
11. B. Dittrich: Partial and Complete Observables for Hamiltonian Constrained Systems, *Class. Quantum Grav.* **23**, 6155–6184 (2006), gr-qc/0411013
12. B. Dittrich: Aspects of Classical and Quantum Dynamics of Canonical General Relativity, PhD thesis, University of Potsdam (2005)
13. D. Alba and L. Lusanna: Generalized Radar 4-Coordinates and Equal-Time Cauchy Surfaces for Arbitrary Accelerated Observers, gr-qc/0501090
14. S. B. Giddings, D. Marolf, and J. B. Hartle: Observables in effective gravity, *Phys. Rev. D* **74**, 064018 (2006), hep-th/0512200
15. S. W. Hawking and R. Penrose: The Singularities of Gravitational Collapse and Cosmology, *Proc. Roy. Soc. Lond. A* **314**, 529–548 (1970)
16. G. T. Horowitz and R. C. Myers: The Value of Singularities, *Gen. Rel. Grav.* **27**, 915–919 (1995), gr-qc/9503062
17. M. Dafermos: Black hole formation from a complete regular past, gr-qc/0310040
18. M. Dafermos: The interior of charged black holes and the problem of uniqueness in general relativity, gr-qc/0307013
19. R. Haag: *Local Quantum Physics* (Springer, Berlin, Heidelberg, New York 1992)
20. M. Bojowald: Absence of a Singularity in Loop Quantum Cosmology, *Phys. Rev. Lett.* **86**, 5227–5230 (2001), gr-qc/0102069
21. M. Bojowald: Homogeneous Loop Quantum Cosmology, *Class. Quantum Grav.* **20**, 2595–2615 (2003), gr-qc/0303073
22. M. Bojowald, G. Date, and K. Vandersloot: Homogeneous Loop Quantum Cosmology: The Role of the Spin Connection, *Class. Quantum Grav.* **21**, 1253–1278 (2004), gr-qc/0311004
23. A. Ashtekar and M. Bojowald: Quantum Geometry and the Schwarzschild Singularity, *Class. Quantum Grav.* **23**, 391–411 (2006), gr-qc/0509075
24. L. Modesto: Loop Quantum Black Hole, *Class. Quantum Grav.* **23**, 5587–5601 (2006), gr-qc/0509078
25. M. Bojowald: Non-singular Black Holes and Degrees of Freedom in Quantum Gravity, *Phys. Rev. Lett.* **95**, 061301 (2005), gr-qc/0506128
26. C. Rovelli and L. Smolin: Spin Networks and Quantum Gravity, *Phys. Rev. D* **52**, 5743–5759 (1995)
27. A. Ashtekar, J. Lewandowski, D. Marolf, J. Mourão, and T. Thiemann: Quantization of Diffeomorphism Invariant Theories of Connections with Local Degrees of Freedom, *J. Math. Phys.* **36**, 6456–6493 (1995), gr-qc/9504018
28. M. Reisenberger and C. Rovelli: Sum over Surfaces form of Loop Quantum Gravity, *Phys. Rev. D* **56**, 3490–3508 (1997), gr-qc/9612035
29. J. Ambjørn, J. Jurkiewicz, and R. Loll: Spectral Dimension of the Universe, *Phys. Rev. Lett.* **95**, 171301 (2005), hep-th/0505113
30. O. Lauscher and M. Reuter: Fractal Spacetime Structure in Asymptotically Safe Gravity, *JHEP* **0510**, 050 (2005), hep-th/0508202
31. A. Vilenkin: Quantum creation of universes, *Phys. Rev. D* **30**, 509–511 (1984)
32. J. B. Hartle and S. W. Hawking: Wave Function of the Universe, *Phys. Rev. D* **28**, 2960–2975 (1983)
33. M. Bojowald: Dynamical Initial Conditions in Quantum Cosmology, *Phys. Rev. Lett.* **87**, 121–301 (2001), gr-qc/0104072
34. M. Bojowald: Initial Conditions for a Universe, *Gen. Rel. Grav.* **35**, 1877–1883 (2003), gr-qc/0305069

Chapter 9
Relativity Theory Does Not Imply that the Future Already Exists: A Counterexample

Rafael D. Sorkin

Abstract It is often said that the relativistic fusion of time with space rules out genuine change or "becoming". I offer the classical sequential growth models of causal set theory as counterexamples.

1. Can one hold a "4D" point of view and still maintain consistently that things really happen? Is a space-time perspective compatible with the idea of "becoming"? Many authors have denied such a possibility, leaving us to choose between a static conception of reality and a return to the pre-relativistic notion of linear time. In contrast, I want to offer a concrete example – a theoretical model of causal set dynamics – that illustrates the possibility of a positive answer to the above questions, according to which reality is more naturally seen as a "growing being" than as a "static thing".

Of course, one might doubt whether the static and dynamic conceptions of reality differ in more than words, given that the distinction between them does not seem to find a home in the mathematics of general relativity. Do the Einstein equations look any different when they are viewed "dynamically" rather than "under the aspect of eternity"?, a skeptic might ask. Or just because the psychological feeling of "the now" impresses itself on our minds, should that really matter to us as physicists? Such questions threaten to lead off into an impassable terrain of metaphysics, metamathematics, and the meaning of meaning. But this does not mean that the questions "being or happening?", "static or dynamic?" lack practical significance for the working scientist, because the answers one gives will inform the *direction* in which one searches for new theoretical structures. Thus, for example, the dynamical scheme that I will be using for illustration has sprung from the search for a theory of quantum gravity. We will see that it does provide a sort of mathematical home for the idea of becoming (as a process of growth or birth) and that, conversely, one would have been hard-pressed to arrive at such a dynamical scheme without starting from the idea of a Markov process unfolding in time.

2. The model I'm referring to is that of *classical sequential growth* (CSG) regarded as a "law of motion" or "dynamical law" for causal sets. You can find in Ref. [1] and [2] a full mathematical description of the model, and in [3] an account

V. Petkov (ed.), *Relativity and the Dimensionality of the World*, 153–161.

of how one resolves in that context the complex of conceptual difficulties known to workers in quantum gravity as "the problem of time". Here I will just try to summarize the basic ideas with an emphasis on those aspects most germane to the present discussion.

A CSG model describes a stochastic process in which elements e of the growing causal set C are born one by one, each with a definite subset of the already born elements as its ancestors. If one records the ancestral relationships among a set of elements produced in this manner, the resulting "family tree" will be an instance of a *causal set* [4]. Mathematically characterized, what one obtains is, more precisely, a past-finite partial order in which x precedes y if x is an ancestor of y, x, and y being arbitrary elements of C. In a CSG dynamics, the specific births that occur are (with trivial exceptions) not determined in advance; rather they happen stochastically, in such a manner as to define a Markov process. A specific member of the family of CSG models is determined by the set of transition probabilities of the Markov process; and these in turn can be expressed in terms of the basic parameters or "coupling constants" of the theory, as explained in [1].

For example, let e_0 be the first-born element, e_1 the next born, etc. The birth of e_0 can be construed as a transition from the empty causal set to the (unique) causal set of one element, and it occurs with probability 1. The next birth, however, can occur in two different ways: either e_0 will be an ancestor of e_1 (written $e_0 \prec e_1$), or it will not; and each of these two events will happen, in general, with non-zero probability. After the third birth the possible outcomes number five, and at subsequent stages the number of possible causal sets rises rapidly. After the fourth birth, one can have any of 16 non-isomorphic causal sets, while after the tenth there are already over two million distinct possibilities (2567284 to be precise). How likely any one of these possibilities is to be realized depends on how the parameters of the model are chosen. At one extreme, each new element acquires all the previous elements as ancestors, and the result is a *chain*, the causal set equivalent of 1D Minkowski space. At the other extreme, none of the elements has ancestors (they are all "spacelike" to each other), and the result is an *antichain*, a causal set which does not correspond to any space-time (although it can have an interpretation as analogous to a spacelike hypersurface when it occurs embedded in a larger causal set.) In between these extremes lie the more interesting regions in parameter space, where one encounters, for example, CSG analogs of cyclical cosmologies, with coupling constants that get renormalized in such a way that the cosmos grows larger with each successive cycle of collapse and reexpansion.

For present purposes, the most important point is that the causal set is analogous to a space-time and the probabilities governing its growth play the role of the "law of motion" for the space-time (i.e. the Einstein equations in the specific case of source-free gravity.) Of course, those of us working with causal sets hope that there's more to all of this than an analogy. We hypothesize that continuous space-time is only an effective description of a deeper reality, a causal set whose dynamics is described by something very like a CSG model. To be physically realistic – and in particular to be able to generate a truly manifold-like causal set – this dynamics could not be classical; it would have to be quantal in an appropriate sense. A dynamical scheme of this sort is the ultimate objective of current work, but even though we don't possess it it yet, it is possible to imagine the kind of formalism to which it would correspond mathematically, namely the formalism of "generalized quantum mechanics" as codified in decoherence functionals and quantal measure theory. A dynamical scheme constructed along such lines would in the end be rather similar to a CSG model. The incorporation of interference (in the quantum sense) would mark a dramatic difference, of course, but the underlying kinematics or "ontology" would differ very little; and even the mathematical structure of the decoherence functional could find itself in close analogy to the probability measure that defines a CSG model.[1] In particular the criterion of "discrete general covariance" could carry over essentially unchanged from the classical to the quantal case. And since considerations of Lorentz invariance and general covariance seem to lie at the heart of the arguments against a dynamical conception of reality, it seems fair enough to reflect on them in the context of CSG models. Indeed, the cosmology of the CSG models is sufficiently realistic that it's hard to imagine a question of principle relating to the "being-becoming" dichotomy that could not be posed in this simplified context.

3. To make my example more convincing however, I should probably try to explain a bit more in what sense a continuum space-time can emerge from a causal set. Since this is essentially a kinematical question, it can be answered fairly satisfactorily in the present state of understanding. Indeed, I claim that the correspondence between certain causal sets and certain space-times is all I really need to make my case, because once you accept it, all that remains is to realize that a causal set can be generated by a process of "growth" or "birth" in a way that does not presuppose any notion of distant simultaneity or any concomitant notion of "space developing in time".

Perhaps a metaphor can bring out the key idea more clearly. Think of the causal set as an idealized growing tree (in the botanical sense, not the combinatorial one). Such a tree grows at the tips of its many branches, and these sites of growth are independent of one another. Perhaps a cluster of two leaves springs up at the tip of one branch (event A) and at the same moment a single leaf unfolds itself at the tip of a second branch (event B). To a good approximation, the words "at the same moment" make sense for real trees, but we know that they are not strictly accurate, because events A and B occur at different locations and distant simultaneity lacks objective meaning. If the tree were broad enough and the growth fast enough, we really could not say whether event A preceded or followed event B. The same should be true for the causal set. It is "growing at the tips" but not in a synchronized manner with respect to any external time. There is no single "now" that spreads itself over the entire process.[2]

"But wait a minute", you might object. "Didn't you just describe the CSG growth process as a succession of births in a definite order, and doesn't the resulting ranking

---

[1] Indeed, one can obtain a non-classical decoherence functional by letting the parameters of the CSG model become complex.

[2] Milič Čapek [5] has proposed a musical metaphor for essentially the same idea: that of a fugue. In such a composition, each voice can seem for a while to unfold in its own region of space, its notes neither later nor earlier than the other's, until the musical lines come back together and intersect.

of the elements of C imply something akin to a distant simultaneity?" The answer to this objection is that a definite birth-order, or an "external time", did figure in the description I gave, but it is to be regarded as an artifact of the description analogous to one's choice of coordinates for writing down the Schwarzschild metric. Only insofar as it reflects the intrinsic causal order of the causal set is this auxiliary time objective. The residue is "pure gauge". Thus, any other order of birth which is compatible with the intrinsic precedence relation $\prec$ is to be regarded as physically equivalent to the first, in the same sense that two diffeomorphic metrics are physically equivalent. So even though a CSG model rests on no background structure in the usual sense (unlike continuum gravity, where the underlying differentiable manifold acts as a background), one still meets with an issue very like that of general covariance, stemming from the entry of an external time-parameter into the mathematical definition of the model. To complete my argument, then, I will have to explain how this issue has been addressed by the formalism, but first let me carry on with the task of explaining how a space-time can emerge from a causal set in the first place.

Geometry = Number + Order

4. What might a stochastic process of the CSG type have to do with space-time and geometry, given that the type of mathematical object involved (a past-finite poset), is not only discrete, but is at first sight far removed from anything like a 4D manifold? Of course, the idea is that the continuity of space-time is illusory, that space-time itself is only an emergent reality, and that its inner basis is a causal set. In order for this to be the case, the apparently rather primitive structure of a discrete partial order must nonetheless conceal within itself the type of information from which a Lorentzian geometry can be recovered naturally, so that causal sets, or at least certain causal sets can be placed in correspondence with certain space-times. Fortunately the basis of this correspondence is easy to understand, at least in broad outline.

With respect to a fixed system of coordinates, the space-time metric appears as a symmetric matrix $g_{\alpha\beta}(x)$ of Lorentzian signature. It is therefore described by 10 real functions of the coordinates, $g_{00}(x)$, $g_{01}(x), \ldots, g_{33}(x)$. Of these, the combination $g_{\alpha\beta}/|\det g|^{1/4}$ is determined if we know the light cones (i.e. the solutions of $g_{\alpha\beta}v^\alpha v^\beta = 0$), and the remaining factor of $\det(g)$ is determined if we know the *volumes* of arbitrary space-time regions R since these are given by integrals of the form $\int_R \sqrt{-\det(g)}\,d^4x$. But we know the light cones once we know the causal ordering among the point-events of space-time, or in other words which point-events can influence which others (the space-time being assumed to carry a time-orientation). Now let us postulate (*i*) that this causal ordering directly reflects the ancestral relation $\prec$ in the causal set; and (*ii*) that space-time volume directly reflects the *number* of causal set elements going to make up the region in question (the number of births "occurring in it"). We then have the ingredients for constructing a four-geometry M, and if the construction succeeds, we may say that resulting M is a good approximation to the underlying causal set C: $M \approx C$. When this is the case, C may be identified with a subset of M, and it turns out to be important (for questions like locality and Lorentz invariance) that this subset needs to be randomly distributed in order to honor the postulate that number = volume.

Implications of Growth Models

5. Having introduced the dynamics of sequential growth, and having pointed out that a causal set growing in accord with such a model is capable in principle of yielding a relativistic space-time (not exactly, but to a sufficient approximation), I am tempted to stop at this point and let the example speak for itself. On one hand sequential growth seems to me to manifest "becoming" to the extent that any mathematical model can. It even provides an objective correlate of our subjective perception of "time passing" in the unceasing cascade of birth-events that build up the causal set, by "accretion" as it were[3]. On the other hand, there is nothing in the model corresponding to a 3D space "evolving in time". Rather, one meets with something which is "four dimensional" from the very beginning, but which at any stage of its growth is still *incomplete*.[4] It is true that if we stop the process at any stage, we can identify the maximal elements of C, and these form a kind of "future boundary" of the growing causal set. But this "boundary" A is an antichain, and as such can support intrinsically none of the metrical structures of physical space. It is only by reference to the many relations of causal precedence connecting the elements of A to their ancestors that geometrical attributes can be attached to A at all.[5] [6] Thus, the space-time character is primary in CSG models, and any approximate notion of "space like hypersurface" is derived from it. Besides, stopping the process at a given stage has no objective meaning within the theory, because with a different choice of birth-order, the causet at the same stage of growth would look entirely different.

The example of the CSG models seems to me to refute the contention that relativistic space-time is incompatible with genuine change, but I suppose that no example can ever bring to a close a debate that remains purely at the level of interpretation. If, on the other hand, we ask not whether becoming is "logically consistent" with 4D, but rather whether the combination of the two notions can be heuristically fruitful, then I think that the development of the CSG models is in fact persuasive evidence that it can.

---

[3] The notation of "accretive time" that arises here seems close to that of C.D. Broad, and also to that of the "Vibhajavadin" school within the Buddhist philosophical tradition.

[4] In this paper, I am using "4D" as a shorthand for "of a space-time character, as opposed to a purely spatial character". There is nothing in the definition of a causal set that limits it to a dimensionality of four, or indeed to any uniform dimensionality.

[5] Arguably, a reference to the enveloping space-time is present in the continuum as well, but there it is disguised by the fact that one need refer only to an arbitrarily small neighborhood of a hypersurface in order to define, for example, its extrinsic curvature, whence reference to any *specific* earlier or later point can be deemed irrelevant.

The only other way to argue would be to refute, one by one, the supposed proofs that becoming and 4D exclude each other; but that would have to be attempted by someone much better versed in those arguments than I am. Perhaps, however, it is fair to say that most such arguments presuppose that the sole alternative to the "block universe" is a doctrine that identifies reality with a 3D instant. If this is so then the conception that emerges from the CSG models of a 4D, but still incomplete reality should be able, if not to settle the debate, then at least to widen its terms in a fruitful manner.

If we attend to our actual experience of time then no difficulty ever arises, as pointed out long ago by Poincaré. Our "now" is (approximately) localized and if we ask whether a distant event spacelike to us has or has not happened yet, this question lacks intuitive sense. But the "opponents of becoming" seem not to content themselves with the experience of a "situated observer". They want to imagine themselves as a "super observer", who would take in all of existence at a glance. The supposition of such an observer *would* lead to a distinguished "slicing" of the causet, contradiction the principle that such a slicing lacks objective meaning ("covariance"). Super-observers do not exist however, and the attempt to put ourselves into their shoes brings the localized human experience of "the now" into conflict with the asynchronous multiplicity of "nows" of a CSG model (cf. the analogy of the growing tree).

6. Returning from metaphor to mathematics, I would like to deal briefly with two related features of relativity theory that arguments against "becoming" seem to rely upon at a technical level, namely Lorentz invariance and general covariance. To what extent might the concept of a growing causet clash with these features? With respect to the first, one can say something definite and quite rigorous: one can quote a theorem. With respect to the second, it is less easy to reach a sure conclusion since the meaning of general covariance in the context of a discrete and stochastic theory is still more elusive than it is in a continuous and deterministic setting.

Lorentz Transformations

In speaking of Lorentz invariance we remain essentially at the level of kinematics, because, so far as one knows, the extant CSG models can give rise to significant portions of Minkowski space-time $\mathbb{M}^4$ only with vanishingly small probability.[6] Let us suppose, nevertheless, that some quantal growth process has produced a causet C resembling a large region of flat space-time which we can idealize as being all of Minkowski space: $C \approx \mathbb{M}^4$. By definition, such a causal set is embeddable "randomly" in $\mathbb{M}^4$, just as if it had been created by running a Poisson process in $\mathbb{M}^4$. With respect to the Poisson probability measure one can then prove that, with probability unity, it is impossible to deduce a distinguished timelike direction from

[6] However, there seem to exist more general Markov processes that do produce, for example, the future of the origin in $\mathbb{M}^4$ [7]. These processes respect discrete general covariance in the sense of [1] but not Bell causality.

the embedding [8]. In this sense, Lorentz invariance is preserved exactly by the causal set, and one sees again how artificial would be its decomposition into any sequence of antichains or other analogs of space-at-a-moment-of-time.

General Covariance

7. Finally, let us return to the question of general covariance. In the familiar context of continuum relativity, this phrase has a double significance. In the first place it implies that only diffeomorphism-invariant quantities possess physical meaning (where the word "quantities" can be replaced, according to taste by "events, "questions", "predicates", or "properties").[7] Given a space-time metric, it is thus meaningful to ask for the maximum area of a black hole horizon but not for the value of the gravitational potential at coordinate radius 17. But aside from thus conditioning our definition of reality, general covariance also demands in the second place that a theory's equations of motion (or its action-functional) be diffeomorphism-invariant. Of course these two facets of general covariance are closely connected. Because general relativity is not a stochastic theory, it distinguishes rigidly between metrics that do and do not solve its field equations. Consequently, the second facet of covariance flows directly from the first as a consistency condition, because it would be senseless to identify two metrics one of which was allowed by the equations of motion and the other of which was forbidden; and conversely, the kinematical identification must be made if one wishes the dynamics to be deterministic. Thus, the first or "ontological" facet of general covariance tends to coalesce with its second or "dynamical" facet.

In relation to causal sets, the context changes because one is dealing with discrete structures and stochastic dynamics. If one tries to rethink the meaning of general covariance in this context, one can perhaps distinguish now three relatively independent facets. At the kinematic (or ontological) level, essentially the same words apply as in general relativity: the causal set elements "carry no inner identifiers", so that what has physical meaning is only the isomorphism equivalence class of the given poset C. General covariance for causets can thus be interpreted as invariance under relabeling, in analogy to the interpretation of general covariance as coordinate–invariance in the continuum.

But because the theory is stochastic, this label-independence does not impose any obvious consistency condition on the assignment of probabilities (or eventually quantal amplitudes) to causets. It simply implies that the only probabilities with physical meaning are those attached to isomorphism equivalence classes of causets. In the CSG models, this is made precise by beginning with a probability measure $\tilde{\mu}$ on a space $\tilde{\Omega}$ of *labeled* causets, and passing from it to the induced measure μ on the quotient space Ω whose members are the equivalence classes [3]. In effect the

[7] Underlying this limitation is the thought that space-time points – or in this case elements of the causal set – possess no individuality beyond what they inherit from their relations to each other. According to John Stachel, they have "quiddity" without "haecceity".

probability of an element of Ω is the sum of the probabilities of all the members of that equivalence class.[8]

The passage from $\tilde{\Omega}$ to Ω expresses (discrete) general covariance kinematically and the induced passage from $\tilde{\mu}$ to μ expresses it dynamically. But what would correspond here to the invariance of the equations of motion – and do we require any such condition? In the analogous situation of gauge theories in $\mathbb{M}^4$, one often "fixes the gauge", but then takes care to integrate the resulting – non gauge-invariant – probability measure over entire gauge equivalence classes. (I am thinking of the Fadeev-Popov approach to the Wick-rotated quantum field theory.) This would be the analog of letting $\tilde{\mu}$ depend on the labeling but only computing probabilities with respect to μ. However, before fixing the gauge one had a measure which *was* gauge invariant (though defined only formally), and this invariance is a crucial physical input to the theory. One might thus expect, in the causet case, that $\tilde{\mu}$ should itself be relabeling invariant, for only in this way would the classical limit of the corresponding quantum theory have a chance to reproduce the Einstein equations. It turns out that a natural invariance condition of this sort can be found, and it is one of the two key inputs to the CSG models. The specific condition[9] states that for any finite causet C of cardinality n, the probability to arrive at C after n births is independent of the birth order of C's elements (provided of course that we limit ourselves to orderings that can actually happen, i.e. to so called natural labelings of C). Since, by construction, the CSG models fulfill this condition, one can conclude that in these models there is no clash: discrete general covariance coexists harmoniously with the concept of a dynamically growing causal set.

8. In the CSG models, a form of spatiotemporal discreteness plays a prominent role, and for that reason alone, one might question whether the example with which we've been working carries over to the space-time continua of special and general relativity. On the other hand, the conception of a dynamically growing reality arguably retains its meaning in a continuous setting, even if it stretches one's intuition to a greater extent there. Accordingly, one might decide that "becoming" is not, after all, in conflict with the 4D Lorentzian manifold of Relativity Theory. In that case, their compatibility would be something that we might have recognized much earlier, without ever taking causal sets into consideration. The simplifying hypothesis of a

---

[8] In order to define μ consistently, one must take $\tilde{\Omega}$ to be a space of infinite causets, ones for which the growth process has "run to completion". We meet here with an echo of the block-universe idea, that is in effect built into mathematicians' formalisation of the concept of stochastic process.

[9] Notice that this condition of "discrete general covariance" is not itself formulated covariantly. Notice also that it says something less than the following formal statement: "the probability of a completed causet C (a causet of countably many elements) is independent of its labeling. This distinction was brought to light by Graham Brightwell, who also pointed out that this stronger statement *does* hold in the CSG models, even though it's not implied by discrete general covariance alone. One should also mention here an important difference between diffeomorphism-invariance and relabeling-invariance: the former is expressed by an invariance *group* arising as the automorphism group of a background structure (the manifold); the latter is not a group (a given permutation need not preserve naturality of the labeling) and there's no background (unless you count the integers $\mathbb{N}$ from which our labels or parameter time n come, but even if you do count them, their automorphism group is trivial and does not generate relabelings).

discrete spatiotemporal substructure would have served only as an inessential aid to our thinking.

On the other hand, one might in the end decide that a space-time *continuum* necessarily is static, even though – as we have just seen – a *discrete* structure can consistently "happen". In that case, an adherent of "becoming" could claim that our intuition of time as a flow, had we but listened to it attentively, was all along speaking to us of the discreteness of whatever process constitutes the inner basis of the phenomenon that we have been accustomed to conceptualizing as a space-time continuum.

Acknowledgment Research at Perimeter Institute for Theoretical Physics is supported in part by the Government of Canada through NSERC and by the Province of Ontario through MRI. This research was partly supported by NSF grant PHY-0404646.

References

1. David P. Rideout and Rafael D. Sorkin, A Classical Sequential Growth Dynamics for Causal Sets, *Phys. Rev. D* **61:** 024002 (2000) `gr-qc/9904062`
2. Madhavan Varadarajan and David Rideout, A General Solution for Classical Sequential Growth Dynamics of Causal Sets, *Phys. Rev. D* **73:** 104021 (2006) `gr-qc/0504066`
3. Graham Brightwell, Fay Dowker, Raquel S. García, Joe Henson, and Rafael D. Sorkin, Observables in Causal Set Cosmology, *Phys. Rev. D* **67:** 084031 (2003) `gr-qc/0210061`;
   Graham Brightwell, H. Fay Dowker, Raquel S. García, Joe Henson, and Rafael D. Sorkin, "General Covariance and the 'Problem of Time' in a Discrete Cosmology", in K.G. Bowden, Ed., *Correlations*, Proceedings of the ANPA 23 conference, held August 16–21, 2001, Cambridge, UK (Alternative Natural Philosophy Association, London, 2002), pp 1–17 `gr-qc/0202097`;
   Fay Dowker and Sumati Surya, "Observables in Extended Percolation Models of Causal Set Cosmology" *Class. Quant. Grav.* **23:** 1381–1390 (2006) `gr-qc/0504069`
4. Luca Bombelli, Joohan Lee, David Meyer, and Rafael D. Sorkin, Spacetime as a Causal Set, *Phys. Rev. Lett.* **59:** 521–524 (1987)
5. Milič Čapek, *The philosophical impact of contemporary physics* (Van Nostrand, 1961)
6. Seth Major, David Rideout, Sumati Surya, Spatial Hypersurfaces in Causal Set Cosmology, *Class.Quant.Grav.* **23:** 4743–4752 (2006) `gr-qc/0506133`;
   Seth Major, David Rideout, and Sumati Surya, On Recovering Continuum Topology from a Causal Set `gr-qc/0604124`
7. Graham Brightwell, unpublished notes on general covariance.
8. Luca Bombelli, Joe Henson, and Rafael D. Sorkin, Discreteness without symmetry breaking: a theorem (in preparation) `gr-qc/0605006`

Chapter 10
Absolute Being versus Relative Becoming

Joy Christian

Abstract Contrary to our immediate and vivid sensation of past, present, and future as continually shifting non-relational modalities, time remains as tenseless and relational as space in all of the established theories of fundamental physics. Here an empirically adequate generalized theory of the inertial structure is discussed in which proper time is causally compelled to be *tensed* within both space-time and dynamics. This is accomplished by introducing the inverse of the Planck time at the conjunction of special relativity and Hamiltonian mechanics, which necessitates energies and momenta to be invariantly bounded from above, and lengths and durations similarly bounded from below, by their respective Planck-scale values. The resulting theory abhors any form of preferred structure, and yet captures the transience of *now* along timelike world lines by causally necessitating a genuinely becoming universe. This is quite unlike the scenario in Minkowski space-time, which is prone to a block universe interpretation. The minute deviations from the special relativistic effects such as dispersion relations and Doppler shifts predicted by the generalized theory remain quadratically suppressed by the Planck energy, but may nevertheless be testable in the near future, for example via observations of oscillating flavor ratios of ultrahigh energy cosmic neutrinos, or of altering pulse rates of extreme energy binary pulsars.

10.1 Introduction

From the very first imprints of awareness, "change" and "becoming" appear to us to be two indispensable norms of the world. Indeed, *prima facie* it seems impossible to make sense of the world other than in terms of changing things and happening events through the incessant passage of time. And yet, the Eleatics, led by Parmenides, forcefully argued that change is nothing but an illusion, thereby rejecting the prevalent view, expounded by Heraclitus, that becoming is all there is. The great polemic that has ensued over these two diametrically opposing views of the world

V. Petkov (ed.), *Relativity and the Dimensionality of the World*, 163–195.

has ever since both dominated and shaped the course of western philosophy [1]. In modern times, influential neo-Eleatics such as McTaggart have sharpened the choice between the being and the becoming universe by distinguishing two different possible modes of temporal discourse, one with and the other without a clear reference to the distinctions of *past*, *present*, and *future*; and it is the former mode with explicit reference to the tenses that is deemed essential for capturing the notions of change and becoming [2]. Conversely, the latter mode – which relies on a *tenseless* linear ordering of temporal moments by a transitive, asymmetric, and irreflexive relation *precedes* – is deemed incapable of describing a genuine change or becoming. Such a sharpening of the temporal discourse, in turn, has inspired two rival philosophies of time, each catering to one of the two possible modes of the discourse [3]. One *tenseless* philosophy of time holds that time is *relational*, much like space, which clearly does not seem to "flow," and hence, what we perceive as the flow or passage of time must be an illusion. The other *tensed* philosophy of time holds, on the other hand, that there is more to time than mere relational ordering of moments. It maintains that time is rather a dynamic or evolving entity *unlike* space, and does indeed "flow" – like a refreshing river – much in line with our immediate experience of it. That is to say, far from being an illusion, our sensation of that sumptuous moment *now*, ceaselessly streaming-in from nowhere and slipping away into the unchanging past, happens to reflect a truly objective feature of the world.

In terms of these two rival philosophies of time, a genuinely becoming universe must then correspond to a notion of time that is more than a mere set of "static" moments, linearly ordered by the relation *precedes*. In addition, it must at least allow a genuine partition of this ordered set into the moments of *past*, *present*, and *future*. From the perspective of physics, the choice of a becoming universe must then necessitate a theory of space and time that not only distinguishes the future events from the past ones intrinsically, but also thereby accounts for the continual passage of the *fleeting present*, from a *nonexisting future* into the *unalterable past*, as a *bona fide* structural attribute of the world. Such a theory of space and time, which would account for the gradual *coming-into-being* of the nonexistent future events – or a continual accumulation of the unalterable past ones – giving rise to a truly becoming universe, may be referred to as a *Heraclitean* theory of space-time, as opposed to a *Parmenidean* one, devoid of any such explicit dictate to becoming.

One such Heraclitean theory of space-time was, of course, that of Newton, for whom "[A]bsolute, true, and mathematical time, of itself, and from its own nature, flow[ed] equably without relation to anything external..." [4]. To be sure, Newton well appreciated the relational attributes of time, and in particular their remarkable similarities with those of space:

> Just as the parts of duration are individuated by their order, so that (for example) if yesterday could change places with today and become the later of the two, it would lose its individuality and would no longer be yesterday, but today; so the parts of space are individuated by their positions, so that if any two could exchange their positions, they would also exchange their identities, and would be converted into each other *qua* individuals. It is only through their reciprocal order and positions that the parts of duration and space are understood to be the very ones that they truly are; and they do not have any other principle of individuation besides this order and position [5].

And yet, Newton did not fail to recognize the *non-relational*, or absolute, attributes of time that go beyond the mere relational ordering of moments. He clearly distinguished his neoplatonic notion of "equably" flowing absolute time, existing independently of changing things, from the Aristotelian notion of "unequably" flowing relative times, determined by their less than perfect empirical measures (such as clocks) [4]. What is more, he well appreciated the closely related need of a temporally founded theory of calculus within mathematics, formulated in terms of his notion of *fluxions* (i.e. continuously generated temporally flowing quantities [6]), and defended this theory vigorously against the challenges that arose from the quiescent theory of calculus put forward by Leibniz [6]. Thus, the notions of flowing time and becoming universe were central to Newton not only for his mechanics, but also for his mathematics [6]. More relevantly for our purposes, according to him the *rate* of flow of time – i.e. the *rate* at which the relationally ordered events succeed each other in the world – is determined by the respective moments of his absolute time, which flows by itself, continuously, uniformly, and unstoppably, without relation to anything external [7]. Alas, as we now well know, such a Newtonian theory of externally flowing absolute time, giving rise to an objectively becoming universe, is no longer physically viable. But is our celebration of Einstein's relativistic revolution complete *only* through an unconditional renunciation of Newton's non-relationally becoming universe?

The purpose of this essay, first, is to disentangle the notion of a becoming universe from that of an absolute time, and then to differentiate two physically viable and empirically distinguishable theories of space-time: namely, special relativity – which is prone to a *Parmenidean* interpretation – and a generalized theory [8] – which is intrinsically *Heraclitean* by construction. The purpose of this essay may also be taken as a case study in *experimental metaphysics*, since it evaluates conceivable experiments that can adjudicate between the two rival philosophies of time under discussion. Experimental metaphysics is a term suggested by Shimony [9] to describe the enterprise of sharpening of the disputes traditionally classified as metaphysical, to the extent that they can be subjected to controlled experimental investigations. A prime example of such an enterprise is the sharpening of a dispute over the novel conceptual implications of quantum mechanics, which eventually led to a point where empirical evidence was brought to bear on the traditionally metaphysical concerns of scientific realism [9]. Historically, recall how resistance to accept the novel implications of quantum mechanics had led to suggestions of alternative theories – namely, hidden variable theories. Subsequently, the efforts by de Broglie, von Neumann, Einstein, Bohr, Bohm, and others led to theoretical sharpening of the central concepts of quantum mechanics, which eventually culminated into Bell's incisive derivation of his inequalities. The latter, of course, was a breakthrough that made it possible to experimentally test the rival metaphysical positions on quantum mechanics [10]. As this well-known example indicates, however, experimental investigations alone cannot be expected to resolve profound metaphysical questions once and for all, without careful conceptual analyses. Indeed, Shimony [9] warns us against overplaying the significance of experimental metaphysics. He points out that without careful conceptual analyses even those questions that are traditionally classified as scientific cannot be resolved by experimental tests alone. Hence, it should not be surprising that questions as slippery as those concerning time and becoming would require more than a mere experimental input. On the other hand, as the above example proves, a judicious experimental input *can, indeed*, facilitate greatly towards a possible resolution of these questions.

Bearing these cautionary remarks in mind, the question answered, affirmatively, in the present essay is: Can the debate over the being versus becoming universe – which is usually also viewed as metaphysical [11] – be sharpened enough to bear empirical input? Of course, as the above example of hidden variable theories suggests, the first step towards any empirical effort in this direction should be to construct a physically viable Heraclitean alternative to special relativity. As alluded to above, this step has already been taken in Ref. [8], with motivations for it stemming largely from the temporal concerns in quantum gravity. What is followed up here is a comparison of these two alternative theories of causal structure with regard to the status of becoming. Accordingly, in the next section we begin by reviewing the status of becoming within special relativity. Then, in section 10.3, we review the alternative to special relativity proposed in Ref. [8], with an emphasis in subsection 10.3.3 on the causal inevitability of the strictly Heraclitean character of this alternative. Finally, before concluding, in section 10.4 we discuss the experimental distinguishability of the two alternatives, and its implications for the status of becoming.

10.2 The Status of Becoming Within Special Relativity

The prevalent theory of the local inertial structure at the heart of modern physics – classical or quantal, non-gravitational or gravitational – is, of course, Einstein's special theory of relativity. This theory, however, happens to be oblivious to any *structural* distinction between the past and the future [12]. To be sure, one frequently comes across references within its formalism to the notions of "absolute past" and "absolute future" of a given event. But these are mere conventional choices, corresponding to assignment of tenseless linear ordering to "static" moments mentioned above, with the ordering now being along the timelike world line of an ideal observer tracing through that event (see Fig. 10.1). There is, of course, no doubt about the objectivity of this ordering. It is preserved under Lorentz transformations, and hence remains unaltered for all inertial observers. But such a sequence of moments has little to do with becoming *per se*, as both physically and mathematically well appreciated by Newton [5, 6], and conceptually much clarified by McTaggart [2]. Worse still, there is no such thing as a worldwide moment "now" in special relativity, let alone the notion of a *passage* of that moment. Due to the relativity of simultaneity, what is a "now-slice" cutting through a given event for one observer would be a "then-slice" for another one moving relative to the first, and vise versa. In other words, what is *past* (or has "already happened") for one observer could be the *future* (or has "not yet happened") for the other, and vise versa [13]. This indeterminacy

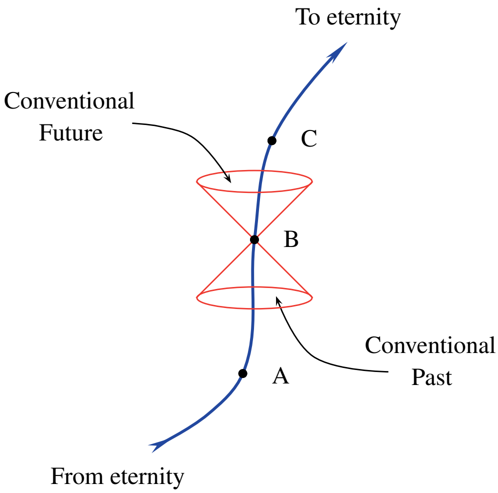

Fig. 10.1 Timelike world line of an observer tracing through an event B in a Minkowski space-time. Events A and C in the conventional past and conventional future of the event B are related to B by the transitive, asymmetric, and irreflexive relation *precedes*. Such a linear ordering of events is preserved under Lorentz transformations.

in temporal order cannot lead to any causal inconsistency however, for it can only occur for spacelike separated events – i.e. for pairs of events lying outside the light-cones of each other. Nevertheless, these facts suggest two rival interpretations for the continuum of events presupposed by special relativity: (1) an *absolute being* interpretation and (2) a *relative becoming* interpretation. According to the first of these interpretations, events in the past, present, and future exist all at once, *with equal ontological status*, across the whole span of time; whereas according to the second, events can be *partitioned*, causally, consistently, and ontologically, into the sets of definite past and indefinite future events, mediated by a fleeting present, albeit only in a relative and observer-dependent manner.

The first of these two interpretations of special relativity is sometimes also referred to as the "block universe" interpretation, because of its resemblance to a 4D block of "already laid out" events. The moments of time in this block are supposed to be no less actual than the locations in space are. Just as London and New York are supposed to be *there* even if you may not be *at* either of these locations, the moments of your birth and death are "there" on your time line, even if you are presently far from being "at" either of these two moments of your life. More precisely, along your timelike worldline all events of your life are fixed once and for all, beyond your control, and in apparent conflict with your freedom of choice. In fact, in special relativity, a congruence of such nonintersecting timelike world lines – sometimes referred to as a fibration of space-time – represents a 3D relative space (or an inertial frame). The 4D space-time is then simply filled by these "lifeless" fibers, with the proper time along any one of them representing the local time associated with the ordered series of events laid out along that fiber. Informally, such a fiber is a track in space-time of an observer moving subluminally *for all eternity*. In particular, for a given moment, all the future instants of time along this track – in exactly the same sense as all the past instants – are supposed to be fixed, once and for all, till eternity.

Such an interpretation of time in special relativity, of course, sharply differs from our everyday conception of time, where we expect the nonexistent future instants to *spring into existence* from nowhere, streaming-in one after another, and then *slipping away* into the unalterable past, thus gradually materializing the past track of our world line, as depicted in Fig. 10.2. In other words, in our everyday life we normally do not think of the future segment of our world line to be preexisting for all eternity; instead, we perceive the events in our lives to be occurring *nonfatalistically*, one after another, rendering our world line to "grow", like a tendril on a wall. But such a "dynamic" conception of time appears to be completely alien to the universe purported by special relativity. Within the Minkowski universe, as Einstein himself has been quoted as saying, "the becoming in three-dimensional space is transformed into a being in the world of four dimensions" [14]. More famously, Weyl has gone one step further in endorsing such a static view of the world: "The objective world simply *is*, it does not *happen*" [15]. Accordingly, the appearances of change and becoming are construed to be mere figments of our conscious experience, as Weyl goes on to explain: "Only to the gaze of my consciousness, crawling upward along

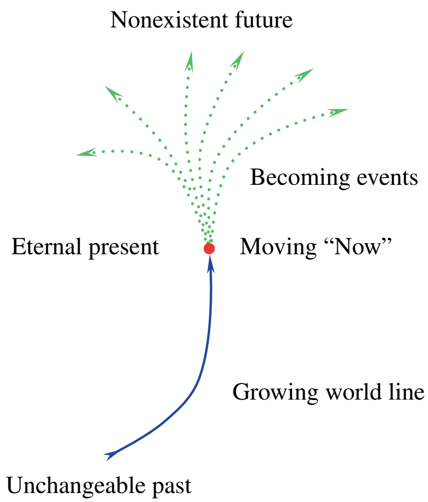

Fig. 10.2 The tensed time of the proverbial man in the street, with a degree in special relativity. His sensation of time is much richer than a mere tenseless linear ordering of events. Future events beyond the moving present are nonexistent to him, whereas he, at least, has a memory of the past events that have occurred along his world line.

the life line of my body, does a section of this world come to life as a fleeting image in space which continuously changes in time." Not surprisingly, some commentators have reacted strongly against such a grim view of reality:

> But this picture of a "block universe", composed of a timeless web of "world-lines" in a four-dimensional space, however strongly suggested by the theory of relativity, is a piece of gratuitous metaphysics. Since the concept of change, of something happening, is an inseparable component of the common-sense concept of time and a necessary component of the scientist's view of reality, it is quite out of the question that theoretical physics should require us to hold the Eleatic view that nothing happens in "the objective world." Here, as so often in the philosophy of science, a useful limitation in the form of representation is mistaken for a deficiency of the universe [16].

The frustration behind these sentiments is, of course, quite understandable. It turns out not to be impossible, however, to appease the sentiments to some extent. It turns out that a formal "becoming relation" of a limited kind can indeed be defined along a timelike world line, *uniquely* and *invariantly*, without in any way compromising the principles of special relativity. The essential idea of such a relation goes back to Putnam [17], who tried to demonstrate that no meaningful binary relation between two events can exist within the framework of special relativity that can ontologically partition a world line into distinct parts of *already settled past* and *not yet settled future*. Provoked by this and related arguments by Rietdijk [18] and Maxwell [19], Stein [20, 21] set out to expose the inconsistencies within such arguments (without unduly leaning on either side of the debate), and proved that a transitive, reflexive, and asymmetric "becoming relation" of a formal nature can indeed be defined consistently between causally connected pairs of events, on a time-orientable Minkowski space-time. Stein's analysis has been endorsed by Shimony [22] in an approach that is different in emphasis but complementary in philosophy, and extended by Clifton and Hogarth [23] to a more natural setting for the becoming along timelike world lines. This coherent set of arguments, taken individually or collectively, amounts to formally proving the permissibility of objective becoming within the framework of special relativity, but only *relative* to a given timelike world line. And since a timelike world line in Minkowski space-time is simply the integral curve of a never vanishing, future-directed, timelike vector field representing the direction of a moving observer, the becoming defended here is meaningful only *relative* to such an observer. There is, of course, no inconsistency in this relativization of becoming, since – thanks to the *absoluteness* of simultaneity for coincident events – different observers would always agree on which events have already "become", and which have not, when their world lines happen to intersect. Consequently, this body of works make it abundantly plain that special relativity *does not* compel us to adopt an interpretation as radical as the block universe interpretation, but leaves room for a rather sophisticated version of our common-sense conception of becoming. To be sure, this counterintuitive notion of a world line-*dependent* becoming permitted within special relativity is a far cry from our everyday experience, where a *worldwide* present seems to perpetually stream-in from a nonexistent future, and then slip away into the unchanging past. But such a pre-relativistic notion of absolute, worldwide becoming, occurring simultaneously for each and every one of us regardless of our motion, has no place in the postrelativistic physics. Moreover, this apparent absolute becoming can be easily accounted for as a gross collective of "local" or "individual" becomings along time-like world lines, emerging cohesively in the nonrelativistic limit. Just as Newtonian mechanics can be viewed as an excellent approximation to the relativistic mechanics for small velocities, our commonly shared "world-wide" becoming can be shown to be an excellent approximation to these relativistic becomings for small distances, thanks to the enormity of the speed of light in everyday units. Consequently, the true choice within special relativity should be taken not as between absolute being and absolute becoming, but between the former (i.e. block universe) and the *relativity of distant becoming*.

There has been rather surprising reluctance to accept this relativization of becoming, largely by the proponents of the block universe interpretation of special relativity. As brought out by Stein [21], some of this reluctance stems from elementary misconceptions regarding the true physical import of the theory, even by philosophers with considerable scientific prowess. There seems to remain a genuine concern, however, because the notion of world line-dependent becoming tends to go against our prerelativistic ideas of existence. This concern can be traced back to Gödel, who flatly refused to accept the relativity of distant becoming on such grounds: "A relative lapse of time, ... if any meaning at all can be given to this phrase, would certainly be something entirely different from the lapse of time in the ordinary sense, which means a change in the existing. The concept of existence, however, cannot be relativized without destroying its meaning completely" [24]. In the similar vein, in a certain book-review Callender remarks: "[T]he relativity of simultaneity poses a problem: existence itself must be relativized to frame. This *may* not be a contradiction, but it is certainly a queer position to hold" [25]. Perhaps. But nature cannot be held hostage to what our pre-relativistic prejudices find queer. Whether we like it or not, the Newtonian notion of absolute worldwide existence has no causal meaning in the post-relativistic physics. Within special relativity, discernibility of events existing at a distance is constrained by the absolute upper-bound on the speeds of causal propagation, and hence, the Newtonian notion of absolute distant existence becomes causally meaningless. To be sure, when we regress back to our everyday Euclidean intuitions concerning the causal structure of the world, the idea of relativized existence seems strange. However, according to special relativity the topology of this causal structure – i.e. the neighborhood relations between causally admissible events – happens *not* to be Euclidean but *pseudo*-Euclidean. Once this aspect of the theory is accepted, it is quite anomalous to hang on to the Euclidean notion of existence, or equivalently to the absoluteness of distant becoming. It is of course logically possible to accept the relativity of distant simultaneity but reject the relativity of distant becoming, as Gödel seems to have done, but conceptually that would be quite inconsistent, since the former relativity appears to us no less queer than the latter. In fact, perhaps unwittingly, some textbook descriptions of the relativity of simultaneity explicitly end up using the language of becoming. Witness for example, Feynman's description of a typical scenario [26]: "[E]vents that *occur* at two separated places at the same time, as seen by Moe in S', do not *happen* at the same time as viewed by Joe in S [emphasis rearranged].” Indeed, keeping the geometrical formalism intact, every statement involving the relativity of distant simultaneity in special relativity can be replaced by an identical statement involving the relativity of distant *becoming*, without affecting either the theoretical or the empirical content of the theory. In other words, Einstein could have written his theory using the latter relativity rather than the former, and that would have made no difference to the relativistic physics – classical or quantal – of the past hundred years. The former would have been then seen as a useful but trivial corollary of the latter. Thus, as Callender so rightly suspects, there is indeed no contradiction in taking the relativity of distant becoming seriously, since any evidence of our perceived co-becoming of objectively existing distant events (i.e., of our perceived *absoluteness* of becoming) is quite indirect and causal [22]. Therefore, the alleged queerness of the relativity of distant becoming by itself cannot be taken as a good reason to opt for an interpretation of special relativity as outrageous as the block universe interpretation.

There do exist other good reasons, however, that, on balance, land the block interpretation the popularity it enjoys. Einstein–Minkowski space-time is pretty “lifeless” on its own, as evident from comparisons of Figs. 10.1 and 10.2 above. If becoming is a truly ontological feature of the world, however, then we expect the sum total of reality to grow incessantly, by objective accretion of entirely newborn events. We expect this to happen as nonexistent future events momentarily come to be the present event, and then slip away into the unchanging past, as we saw in Fig. 10.2. No such objective growth of reality can be found within the Einstein–Minkowski framework for the causal structure. It is all very well for Stein to prove the definability of a two-place “becoming relation” within Minkowski space-time, but in a genuinely becoming universe no such relation between future events and a present event can be meaningful. Indeed, as recognize by Broad long ago, “[T]he essence of a present event is, not that it precedes future events, but that there is quite literally *nothing* to which it has the relation of precedence” [27]. Even more tellingly, in the Einstein–Minkowski framework there is no *causal compulsion* for becoming. In a genuinely becoming universe we would expect the accretion of new events to be necessitated *causally*, not left at the mercy of our interpretive preferences. In other words, we would expect the entire spatio-temporal structure to not only grow, but this growth to be also necessitated by causality itself. No such causal dictate to becoming is there in the Einstein–Minkowski framework of causality. A theory of local inertial structure with just such a causal necessity for objective temporal becoming is the subject matter of our next section.

10.3 A Purely Heraclitean Generalization of Relativity

Despite the fact that temporal transience is one of the most immediate and constantly encountered aspects of the world [11], Newton appears to be the last person to have actively sought to capture it, at the most fundamental level, within a successful physical theory. Equipped with his hypothetico-deductive methodology, he was not afraid to introduce metaphysical notions into his theories as long as they gave rise to testable experimental consequences. After the advent of excessively operationalistic trends within physics since the dawn of the last century, however, questions of metaphysical flavor – questions even as important as those concerning time – have tended to remain on the fringe of serious physical considerations.[1] Perhaps this explains why most of the popular approaches to the supposed quantum gravity are entirely oblivious to the profound controversies concerning the status of temporal becoming.[2] If, however, temporal becoming is indeed a genuinely ontological attribute of the world, then no approach to quantum gravity can afford to ignore it. After all, by quantum gravity one usually means a *complete* theory of nature. How can a complete theory of nature be oblivious to one of the most immediate and ubiquitous features of the world? Worse still: if temporal becoming is a genuine feature of the world, then how can any approach to quantum gravity possibly hope to succeed while remaining in total denial of its reality?

Partly in response to such ontological and methodological questions, an intrinsically Heraclitean generalization of special relativity was constructed in Ref. [8]. The strategy behind this approach was to judiciously introduce the inverse of the Planck time, namely t_P^{-1}, at the *conjunction* of special relativity and Hamiltonian mechanics, with a bottom-up view to a *complete theory of nature*, in a manner similar to how general relativity was erected by Einstein on special relativity (see Fig. 10.3). The resulting theory of the causal structure has already exhibited some remarkable physical consequences. In particular, such a judicious introduction of t_P^{-1} necessitates energies and momenta to be invariantly bounded from above, and lengths and durations similarly bounded from below, by their respective Planck scale values. By contrast, within special relativity nothing prevents physical quantities such as energies and momenta to become unphysically large – i.e. infinite – in a rapidly moving frame. In view of the primary purpose of the present essay, however, we shall refrain form dwelling too much into these physical consequences (details of which may be found in Ref. [8]). Instead, we shall focus here on those features of the generalized theory that accentuates its purely Heraclitean character.

10.3.1 Fresh Look at the Proper Duration in Special Relativity

To this end, let us reassess the notion of *proper duration* residing at the very heart of special relativity. Suppose an object system, equipped with an ideal classical clock

[1] There are, of course, a few brave hearts, such as Shimony [28] and Elitzur [29], who have time and again urged the physics community to take temporal becoming seriously. However, there are also those who have preferred to explain it away as a counterfeit, resulting from some sort of "macroscopic irreversibility" [30–32].

[2] A welcome exception is the causal set approach initiated by Sorkin [33]. However, the stochasticity of "growth dynamics" discovered *a posteriori* within this discrete approach is a far cry from the inevitable continuity of becoming recognized by Newton [6]. Such a deficiency seems unavoidable within any discrete approach to quantum gravity, due to the "inverse problem" of recovering the continuum [34].

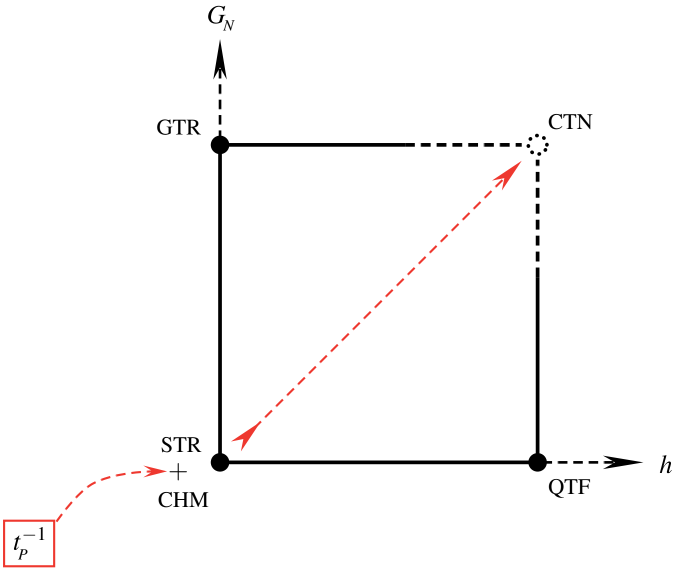

Fig. 10.3 Introducing the inverse of the Planck time at the conjunction of Special Theory of Relativity (STR) and Classical Hamiltonian Mechanics (CHM), with a bottom-up view to a Complete Theory of Nature (CTN). Both General Theory of Relativity (GTR) and Quantum Theory of Fields (QTF) are viewed as limiting cases, corresponding to negligible quantum effects (represented by Planck's constant h) and negligible gravitational effects (represented by Newton's constant G_N), respectively.

of unlimited accuracy, is moving with a uniform velocity $\mathbf{v}$ in a Minkowski spacetime $\mathcal{M}$, from an event e_1 at the origin of a reference frame to a nearby event e_2 in the future light cone of e_1, as shown in Fig. 10.4a. For our purposes, it would suffice to refer to this system, say of n degrees of freedom, simply as "the clock." As it moves, the clock will also *necessarily* evolve, as a result of its external motion, say at a uniform rate ω, from one state, say s_1, to another state, say s_2, within its own relativistic phase space, say $\mathcal{N}$. In other words, the inevitable evolution of the clock from s_1 to s_2 – or rather that of its state – will trace out a unique trajectory in the phase space $\mathcal{N}$, as shown in Fig. 10.4b. For simplicity, we shall assume that this phase space of the clock is finite dimensional; apart from possible mathematical encumbrances, the reasoning that follows would go through unabated for the case of infinite dimensional phase spaces (e.g. for clocks made out of relativistic fields).

Now, nothing prevents us from thinking of this motion and evolution of the clock conjointly, as taking place in a combined $4+2n-$dimensional space, say $\mathcal{E}$, the elements of which may be called *event-states* and represented by pairs $(\mathrm{e}_i, \mathrm{s}_i)$, as depicted in Fig. 10.5. Undoubtedly, it is this combined space that truly captures the complete specification of all possible physical attributes of our classical clock. Therefore, we may ask: *What will be the time interval actually registered by the*

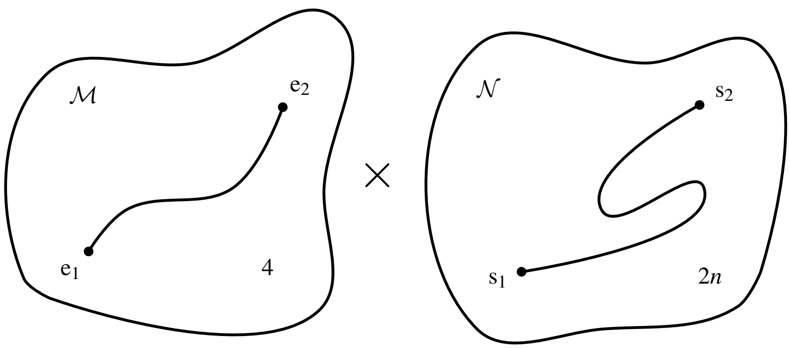

Fig. 10.4 (a) The motion of a clock from event e_1 to event e_2 in a Minkowski Space-time $\mathcal{M}$. (b) As the clock moves from e_1 to e_2, it also inevitably evolves, as a result of its external motion, from state s_1 to state s_2 in its own $2n$-dimensional phase space $\mathcal{N}$.

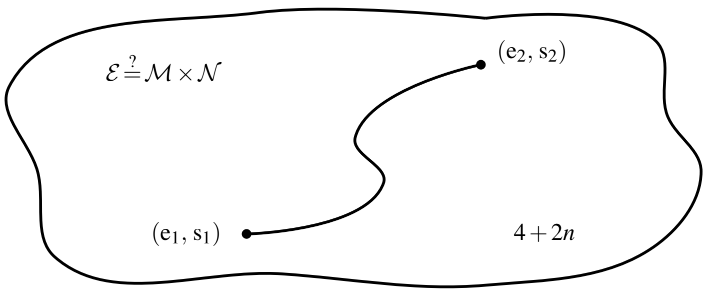

Fig. 10.5 What is the correct metric-topology of the combined space $\mathcal{E}$ – made up of the external Minkowski space-time $\mathcal{M}$ and the internal space of states $\mathcal{N}$ – in which our clock moves as well as evolves from event-state (e_1, s_1) to event-state (e_2, s_2)?

clock as it moves and evolves from the event-state (e_1, s_1) *to the event-state* (e_2, s_2) *in this combined space* $\mathcal{E}$? It is only by answering such a physical question can one determine the correct topology and geometry of the combined space in the form of a metric, analogous to the Minkowski metric corresponding to the line element

$$d\tau_{_E}^2 = dt^2 - c^{-2}d\mathbf{x}^2 \geq 0\,, \tag{10.1}$$

where the inequality asserts the causality condition. Of course, after Einstein the traditional answer to the above question, in accordance with the line element in Eq. (10.1), is simply

$$\Delta\tau_E = \int_{(\mathrm{e}_1,\mathrm{s}_1)}^{(\mathrm{e}_2,\mathrm{s}_2)} d\tau_E = \int_{\mathrm{t}_1}^{\mathrm{t}_2} \frac{1}{\gamma(v)}\, dt\,, \tag{10.2}$$

with the usual Lorentz factor

$$\gamma(v) := \frac{1}{\sqrt{1 - v^2/c^2}} > 1. \tag{10.3}$$

In other words, the traditional answer is that the metrical topology of the space $\mathcal{E}$ is of a product form, $\mathcal{E} = \mathcal{M} \times \mathcal{N}$, and – more to the point – the clock that records the duration $\Delta\tau_E$ in question remains *insensitive* to the passage of time that marks the evolution of variables within its own phase space $\mathcal{N}$.

But from the above perspective – i.e. from the perspective of Fig. 10.5 – it is evident that Einstein made an implicit assumption while proposing the proper duration in Eq. (10.2). He tacitly assumed that the rate at which a given physical state can evolve remains *unbounded*. Of course, he had no particular reason to question the limitlessness of how fast a physical state can evolve. However, for us – from what we have learned from our efforts to construct a theory of quantum gravity – it is not unreasonable to suspect that the possible rate at which a physical state can evolve is invariantly bounded from above. Indeed, it is generally believed that the Planck-scale marks a threshold beyond which our theories of space and time, and possibly also of quantum phenomena, are unlikely to survive [8, 35, 36]. In particular, the Planck time t_P is widely thought to be *the* minimum possible duration. It is then only natural to suspect that the inverse of the Planck time – namely t_P^{-1}, with its approximate value of 10^{+43} Hertz in ordinary units – must correspond to the absolute upper bound on how fast a physical state can possibly evolve. In this context, it is also worth noting that the speed of light is simply a ratio of the Planck length over the Planck time, $c := l_P/t_P$, which suggests that perhaps the assumption of absolute upper bound t_P^{-1} on possible rates of evolution should be taken to be more primitive in physical theories than the usual assumption of absolute upper bound c on possible speeds of motion. In fact, as we shall see, the assumption of upper bound c on speeds of motion can indeed be viewed as a special case of our assumption of upper bound t_P^{-1} on rates of evolution.

To this end, let us then systematize the above thoughts by incorporating t_P^{-1} into a physically viable and empirically adequate theory of the local causal structure. One way to accomplish this task is to first consider a simplified picture, represented by what is known as the extended phase space, constructed within a global inertial frame in which the clock is at rest (see Fig. 10.6). Now, in such a frame the proper time interval the clock would register is simply the Newtonian time interval Δt. Using this time $t \in \mathbb{R}$ as an external parameter, within this frame one can determine the extended phase space $\mathcal{O} = \mathbb{R} \times \mathcal{N}$ for the dynamical evolution of the clock using the usual Hamiltonian prescription. Suppose next we consider time-dependent canonical transformations of the dimensionless phase space coordinates $\mathrm{y}^\mu(t)$ $(\mu = 1, \ldots, 2n)$, expressed in Planck units, into coordinates $\mathrm{y}'^\mu(t)$ of the following general linear form:

$$\mathrm{y}'^\mu(\mathrm{y}^\mu(0), t) = \mathrm{y}^\mu(0) + \omega_r^\mu(\mathbf{y}(0))\, t + b^\mu\,, \tag{10.4}$$

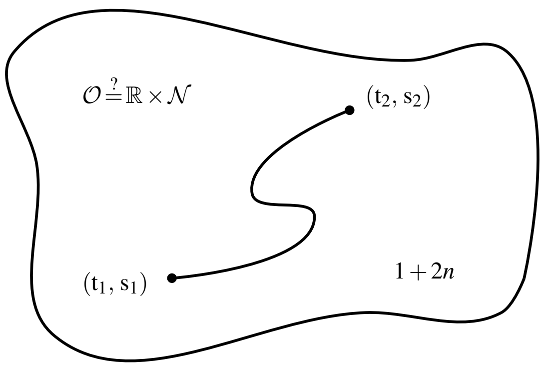

Fig. 10.6 The evolution of our clock from instant-sate $(\mathrm{t}_1, \mathrm{s}_1)$ to instant-state $(\mathrm{t}_2, \mathrm{s}_2)$ in the odd dimensional extended phase space $\mathcal{O}$. What is the correct topology of $\mathcal{O}$?

where ω_r^μ and b^μ do not have explicit time dependence, and the reason for the subscript r in ω_r^μ, which stands for "relative", will become clear soon. Interpreted actively, these are simply the linearized solutions of the familiar Hamiltonian flow equations,

$$\frac{d\mathrm{y}^\mu}{dt}\frac{d\mathrm{y}^\mu}{dt} = \omega_r^\mu\left(\mathbf{y}(t)\right) := \Omega^{\mu\nu}\frac{\partial H}{\partial \mathrm{y}^\nu}, \tag{10.5}$$

where ω_r is the Hamiltonian vector field generating the flow, $\mathbf{y}(t)$ is a $2n$-dimensional local Darboux vector in the phase space $\mathcal{N}$, Ω is the symplectic 2-form on $\mathcal{N}$, and H is a Hamiltonian function governing the evolution of the clock. If we now denote by ω^μ the uniform time rate of change of the canonical coordinates y^μ, then the linear transformations of Eq. (10.4) imply the composition law

$$\omega'^\mu = \omega^\mu + \omega_r^\mu \tag{10.6}$$

for the evolution rates of the two sets of coordinates, with $-\omega_r^\mu$ interpreted as the rate of evolution of the transformed coordinates with respect to the original ones. Crucially for our purposes, what is implicit in the law in Eq. (10.6) is the assumption that there is no upper bound on the rates of evolution of physical states. Indeed, successive transformations of the type in Eq. (10.4) can be used, along with Eq. (10.6), to generate arbitrarily high rates of evolution for the state of the clock. More pertinently, the assumed validity of the composition law of Eq. (10.6) turns out to be equivalent to assuming the absolute simultaneity of "instant-states" $(\mathrm{t}_i, \mathrm{s}_i)$ within the $1+2n-$dimensional extended phase space $\mathcal{O}$. In other words, within the $1+2n-$dimensional manifold $\mathcal{O}$, the $2n-$dimensional phase spaces simply constitute strata of hypersurfaces of simultaneity, much like the strata of spatial hypersurfaces within a Newtonian space-time. Indeed, the extended phase spaces such as $\mathcal{O}$

are usually taken to be contact manifolds, with topology presumed to be a product of the form $\mathbb{R} \times \mathcal{N}$.

Thus, not surprisingly, the assumption of absolute time in contact spaces is equivalent to the assumption of "no upper bound" on the possible rates of evolution of physical states. Now, in accordance with our discussion above, suppose we impose the following upper bound on the evolution rates[3]:

$$\left| \frac{d\mathbf{y}}{dt} \right| =: \omega \leq t_P^{-1}\,. \tag{10.7}$$

If this upper bound is to have any physical significance, however, then it must hold for *all* possible evolving phase space coordinates $\mathrm{y}^\mu(t)$, and that is amenable if and only if the composition law in Eq. (10.6) is replaced by

$$\omega'^\mu = \frac{\omega^\mu + \omega_r^\mu}{1 + t_P^2\, \omega^\mu\, \omega_r^\mu}\,, \tag{10.8}$$

which implies that as long as neither ω^μ nor ω_r^μ exceeds the causal upper bound t_P^{-1}, ω'^μ also remains within t_P^{-1}. Of course, this generalized law of composition has been inspired by Einstein's own such law for velocities, which states that the velocity, say v^k ($k = 1, 2$, or 3), of a material body in a given direction in one inertial frame is related to its velocity, say v'^k, in another frame, moving with a velocity $-v_r^k$ with respect to the first, by the relation

$$v'^k = \frac{v^k + v_r^k}{1 + c^{-2}\, v^k\, v_r^k}\,. \tag{10.9}$$

Thus, as long as neither v^k nor v_r^k exceeds the upper bound c, v'^k also remains within c. It is this absoluteness of c that lends credence to the view that it is merely a conversion factor between the dimensions of time and space. This fact is captured most conspicuously by the quadratic invariant of Eq. (10.1) of space-time. In exact analogy, if we require the causal relationships among the possible instant-states $(\mathrm{t}_i,\, \mathrm{s}_i)$ in $\mathcal{O}$ to respect the upper bound t_P^{-1} in accordance with the law (10.8), then the usual product metric of the space $\mathcal{O}$ would have to be replaced by the pseudo-Euclidean metric corresponding to the line element

$$d\tau_H^2 = dt^2 - t_P^2 d\mathbf{y}^2 \geq 0, \tag{10.10}$$

where the phase space line element $d\mathbf{y}$ was discussed in the footnote 3 below. But then, in the resulting picture, different canonical coordinates evolving with nonzero relative rates would differ in general over which instant-states are simultaneous with a given instant-state. As unorthodox as this new picture may appear to be, it is an inevitable consequence of the upper-bound Eq. (10.7).

---

[3] The "flat" Euclidean metric on the phase space that is being used here is the "quantum shadow metric", viewed as a classical limit of the Fubini-Study metric of the quantum state space (namely, the projective Hilbert space), in accordance with our bottom-up philosophy depicted in Fig. 10.3. See Ref. [8] for further details.

Let us now raise a question analogous to the one raised earlier: In its rest frame, what will be the time interval registered by the clock as it evolves from an instant-state (t_1, s_1) to an instant-state (t_2, s_2) within the space $\mathcal{O}$? The answer, according to the pseudo-Euclidean line element in Eq. (10.10), is clearly

$$\Delta\tau_{_H} = \int_{(\mathrm{t}_1,\mathrm{s}_1)}^{(\mathrm{t}_2,\mathrm{s}_2)} d\tau_{_H} = \int_{\mathrm{t}_1}^{\mathrm{t}_2} \frac{1}{\gamma(\omega)}\, dt = \frac{|\,\mathrm{t}_2 - \mathrm{t}_1|}{\gamma(\omega)} = \frac{\Delta t}{\gamma(\omega)}\,, \tag{10.11}$$

where

$$\gamma(\omega) := \frac{1}{\sqrt{1 - t_{_P}^2\,\omega^2}} > 1. \tag{10.12}$$

Thus, if the state of the clock is evolving, then we will have the phenomenon of "time dilation" even in the rest frame. Similarly, we will have a phenomenon of "*state* contraction" in analogy with the phenomenon of "length contraction":

$$\Delta\mathrm{y}' = \omega\,\Delta\tau_{_H} = \frac{\omega\,\Delta t}{\gamma(\omega)} = \frac{\Delta\mathrm{y}}{\gamma(\omega)}\,. \tag{10.13}$$

It is worth emphasizing here, however, that, as in ordinary special relativity, nothing is actually "dilating" or "contracting". All that is being exhibited by these phenomena is that the two sets of mutually evolving canonical coordinates happen to differ over which instant-states are simultaneous.

So far, to arrive at the expression of Eq. (10.10) for the proper duration, we have used a specific Lorentz frame, namely the rest frame of the clock. In a frame with respect to which the same clock is uniformly moving, the expression for the actual proper duration can be obtained at once from Eq. (10.10), by simply using the Minkowski line element in Eq. (10.1), yielding

$$d\tau^2 = dt^2 - c^{-2}d\mathbf{x}^2 - t_{_P}^2 d\mathbf{y}^2 \geq 0. \tag{10.14}$$

This, then, is the $4+2n$−dimensional quadratic invariant of our combined space $\mathcal{E}$ of Fig. 10.5. We may now return to our original question and ask: What, according to this generalized theory of relativity, will be the proper duration registered by a given clock as it moves and evolves from an event-state $(\mathrm{e}_1, \mathrm{s}_1)$ to an event-state $(\mathrm{e}_2, \mathrm{s}_2)$ in the combined space $\mathcal{E}$? Evidently, according to the quadratic invariant of Eq. (10.14), the answer is simply:

$$\Delta\tau = \int_{(\mathrm{e}_1,\mathrm{s}_1)}^{(\mathrm{e}_2,\mathrm{s}_2)} d\tau = \int_{\mathrm{t}_1}^{\mathrm{t}_2} \frac{1}{\gamma(v,\,\omega)}\, dt\,, \tag{10.15}$$

with

$$\gamma(v,\,\omega) := \frac{1}{\sqrt{1 - c^{-2}\,v^2 - t_{_P}^2\,\omega^2}} > 1. \tag{10.16}$$

We are now in a position to isolate the two basic postulates on which the generalized theory of relativity developed above can be erected in the manner analogous to the usual special relativity. In fact, the first of the two postulates can be taken to be Einstein's very own first postulate, except that we must now revise the meaning of inertial coordinate system. In the present theory it is taken to be a system of $4+2n$ dimensions, "moving" uniformly in the combined space $\mathcal{E}$, with 4 being the external space-time dimensions, and $2n$ being the internal phase space dimensions of the system. Again, the internal dimensions of the object system can be either finite or infinite in number. Next, note that by eliminating the speed of light in favor of pure Planck-scale quantities the quadratic invariant of Eq. (10.14) can be expressed in the form

$$d\tau^2 = dt^2 - t_P^2 \left\{ l_P^{-2} d\mathbf{x}^2 + d\mathbf{y}^2 \right\} \geq 0\,, \tag{10.17}$$

where l_P is the Planck length of the value $\sim 10^{-33}$ cm in ordinary units. The two postulates of generalized relativity may now be stated as follows:

1. The laws governing the states of a physical system are insensitive to "the state of motion" of the $4+2n-$dimensional reference coordinate system in the pseudo-Euclidean space $\mathcal{E}$, as long as it remains "inertial".
2. No time rate of change of a dimensionless physical quantity, expressed in Planck units, can exceed the inverse of the Planck time.

Clearly, the generalized invariance embedded within this new causal theory of local inertial structure is much broader in its scope – both physically and conceptually – than the invariance embedded within special relativity. For example, in the present theory even the four dimensional continuum of space-time no longer enjoys the absolute status it does in Einstein's theories of relativity. Einstein dislodged Newtonian concepts of absolute time and absolute space, only to replace them by an analogous concept of *absolute space-time* – namely, a continuum of *in principle* observable events, idealized as a connected pseudo-Riemannian manifold, with observer-independent space-time intervals. Since it is impossible to directly observe this remaining absolute structure without recourse to the behavior of material objects, perhaps it is best viewed as the "ether" of the modern times, as Einstein himself occasionally did [37]. By contrast, it is evident that in the present theory even this 4D space-time continuum has no absolute, observer-independent meaning. In fact, apart from the laws of nature, there is very little absolute structure left in the present theory, for now even the quadratic invariant of Eq. (10.14) is dependent on the phase space structure of the material system being employed.

10.3.2 Physical Implications of the Generalized Theory of Relativity

Although not our main concern here,[4] it is worth noting that the generalized theory of relativity described above is both a physically viable and empirically adequate theory. In fact, in several respects the present theory happens to be physically better

[4] In this subsection we shall only briefly highlight the physical implications of the generalized theory of relativity. For a complete discussion see section VI of Ref. [8].

behaved than Einstein's Special Theory of Relativity. For instance, unlike in special relativity, in the present theory physical quantities such as lengths, durations, energies, and momenta remain bounded by their respective Planck-scale values. This physically sensible behavior is due to the fact that present theory assumes even less preferred structure than special relativity, by positing democracy among the internal phase space coordinates in addition to that among the external space-time coordinates.

Mathematically, this demand of combined democracy among space-time and phase space coordinates can be captured by requiring invariance of the physical laws under the $4+2n-$dimensional coordinate transformations [8]

$$z^A = \Lambda^A{}_B \, z'^B + b^A \tag{10.18}$$

analogous to the Poincaré transformations, with the index $A = 0, \ldots, 3+2n$ now running along the $4+2n$ dimensions of the manifold $\mathcal{E}$ of Fig. 10.5. These transformations would preserve the quadratic invariant in Eq. (10.17) if the constraints

$$\Lambda^A{}_C \Lambda^B{}_D \, \xi_{AB} = \xi_{CD} \tag{10.19}$$

are satisfied, where ξ_{AB} are the components of the metric on the manifold $\mathcal{E}$. At least for simple finite dimensional phase spaces, the coefficients $\Lambda^A{}_B$ are easily determinable. For example, consider a massive relativistic particle at rest (and hence, also not evolving) with respect to a primed coordinate system in the external space-time, which is moving with a uniform velocity $\mathbf{v}$ with respect to another unprimed coordinate system. Since, as it moves, the state of the particle will also be evolving in its 6D phase space, say at a uniform rate ω, we can view its motion and evolution together with respect to a $4+6-$dimensional unprimed coordinate system in the space $\mathcal{E}$.

Restricting now to the external spatiotemporal sector where we actually perform our measurements, it is easy to show [8] that the coefficients $\Lambda^A{}_B$ are functions of the generalized gamma factor of Eq. (10.16), with the corresponding expression for the length contraction being

$$\Delta \mathrm{x}' = \frac{\Delta \mathrm{x}}{\gamma(v,\, \omega)}\,, \tag{10.20}$$

which can be further evaluated to yield

$$\Delta \mathrm{x}' = \Delta \mathrm{x} \sqrt{1 - \frac{v^2}{c^2} - l_P^2 \left(\frac{\Delta \mathrm{x} - \Delta \mathrm{x}'}{\Delta \mathrm{x}' \Delta \mathrm{x}} \right)^2}\,. \tag{10.21}$$

Although nonlinear, this expression evidently reduces to the special relativistic expression for length contraction in the limit of vanishing Planck length. For the physically interesting case of $\Delta \mathrm{x}' \ll \Delta \mathrm{x}$, it can be simplified and solved exactly, yielding the "linearized" expression for the "contracted" length,

$$\Delta \mathrm{x}' = \Delta \mathrm{x} \sqrt{\frac{1}{2}\left(1 - \frac{v^2}{c^2}\right) + \sqrt{\frac{1}{4}\left(1 - \frac{v^2}{c^2}\right)^2 - \frac{l_P^2}{(\Delta \mathrm{x})^2}}}\,, \tag{10.22}$$

provided the reality condition

$$\frac{1}{4}\left(1-\frac{v^2}{c^2}\right)^2 \;\geq\; \frac{l_P^2}{(\Delta \mathrm{x})^2} \tag{10.23}$$

is satisfied. Substituting this condition back into the solution of Eq. (10.22) then gives

$$\Delta \mathrm{x}' \;\geq\; \sqrt{l_P \Delta \mathrm{x}}\,, \tag{10.24}$$

which implies that as long as $\Delta \mathrm{x}$ remains greater than l_P the "contracted" length $\Delta \mathrm{x}'$ also remains greater than l_P, in close analogy with the invariant bound c on speeds in special relativity. That is to say, in addition to the upper bound $\Delta \mathrm{x}$ on lengths implied by the condition $\gamma(v,\,\omega) > 1$ above, the "contracted" length $\Delta \mathrm{x}'$ also remains invariantly bounded from below, by l_P:

$$\Delta \mathrm{x} \;>\; \Delta \mathrm{x}' \;>\; l_P\,. \tag{10.25}$$

Starting again from the expression for time dilation analogous to that for the length contraction,

$$\Delta\tau = \frac{\Delta t}{\gamma(v,\,\omega)}\,, \tag{10.26}$$

and using almost identical line of arguments as above, one analogously arrives at a generalized expression for the time dilation,

$$\Delta\tau = \Delta t\sqrt{\frac{1}{2}\left(1-\frac{v^2}{c^2}\right)+\sqrt{\frac{1}{4}\left(1-\frac{v^2}{c^2}\right)^2-\frac{t_P^2}{(\Delta t)^2}}}\,, \tag{10.27}$$

together with the corresponding invariant bounds on the "dilated" time:

$$\Delta t \;>\; \Delta\tau \;>\; t_P\,. \tag{10.28}$$

Thus, in addition to being bounded from above by the time Δt, the "dilated" time $\Delta\tau$ remains invariantly bounded also from below, by the Planck time t_P.

So far we have not assumed or proved explicitly that the constant "c" is an upper bound on possible speeds. As alluded to above, in the present theory the observer-independence of the upper bound c turns out to be a derivative notion. This can be easily appreciated by considering the ratio of the "contracted" length of Eq. (10.22) and "dilated" time of Eq. (10.27), along with the definitions

$$u := \frac{\Delta \mathrm{x}}{\Delta t} \quad \text{and} \quad u' := \frac{\Delta \mathrm{x}'}{\Delta\tau} \tag{10.29}$$

for velocities, leading to the upper bound on velocities in the moving frame:

$$u' \leq u\sqrt{1+\sqrt{1-c^2 u^{-2}}}\,. \tag{10.30}$$

Hence, as long as u does not exceed c, u' also remains within c. In other words, in the present theory c retains its usual status of the observer-independent upper bound on causally admissible speeds, but in a rather derivative manner.

In addition to the above kinematical implications, the basic elements of the particle physics are also modified within our generalized theory, the central among which being the Planck-scale ameliorated dispersion relation

$$p^2 c^2 + m^2 c^4 = E^2 \left[1 - \frac{\left(E - mc^2\right)^2}{E_P^2} \right], \tag{10.31}$$

where E_P is the Planck energy. It is worth emphasizing here that this is an *exact* relation between energies and momenta, which in the rest frame of the massive particle reproduces Einstein's famous mass-energy equivalence:

$$E = mc^2. \tag{10.32}$$

Moreover, in analogy with the invariant lower bounds on lengths and durations we discussed above, in the present theory energies and momenta can also be shown to remain invariantly bounded from above by their Planck values:

$$E' \leq \sqrt{E_P E} \quad \text{and} \quad p' \leq \sqrt{k_P\, p} \times \sqrt{\frac{v}{c}}\,, \tag{10.33}$$

where k_P is the Planck momentum. Thus, as long as the unprimed energy E does not exceed E_P, the primed energy E' also remains within E_P. That is to say, in addition to the lower bound E on energies implied by the condition $\gamma(v,\,\omega) > 1$, the energies E' remain invariantly bounded also from above, by the Planck energy E_P:

$$E < E' < E_P\,. \tag{10.34}$$

Similarly, as long as the relative velocity v does not exceed c and the unprimed momentum p does not exceed k_P, the primed momentum p' also remains within k_P. Hence, in addition to the lower bound p on momenta set by the condition $\gamma(v,\,\omega) > 1$, the momenta p' remain invariantly bounded also from above, by the Planck momentum k_P:

$$p < p' < k_P\,. \tag{10.35}$$

Thus, unlike in special relativity, in the present theory *all* physical quantities remain invariantly bounded by their respective Planck-scale values.

Next, consider an isolated system of mass m_{sys} composed of a number of constituents undergoing an internal reaction. It follows from the quadratic invariant in Eq. (10.17) that the $4+2n-$vector $\mathcal{P}_{sys}$, defined as the abstract momentum of the system as a whole, would be conserved in such a reaction (cf. [8]),

$$\Delta \mathcal{P}_{sys} = 0\,, \tag{10.36}$$

where Δ denotes the difference between initial and final states of the reaction, and $\mathcal{P}_{sys}$ is defined by

$$m_{sys}\,\frac{dz^A}{d\tau} =: \mathcal{P}^A_{sys} := \left(E_{sys}/c\,,\; p^k_{sys}\,,\; P^\mu_{sys}\right), \tag{10.37}$$

with $k = 1,2,3$ denoting the external 3D and $\mu = 4,5,\ldots,3+2n$ denoting the phase space dimensions of the system as a whole. It is clear from this definition that, since dz^A is a $4+2n-$vector whereas m_{sys} and $d\tau$ are invariants, $\mathcal{P}^A_{sys}$ is also a $4+2n-$vector, and hence, transforms under Eq. (10.18) as

$$\mathcal{P}'^A_{sys} = \Lambda^A{}_B\,\mathcal{P}^B_{sys}\,. \tag{10.38}$$

Moreover, since Λ depends only on the overall coordinate transformations being performed within the space $\mathcal{E}$, the difference on the left hand side of Eq. (10.36) is also a $4+2n-$vector, and therefore transforms as

$$\Delta\mathcal{P}'^A_{sys} = \Lambda^A{}_B\,\Delta\mathcal{P}^B_{sys}\,. \tag{10.39}$$

Thus, if the conservation law of Eq. (10.36) holds for one set of coordinates within the space $\mathcal{E}$, then, according to Eq. (10.39), it does so for *all* coordinates related by the transformations of Eq. (10.18). Consequently, the conservation law in Eq. (10.36), once unpacked into its external, internal, and constituent parts as

$$0 = \Delta\mathcal{P}_{sys} = \left(\Delta E_{sys}/c\,,\; \Delta\mathbf{p}_{sys}\,,\; \Delta\mathbf{P}^{int}_{sys}\right), \tag{10.40}$$

leads to the familiar conservation laws for energies and momenta:

$$0 = \Delta E_{sys} := \sum_f E_f - \sum_i E_i \tag{10.41}$$

and

$$0 = \Delta\mathbf{p}_{sys} := \sum_f \mathbf{p}_f - \sum_i \mathbf{p}_i\,, \tag{10.42}$$

where the indices f and i stand for the final and initial number of constituents of the system. Thus, in the present theory the energies and momenta remain as *additive* as in special relativity. In other words, in the present theory not only are there no preferred class of observers, but also the usual conservation laws of special relativity remain essentially unchanged, contrary to expectation.

10.3.3 The Raison D'être of Time: Causal Inevitability of Becoming

With the physical structure of the generalized relativity in place, we are now in a position to address the central concern of the present essay: namely, the *raison d'être* of the *tensed* time, as depicted in Fig. 10.2. To this end, let us first note that the causal structure embedded within our generalized relativity is profoundly unorthodox. One way to appreciate this unorthodoxy is to recall the blurb for space-time put forward by Minkowski in his seminal address at Cologne, in 1908. "Nobody has ever noticed a place except at a time, or a time except at a place", he ventured [38]. But, surely, this famous quip of Minkowski hardly captures the complete picture. Perhaps it is more accurate to say something like: *Nobody has ever noticed a place except at a certain time while being in a certain state, or noticed a time except at a certain place while being in a certain state, or been in some state except at a certain time, and a certain place.* At any rate, this revised statement is what better captures the notion of time afforded by our generalized theory of relativity. For, as evident from the quadratic invariant of Eq. (10.14), in addition to space, time in our generalized theory is as much a state-dependent attribute as states are time-dependent attributes, and as states of the world do happen and become, so does time. Intuitively, this dynamic state of affairs can be summarized as follows:

$$\begin{aligned} \mathrm{x} &= \mathrm{x}(t,\, \mathrm{y}) \\ t &= t(\mathrm{x},\, \mathrm{y}) \\ \mathrm{y} &= \mathrm{y}(t,\, \mathrm{x}), \end{aligned} \tag{10.43}$$

where y is the phase space coordinate as before. In other words, place in the present theory is regarded as a function of time and state; time is regarded as a function of place and state; and state is regarded as a function of time and place. As we shall see, it is this state-dependence of time that is essentially what mandates the causal necessity for becoming in the present theory.

To appreciate this dynamic or tensed nature of time in the present theory, let us return once again to our clock that is moving and evolving, say, from an event-state $(\mathrm{e}_1,\, \mathrm{s}_1)$ to an event-state $(\mathrm{e}_5,\, \mathrm{s}_5)$ in the combined space $\mathcal{E}$, as depicted in the space-time-state[5] diagram of Fig. 10.7. According to the line element of Eq. (10.14), the proper duration recorded by the clock would be given by

$$\Delta\tau = \int_{\mathrm{t}_1}^{\mathrm{t}_5} \frac{1}{\gamma(v,\, \omega)}\; dt\,, \tag{10.44}$$

where $\gamma(v,\, \omega)$ is defined by Eq. (10.16). Now, assuming for simplicity that the clock is not massless, we can represent its journey by the integral curve of a timelike $4+2n-$velocity vector field V^A on the space $\mathcal{E}$, defined by

$$V^A := l_P \frac{dz^A}{d\tau}\,, \tag{10.45}$$

such that its external components V^a $(a = 0, 1, 2, 3)$ would trace out, for each possible state s_i of the clock, the familiar 4D timelike world lines in the corresponding Minkowski space-time. In other words, the overall velocity vector field V^A would give rise to the familiar timelike, future-directed, never vanishing, 4-velocity vector field V^a, tangent to each of the external timelike world lines. As a result, the "length" of the overall enveloping world line would be given by the proper duration

[5] Here perhaps "space-time-phase space diagram" would be a much more accurate neologism, but it would be even more mouthful than "space-time-state diagram."

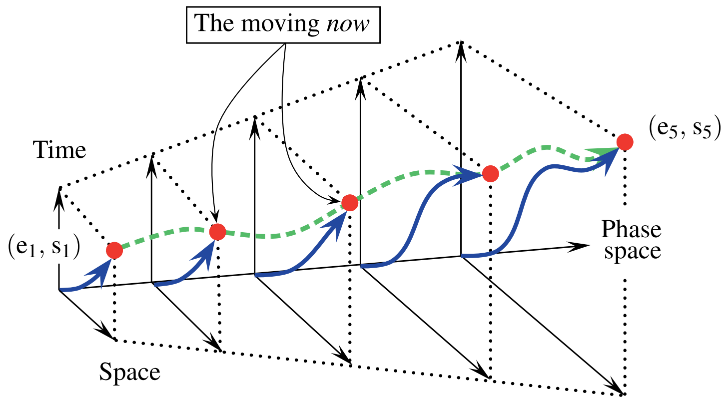

Fig. 10.7 Space-time-state diagram depicting the flow of time. The solid blue curves represent growing timelike world lines at five successive stages of growth, from s_1 to s_5, whereas the dashed green curve represents the growing overall world line from (e_1, s_1) to (e_5, s_5). The red dot represents the necessarily moving present. In fact, the entire space-time-state structure is causally necessitated to expand continuously.

in Eq. (10.44), whereas the "length" of the external worldline, for each s_i, would be given by the Einsteinian proper duration

$$\Delta\tau_E^i = \int_{t_1}^{t_i} \frac{1}{\gamma(v)}\, dt\,, \tag{10.46}$$

where $\gamma(v)$ is the usual Lorentz factor given by Eq. (10.3). In Fig. 10.7, five of such external timelike world lines – one for each s_i ($i = 1,2,3,4,5$) – are depicted by the blue curves with arrowheads going "upwards", and the overall enveloping worldline traced out by V^A is depicted by the dashed green curve going from the "initial" event-state (e_1, s_1) to the "final" event-state (e_5, s_5).

It is perhaps already clear from this picture that the external world line of our clock is not given all at once, stretched out till eternity, but grows continuously, along with each temporally successive stage of the evolution of the clock, like a tendril on a wall. That is to say, as anticipated in Fig. 10.2, the future events along the external world line of the clock simply do not exist. Hence the "now" of the clock cannot even be said to be preceding the future events, since, quite literally, there exists nothing to which it has the relation of precedence [27]. Moreover, since the external Minkowski space-time is simply a congruence of nonintersecting timelike world lines of idealized observers, according to the present theory the entire sum total of existence must increase continuously [27]. In fact, this continuous growth of existence turns out to be *causally necessitated* in the present theory, and can be represented by a *Growth Vector Field* quantifying the instantaneous directional rate of this growth:

$$U^a := \hat{V}^a \, \frac{d\tau_{_E}}{d\text{y}} \,, \tag{10.47}$$

where $\hat{V}^a$ is a unit vector field in the direction of the 4-velocity vector field V^a, $d\text{y} := |d\mathbf{y}|$ is the infinitesimal dimensionless phase-space distance between the two successive states of the clock discussed before, and $d\tau_{_E}$ is the infinitesimal Einsteinian proper duration defined by Eq. (10.1). It is crucial to note here that *in special relativity this Growth Vector Field would vanish identically everywhere, whereas in our generalized theory it cannot possibly vanish anywhere*. This is essentially because of the mutual dependence of place, time, and state in the present theory we discussed earlier (cf. Eq. (10.43)). More technically, this is because the 4-velocity vector V^a of an observer in Minkowski space-time, such as the one in Eq. (10.47), can never vanish, whereas the causality constraint in Eq. (10.14) of the present theory imposes the lower bound $t_{_P}$ on the rate of change of Einsteinian proper duration with respect to the phase-space coordinates,

$$\frac{d\tau_{_E}}{d\text{y}} \geq t_{_P} \,, \tag{10.48}$$

which, taken together, *causally* necessitates the never-vanishing of the Growth Vector Filed U^a. Consequently, the "now" of the clock (the red dot in Fig. 10.7) moves in the future direction along its external world line, at the rate of no less than one Planck unit of time per Planck unit of change in its physical state. And, along with the nonvanishing of the 4-velocity vector field V^a, the lower bound $t_{_P}$ on the growth rate of any external world line implies that not only do all such "nows" move, but they *cannot* not move – i.e. not only does the sum total of existence increase, but it *cannot* not increase. To parody Weyl quoted above, the objective world cannot simply *be*, it can only *happen*.

This conclusion can be further consolidated by realizing that in the present theory even the overall enveloping world line (the dashed green curve in Fig. 10.7) cannot help but grow non-relationally and continuously. This can be confirmed by first parallelling the above analysis for the $1+2n-$dimensional internal space $\mathcal{O}$ instead of the external space-time $\mathcal{M}$, which amounts to slicing up the combined space $\mathcal{E}$ of Fig. 10.7 along the spatial axis instead of the phase space axis, and then observing that even the "internal world line" (not shown in the figure) must necessarily grow progressively further as time passes, at the rate given by the *internal* growth vector field

$$U^\alpha = l_{_P} \, \hat{V}^\alpha \, \frac{dt_{_H}}{d\text{x}} \,. \tag{10.49}$$

Here $\hat{V}^\alpha$ is a unit vector field in the direction of the $1+2n-$velocity vector filed V^α corresponding to the internal part of the overall velocity vector field V^A, $d\text{x} := |d\mathbf{x}|$ is the infinitesimal spatial distance between two slices, and $dt_{_H}$ is the infinitesimal internal proper duration defined by Eq. (10.10). Once again, it is easy to see that the causality condition of Eq. (10.14) gives rise to the lower bound

$$l_{_P} \, \frac{dt_{_H}}{d\text{x}} \geq t_{_P} \,. \tag{10.50}$$

Thus, "now" of the clock necessarily moves in the future direction also along its internal world line within the internal space $\mathcal{O}$. As a result, even the overall world line – namely, the dashed green curve in Fig. 10.7 – can be easily shown to be growing non-relationally and continuously. Indeed, using Eqs. (10.47)–(10.50), an elementary geometrical analysis [8] shows that the instantaneous directional rate of this growth is given by the *overall* growth vector field

$$U^A = \left(\hat{V}^a \, \frac{dt_E}{d\mathrm{y}} \, , \, l_P \hat{V}^\mu \, \frac{dt_H}{d\mathrm{x}} \right), \tag{10.51}$$

whose magnitude also remains bounded from below by the Planck time t_P:

$$\sqrt{-\xi_{AB} U^A U^B} \; \geq \; t_P \,. \tag{10.52}$$

Thus, in the present theory, not only are the external events in $\mathcal{E}$ not all laid out once and for all, for all eternity, but there does not remain even an overall $4+2n$-dimensional "block" that could be used to support a "block" view of the universe. In fact, the causal necessity of the lower-bound Eq. (10.52) on the magnitude of the overall growth vector field U^A – which follows from the causality constraint of Eq. (10.14) – exhibits that in the present theory the sum total of existence itself is causally necessitated to increase continuously. That is to say, the very structure of the present theory *causally necessitates* the Universe to be purely *Heraclitean*, in the sense discussed in the Introduction.

10.4 Prospects for the Experimental Metaphysics of Time

As alluded to in the Introduction, any empirical confrontation of the above generalized relativity with special relativity would amount to a step towards what may be called the experimental metaphysics of time. However, since the generalized theory is deeply rooted in the Planck regime, any attempt to experimentally discriminate it from special relativity immediately encounters a formidable practical difficulty. To appreciate this difficulty, consider the following series expansion of expression in Eq. (10.27) for the generalized proper time, up to second order in the Planck time:

$$\Delta\tau = \Delta t \sqrt{1 - \frac{v^2}{c^2}} \, - \, \frac{1}{2} \, \frac{t_P^2}{\Delta t} \left(1 - \frac{v^2}{c^2} \right)^{-\frac{3}{2}} + \ldots \tag{10.53}$$

The first term on the right-hand side of this expansion is, of course, the familiar special relativistic term. The difficulty arises in the second term, i.e. in the first largest correction term to the special relativistic time dilation effect, since this term is modulated by the *square* of the Planck time, which in ordinary units amounts to some 10^{-87} sec^2. Clearly, the precision required to directly verify such a miniscule correction to the special relativistic prediction is well beyond the scope of any foreseeable precision technology.

Fortunately, in recent years an observational possibility has emerged that might save the day for the experimental metaphysics of time. The central idea that has emerged during the past decade within the context of quantum gravity is to counter the possible Planck-scale suppression of physical effects by appealing to ultrahigh energy particles cascading the earth that are produced at cosmological distances. One strategy along this line is to observe oscillating flavor ratios of ultrahigh energy cosmic neutrinos to detect possible deviations in the energy–momentum relations predicted by special relativity [39]. Let us briefly look at this strategy, as it is applied to our generalized theory of relativity (further details can be found in Refs. [36] and [39]; as in these references, from now on we shall be using the Planck units: $\hbar = c = G = 1$).

10.4.1 Testing Heraclitean Relativity Using Cosmic Neutrinos

The remarkable phenomena of neutrino oscillations are due to the fact that neutrinos of definite flavor states $|\nu_\alpha\rangle$, $\alpha = e, \mu$, or τ, are *not* particles of definite mass states $|\nu_j\rangle$, $j = 1, 2$, or 3, but are superpositions of the definite mass states. As a neutrino of definite flavor state propagates through vacuum for a long enough laboratory time, its heavier mass states lag behind the lighter ones, and the neutrino transforms itself into an altogether different flavor state. The probability for this "oscillation" from a given flavor state, say $|\nu_\alpha(0)\rangle$, to another flavor state, say $|\nu_\beta(t)\rangle$, is famously given by

$$P_{\alpha\beta}(E,L) = \delta_{\alpha\beta} - \sum_{j\neq k} U^*_{\alpha j} U_{\alpha k} U_{\beta j} U^*_{\beta k} \left[1 - e^{-i(\Delta m^2_{jk}/2E)L}\right]. \tag{10.54}$$

Here $\Delta m^2_{jk} \equiv m^2_k - m^2_j > 0$ is the difference in the squares of the two neutrino masses, U is the time-independent leptonic mixing matrix, and E and L are, respectively, the energy and distance of propagation of the neutrinos. It is clear from this transition probability that the experimental observability of the flavor oscillations is dependent on the quantum phase

$$\Phi := 2\pi \frac{L}{L_O}, \tag{10.55}$$

where

$$L_O := \frac{2\pi}{\Delta p} = \frac{4\pi E}{\Delta m^2_{jk}} \tag{10.56}$$

is the energy-dependent oscillation length. Thus, changes in neutrino flavors would be observable whenever the propagation distance L is of the order of the oscillation length L_O. However, in definition Eq. (10.56) the difference in momenta, $\Delta p \equiv p_j - p_k$, was obtained by using the special relativistic relation

$$p_j = \sqrt{E^2 - m^2_j} \approx E - \frac{m^2_j}{2E}. \tag{10.57}$$

In the present theory this relation between energies and momenta is, of course, generalized, and given by Eq. (10.31), replacing the above approximation by

$$p_j \approx E - \frac{m_j^2}{2E} + \frac{E^2}{m_P^2} m_j \tag{10.58}$$

up to the second order, with m_P being the Planck mass. The corresponding modified oscillation length analogous to Eq. (10.56) is then given by

$$L_O^{'} := \frac{2\pi}{\Delta p} = \frac{2\pi}{\frac{1}{2E}\,\Delta m_{jk}^2 - \frac{E^2}{m_P^2}\,\Delta m_{jk}}\,, \tag{10.59}$$

where $\Delta m_{jk}^2 \equiv m_k^2 - m_j^2$ as before, and $\Delta m_{jk} \equiv m_k - m_j > 0$. Consequently, according to our generalized relativity the transition probability of Eq. (10.54) would be quite different in general, as a function of E and L, from how it is according to special relativity. And despite the *quadratic* Planck-energy suppression of the correction to the oscillation length, this difference would be observable for neutrinos of sufficiently high energies and long propagation distances. Indeed, it can be easily shown [39] that the relation

$$L \sim \frac{\pi m_P^4}{E^5} \tag{10.60}$$

is the necessary constraint between the neutrino energy E and the propagation distance L for the observability of possible deviations from the standard flavor oscillations. For instance, it can be readily calculated from this constraint that the Planck-scale deviations in the oscillation length predicted by our generalized relativity would be either observable, or can be ruled out, for neutrinos of energy $E \sim 10^{17}$ eV, provided that they have originated from a cosmic source located at some 10^5 light years away from a terrestrial detector. The practical means by which this can be achieved in the foreseeable future have been discussed in some detail in the Refs. [36] and [39] cited above.

10.4.2 Testing Heraclitean Relativity Using γ-ray Binary Pulsars

The previous method of confronting the generalized theory of relativity with special relativity is clearly phenomenological. Fortunately, a much more direct test of the generalized theory may be possible, thanks to the precise deviations it predicts from the special relativistic Doppler shifts [8]:

$$\frac{E'}{E} = \frac{\varepsilon' \left[\, \varepsilon' - \frac{v}{c}\cos\phi \,\right]}{\sqrt{(\varepsilon')^2 - \frac{v^2}{c^2}}}\,, \tag{10.61}$$

with

$$\varepsilon' := \sqrt{1 - \frac{E^2}{E_P^2}\left(1 - \frac{E'}{E}\right)^2}\,. \tag{10.62}$$

Here v is the relative speed of a receiver receding from a photon source, E and E', respectively, are the energy of the photon and that observed by the receiver, and ϕ is the angle between the velocity of the receiver and the photon momentum. Note that ε' here clearly reduces to unity for $E' - E \ll E_P$, thus reducing the generalized expression of Eq. (10.61) to the familiar *linear* relation for Doppler shifts predicted by special relativity.

Even without solving the relation in Eq. (10.61) for E' in terms of E, it is not difficult to see that, since $\varepsilon' < 1$, at sufficiently high energies any red-shifted photons would be somewhat more red-shifted according to Eq. (10.61) than predicted by special relativity. But one can do better than that. A Maclaurin expansion of the right hand side of Eq. (10.61) around the value $E/E_P = 0$, after keeping terms only up to the second order in the ratio E/E_P, gives

$$\frac{E'}{E} \approx \frac{1-\frac{v}{c}\cos\phi}{\sqrt{1-\frac{v^2}{c^2}}} + \frac{1}{2}\frac{E^2}{E_P^2}\left[\frac{1-\frac{v}{c}\cos\phi}{\left(1-\frac{v^2}{c^2}\right)^{3/2}} - \frac{2-\frac{v}{c}\cos\phi}{\left(1-\frac{v^2}{c^2}\right)^{1/2}}\right]\left(1-\frac{E'}{E}\right)^2 + \ldots \tag{10.63}$$

This truncation is an excellent approximation to Eq. (10.61). The quadratic equation of Eq. (10.63) can now be solved for the desired ratio E'/E, and then the physical root once again expanded, now in the powers of v/c. In what results if we again keep terms only up to the second order in the ratios E/E_P and v/c, then, after some tedious but straightforward algebra, we arrive at

$$\frac{E'}{E} \approx 1 - \frac{v}{c}\cos\phi + \frac{1}{2}\left[1 - \frac{E^2}{E_P^2}\cos^2\phi\right]\frac{v^2}{c^2} \pm \ldots, \tag{10.64}$$

which, in the limit $E \ll E_P$, reduces to the special relativistic result

$$\frac{E'}{E} \approx 1 - \frac{v}{c}\cos\phi + \frac{1}{2}\frac{v^2}{c^2} \pm \ldots. \tag{10.65}$$

Comparing Eqs. (10.64) and (10.65) we see that up to the first order in v/c there is no difference between the special relativistic result and that of the present theory. The first deviation between the two theories occur in the second-order coefficient, precisely where special relativity differs also from the classical theory. What is more, this second-order deviation depends non-trivially on the angle between the relative velocity and photon momentum. For instance, up to the second order both red-shifts ($\phi = 0$) and blue-shifts ($\phi = \pi$) predicted by Eq. (10.64) significantly differ from those predicted by special relativity. In particular, the red-shifts are now somewhat more red-shifted, whereas the blue-shifts are somewhat less blue-shifted. On the other hand, the transverse red-shifts ($\phi = \pi/2$ or $\phi = 3\pi/2$) remain identical to those predicted by special relativity. As a result, even for the photon energy approaching the Planck energy an Ives–Stilwell type classic experiment [40] would not be able to distinguish the predictions of the present theory from those of special relativity. The complete angular distribution of the second-order coefficient predicted by the two theories, along with its energy dependence, is displayed in Fig. 10.8.

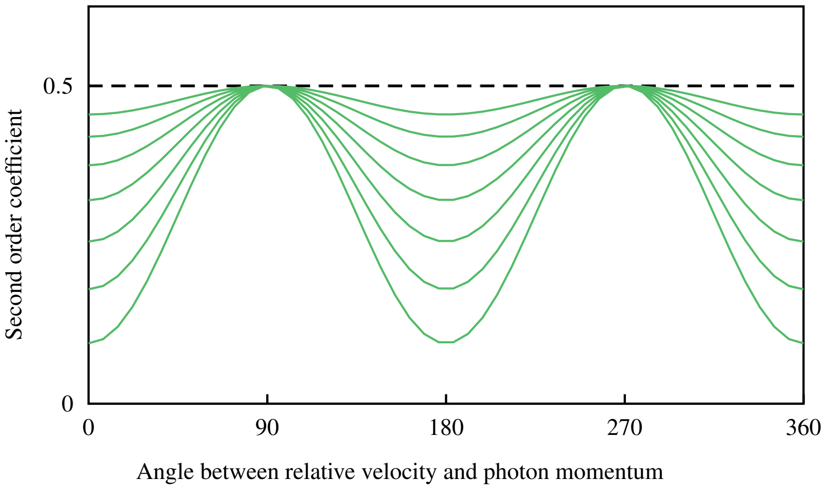

Fig. 10.8 The energy-dependent signatures of Heraclitean relativity. The green curves are based on the predictions of the present theory, for $E/E_P = 0.3$–0.9 in the descending order, whereas the dashed black line is the prediction of special relativity.

In spite of this rather nontrivial angular dependence of Doppler shifts, in practice, due to the quadratic suppression by Planck energy, distinguishing the expansion in Eq. (10.64) from its special relativistic counterpart in Eq. (10.65) would be a formidable task. The maximum laboratory energy available to us is of the order of 10^{12} eV, yielding $E^2/E_P^2 \sim 10^{-32}$. This represents a correction of one part in 10^{32} from Eq. (10.65), demanding a phenomenal sensitivity of detection well beyond the means of any foreseeable precision technology. However, an extraterrestrial source, such as an extreme energy γ-ray binary pulsar, may turn out to be accessible for distinguishing the second-order Doppler shifts predicted by the two theories. It is well known that binary pulsars not only exhibit Doppler shifts, but the second-order shifts resulting from the periodic motion of such a pulsar about its companion can be isolated, say, from the first order shifts, because they depend on the *square* of the relative velocity, which varies as the pulsar moves along its two-body elliptical orbit [41]. Due to these Doppler shifts, the rate at which the pulses are observed on Earth reduces slightly when the pulsar is receding away from the Earth, compared to when it is approaching towards it. As a result, the period, its variations, and other orbital characteristics of the pulsar, as they are determined on Earth, crucially depend on these Doppler shifts. In practice, the parameter relevant in the arrival-time analysis of the pulses received on Earth turns out to be a nontrivial function of the gravitational red-shift, the masses of the two binary stars, and other Keplerian parameters of their orbits, and is variously referred to as the red-shift-Doppler parameter or the time dilation parameter [41]. For a pulsar that is also following a periastron precession similar to the perihelion advance of Mercury, it can be determined with excellent precision.

The arrival-time analysis of the pulses begins by considering the time of emission of the Nth pulse, which is given by

$$N = N_o + \nu\tau + \frac{1}{2}\dot{\nu}\tau^2 + \frac{1}{6}\ddot{\nu}\tau^3 + \dots\,, \tag{10.66}$$

where N_o is an arbitrary integer, τ is the proper time measured by a clock in an inertial frame on the surface of the pulsar, and ν is the rotation frequency of the pulsar, with $\dot{\nu} \equiv d\nu/d\tau|_{\tau=0}$ and $\ddot{\nu} \equiv d^2\nu/d\tau^2|_{\tau=0}$. The proper time τ is related to the coordinate time t by

$$d\tau = dt\left[1 - \frac{\alpha_2^* m_2}{r} - \frac{1}{2}\frac{v_1^2}{c^2} + \dots\right], \tag{10.67}$$

where the first correction term represents the gravitational red-shift due to the field of the companion, and the second correction term represents the above mentioned second-order Doppler shift due to the orbital motion of the pulsar itself. The time of arrival of the pulses on Earth differs from the coordinate time t taken by the signal to travel from the pulsar to the barycenter of the solar system, due to the geometrical intricacies of the pulsar binary and the solar system [41]. More relevantly for our purposes, the time of arrival of the pulses is directly affected by the second-order Doppler shift appearing in Eq. (10.67), which thereby affects the observed orbital parameters of the pulsar.

Now, returning to our Heraclitean generalization of relativity, it is not difficult to see that the generalized Doppler shift expression in Eq. (10.64) immediately gives the following generalization of the infinitesimal proper time of Eq. (10.67):

$$d\tau = dt\left[1 - \frac{\alpha_2^* m_2}{r} - \frac{1}{2}\left\{1 - \frac{E^2}{E_P^2}\cos^2\phi\right\}\frac{v_1^2}{c^2} + \dots\right]. \tag{10.68}$$

Thus, in our generalized theory the second-order Doppler shift acquires an *energy-dependent* modification. The question then is: At what radiation energy this non-trivial modification will begin to affect the observable parameters of the pulsar? The most famous pulsar, namely, PSR B1913+16, which has been monitored for three decades with exquisite accumulation of timing data, is a radio pulsar, and hence, for it the energy-dependent modification predicted in Eq. (10.68) is utterly negligible, thanks to the quadratic suppression by the Planck energy. However, for a γ-ray pulsar with sufficiently high radiation energy the modification predicted in Eq. (10.68) should have an impact on its observable parameters, such as the orbital period and its temporal variations.

The overall precision in the timing of the pulses from PSR B1913+16, and consequently in the determination of its orbital period, is famously better than one part in 10^{14} [42]. Indeed, the monitoring of the decaying orbit of PSR B1913+16 constitutes one of the most stringent tests of general relativity to date. It is therefore not inconceivable that similar careful observations of a suitable γ-ray pulsar may be able to distinguish the predictions of the present theory from those of special relativity. Unfortunately, the highest energy of radiation from a pulsar known to date happens to be no greater than 10^{13} eV, giving the discriminating ratio E^2/E_P^2 to be of the order of 10^{-30}, which is only two orders of magnitude improvement over a possible terrestrial scenario. On the other hand, the γ-rays emitted by a binary pulsar would have to be of energies exceeding 10^{21} eV for them to have desired observable consequences, comparable to those of PSR B1913+16. Moreover, the desired pulsar have to be located sufficiently nearby, since above the 10^{13} eV threshold γ-rays are expected to attenuate severely through pair-production if they are forced to pass through the cosmic infrared background before reaching the Earth. It is not inconceivable, however, that a suitable binary pulsar emitting radiation of energies exceeding 10^{21} eV is found in the near future, allowing experimental discrimination of our generalized relativity from special relativity.

10.5 Concluding Remarks

One of the perennial problems in natural philosophy is the problem of change; namely, *How is change possible?* Over the centuries, this problem has fostered two diametrically opposing views of time and becoming. While these two views tend to agree that time presupposes change, and that genuine change requires becoming, one of them actually denies the reality of change and time, by rejecting becoming as a "stubbornly persistent illusion" [43]. The other view, by contrast, accepts the reality of change and time, by embracing becoming as a *bona fide* attribute of the world. Since the days of Aristotle within physics we have been rather successful in explaining *how* the changes occur in the world, but seem to remain oblivious to the deeper question of *why* do they occur at all. The situation has been aggravated by the advent of Einstein's theories of space and time, since in these theories there is no room to *structurally* accommodate the distinction between the past and the future – a prerequisite for the genuine onset of change. By contrast, the causal structure of the Heraclitean relativity discussed above not only naturally distinguishes the past form the future by causally necessitating becoming, but also *forbids* inaction altogether, thereby providing an answer to the deeper question of change. Moreover, since it is not impossible to experimentally distinguish the Heraclitean relativity from special relativity, and since the ontology underlying only the latter of these two relativities is prone to a block universe interpretation, the enterprise of *experimental metaphysics of time* becomes feasible now, for the first time, within a relativistic context. At the very least, such an enterprise should help us decide whether time is best understood relationally, or non-relationally.

Acknowledgments I would like to thank Huw Price and Abner Shimony for their comments on Ref. [8], of which this essay is an apologia. I would also like to thank Lucien Hardy, Lee Smolin, Antony Valentini, and other members of the Foundations of Physics group at the Perimeter Institute for their hospitality and support.

References

1. K.R. Popper: *The World of Parmenides: Essays on the Presocratic Enlightenment*, A.F. Petersen (ed.) (Routledge, New York 2001)
2. J.M.E. McTaggart: Mind (New Series) **17**, 68, 457 (1908)
3. Q. Smith: General Introduction: The Implications of the Tensed and Tenseless Theories of Time. In: *The New Theory of Time*, L.N. Oaklander and Q. Smith (eds) (Yale University Press, New Haven, CT 1994) p 1
4. I. Newton: *Mathematical Principles of Natural Philosophy*, translated by A. Motte, revised and annotated by F. Cajori (University of California Press, Berkeley, CA 1962) Vol. I, p 6
5. A.R. Hall and M.B. Hall (eds): *Unpublished Scientific Papers of Isaac Newton* (Cambridge University Press, Cambridge 1962) p 103
6. R.T.W. Arthur: Studies in History and Philosophy of Science **26**, 2, 323 (1995)
7. G.J. Whitrow: *The Natural Philosophy of Time*, Second Edition (Oxford University Press, Oxford 1980) p 41
8. J. Christian: International Journal of Modern Physics D **13**, 1037 (2004); see also J. Christian: [arXiv:gr-qc/0308028]
9. A. Shimony: Search for a Worldview which can Accommodate Our Knowledge of Microphysics. In: *Search for a Naturalistic World View*, Vol. I (Cambridge University Press, Cambridge 1993) p 64
10. A. Shimony: Experimental Test of Local Hidden-Variable Theories. In: *Search for a Naturalistic World View*, Vol. II (Cambridge University Press, Cambridge 1993) p 77
11. G.N. Schlesinger: *Aspects of Time* (Hackett, Indianapolis, IN 1980) p 24
12. P.C.W. Davies: *The Physics of Time Asymmetry* (Surrey University Press, 450 Edgware Road, London 1974) p 12
13. R. Penrose: *The Emperor's New Mind* (Oxford University Press, Oxford 1989) p 303
14. É. Meyerson: *The Relativistic Deduction: Epistemological Implications of the Theory of Relativity*, D.A. and M-A. Sipfle (trans.) (D. Reidel, Dordrecht, The Netherlands 1985) p xl
15. H. Weyl: *Philosophy of Mathematics and Natural Science* (Princeton University Press, Princeton 1949) p 116
16. M. Black: Scientific American **206**, 4, 179 (April 1962) p 181
17. H. Putnam: The Journal of Philosophy **64**, 240 (1967)
18. C.W. Rietdijk: Philosophy of Science **33**, 341 (1966)
19. N. Maxwell: Philosophy of Science **52**, 23 (1985)
20. H. Stein: The Journal of Philosophy **65**, 5 (1968)
21. H. Stein: Philosophy of Science **58**, 147 (1991)
22. A. Shimony: The Transient *Now*. In: *Search for a Naturalistic World View*, Vol. II (Cambridge University Press, Cambridge 1993) p 271
23. R. Clifton and M. Hogarth: Synthese **103**, 355 (1995)
24. K. Gödel: A Remark about the Relationship between Relativity Theory and Idealistic Philosophy. In: *Albert Einstein: Philosopher-Scientist*, P.A. Schilpp (ed.) (Open Court, La Salle, IL 1949) p 558
25. C. Callender: British Journal for the Philosophy of Science **48**, 117 (1997)
26. R.P. Feynman, R.B. Leighton, and M. Sands: *The Feynman Lectures on Physics* (Addison-Wesley, London 1963) p 15–17
27. C.D. Broad: *Scientific Thought* (Kegan Paul, London 1923) pp 66
28. A. Shimony: Implications of Transience for Spacetime Structure. In: *The Geometric Universe: Science, Geometry, and the Work of Roger Penrose*, S.A. Huggett et al. (eds) (Oxford University Press, Oxford 1998) p 161
29. A. Elitzur and S. Dolev: Foundations of Physics Letters **12**, 309 (1999)
30. H. Primas: Mind and Matter **1**, 81 (2003)
31. J.B. Hartle: American Journal of Physics **73**, 2, 101 (2005)
32. G.F.R. Ellis: Physics in the Real Universe: Time and Spacetime. In this volume

33. D.P. Rideout and R.D. Sorkin: Physical Review D **61**, 024002 (2000); See also R.D. Sorkin: In this volume
34. L. Smolin: The case for background independence [arXiv:hep-th/0507235]
35. J. Christian: Why the Quantum Must Yield to Gravity. In: *Physics Meets Philosophy at the Planck Scale*, C. Callender and N. Huggett (eds) (Cambridge University Press, Cambridge 2001) p 305
36. J. Christian: Physical Review Letters **95**, 160403 (2005)
37. A. Einstein: Ether and the Theory of Relativity. In: *The Collected Papers of Albert Einstein*, Vol. 7, Alfred Engel (trans.) (Princeton University Press, Princeton 2002) p 161
38. H. Minkowski: Space and Time. In: *The Principle of Relativity*, W. Perrett and G.B. Jeffery (trans.) (Dover, New York 1952) p 75
39. J. Christian: Physical Review D **71**, 024012 (2005)
40. H.E. Ives and G.R. Stilwell: Journal of the Optical Society of America **28**, 215 (1938).
41. C.M. Will: *Theory and Experiment in Gravitational Physics*, Revised Edition (Cambridge University Press, Cambridge 1993) pp 283
42. C.M. Will: Living Reviews in Relativity **9**, 3 (2006) [Online article, http://www.livingreviews.org//lrr-2006-3] pp 58
43. A. Einstein: letter of condolence to V. Besso. In: *Albert Einstein–Michele Besso Correspondence*, P. Speziali (ed.) (Hermann, Paris 1972) p 537

Chapter 11
An Argument for 4D Block World from a Geometric Interpretation of Nonrelativistic Quantum Mechanics

Michael Silberstein, W.M. Stuckey, and Michael Cifone

11.1 Introduction

We use a new, distinctly "geometrical" interpretation of nonrelativistic quantum mechanics (NRQM) to argue for the fundamentality of the 4D block world ontology. We argue for a geometrical interpretation whose fundamental ontology is one of space-time relations as opposed to constructive entities whose time-dependent behavior is governed by dynamical laws. Our view rests on two formal results: Kaiser (1981, 1990), Bohr and Ulfbeck (1995), and Anandan (2003) showed independently that the Heisenberg commutation relations of NRQM follow from the relativity of simultaneity (RoS) per the Poincaré Lie algebra. And, Bohr, Ulfbeck, and Mottelson (2004a, 2004b) showed that the density matrix for a particular NRQM experimental outcome may be obtained from the space-time symmetry group of the experimental configuration. This shows how the block world view is not only consistent with NRQM, not only an implication of our geometrical interpretation of NRQM, but it is necessary in a nontrivial way for explaining quantum interference and "non-locality" from the space-time perspective. Together the formal results imply that contrary to accepted wisdom, NRQM, the measurement problem and so-called quantum nonlocality do not provide reasons to abandon the 4D block world implication of RoS. Rather, the deep noncommutative structure of the quantum and the deep structure of space-time as given by the Minkowski interpretation of special relativity (STR) are deeply unified in a 4D space-time regime that lies between Galilean space-time (G4) and Minkowski space-time (M4).

Taken together the aforementioned formal results allow us to model NRQM phenomena such as interference without the need for realism about 3N Hilbert space, establishing that the world is really 4D and that configuration space is nothing more than a calculational device. Our new geometrical interpretation of NRQM provides a geometric account of quantum entanglement and so-called nonlocality free of conflict with STR and free of interpretative mystery.

V. Petkov (ed.), *Relativity and the Dimensionality of the World*, 197–216.

In section 11.2 we discuss the various tensions between STR and NRQM with respect to the dimensionality of the world. Section 11.3 is devoted to an explication of the Kaiser et al. results and their philosophical implications. Likewise, the Bohr et al. results and their implications are the subject of section 11.4. In section 11.5, we present our geometric resolution of the measurement problem and interpretation of quantum entanglement and "non-locality."

11.2 Motivating the Geometric Interpretation: STR versus NRQM on the Dimensionality of the World

In relativity theory, we have two physical postulates (relativity and light postulates) and we have a geometric model or "interpretation" of those postulates – Minkowski's hyperbolic 4-geometry that gives us a geometry of "light-cones." The "blockworld" (BW) view tries to establish a metaphysical interpretation of the Minkowski geometrical rendition of special relativity. It is a view that tries to establish the reality of all space-time events (*contra* presentism), whose structure is given by the special relativistic metric. We shall not rehearse the familiar arguments for the BW implication from the relativity of simultaneity (see Stuckey et al. 2006), but only describe it herein:

> There is no dynamics within space-time itself: nothing ever moves therein; nothing happens; nothing changes. In particular, one does not think of particles as moving through space-time, or as following along their world-lines. Rather, particles are just in space-time, once and for all, and the world-line represents, all at once, the complete life history of the particle. Robert Geroch, *General Relativity from A to B*. (University of Chicago Press, Chicago, 1978) p. 20–21

When Geroch says there is no dynamics within space-time itself, he is not denying that the mosaic of the BW possesses patterns that can be described with dynamical laws. Nor is he denying the predictive and explanatory value of such laws. Rather his point is that in a BW (given the reality of all events) dynamics such as Schr*ö*dinger dynamics are not event factories that bring heretofore nonexistent events (such as measurement outcomes) into being. Dynamical laws are not brute unexplained explainers that "produce" events. Geroch is advocating for what philosophers call Humeanism about laws. Namely, the claim is that dynamical laws are *descriptions* of regularities and not the brute explanation for such regularities. His point is that in a BW, Humeanism about laws is an obvious position to take because everything is just "there."

Some have actually suggested that we ought to take the fact of BW seriously when doing physics and modeling reality. Huw Price (1996) for example calls it the "Archimedean view from nowhen" (260) and it has motivated him to take seriously the idea of a time-symmetric quantum mechanics. Price is primarily concerned with seeing if one can construct a local hidden-variables interpretation of NRQM that explains so-called quantum nonlocality with purely timelike dynamics or backwards causation.

Not only is the BW strikingly at odds with NRQM dynamically conceived, but NRQM and STR appear to disagree about the very dimensionality of the world. For as David Albert says:

> [T]he space in which any realistic interpretation of quantum mechanics is necessarily going to depict the history of the world as playing itself out ... is configuration-space. And whatever impression we have to the contrary (whatever impression we have, say, of living in a three-dimensional space, or in a four-dimensional space) is somehow flatly illusory. (1996, p. 277)

Is the world a 4D Minkowski block world as relativity tells us? Or, is it a 3N-dimensional configuration space of possibly infinite dimensions as quantum mechanics tells us? How can we resolve this apparent conflict? If we assume that it is in fact a 4D BW as we do here, then what should we make of Hilbert space?

Most natural philosophers are inclined to accept that special relativity unadorned implies the block world view. Among those who might agree that special relativity unadorned implies a block world are those who think that quantum theory provides an excellent reason to so adorn it even apart from Hilbert space realism. That is, there are those who claim that quantum nonlocality or some particular solution to the measurement problem (such as collapse interpretations) require the addition of, or imply the existence of, some variety of preferred frame (a preferred foliation of space-time into space and time)[1] in order to render quantum mechanics covariant and resolve potential conflicts between observers in different frames of reference. This trick could be done in a number of ways and need not involve postulating something like the "luminiferous aether." For example, one could adopt the Newtonian or neo-Newtonian space-time of Lorentz or one could add a physically preferred foliation to M4. With a constructive theory of STR in hand one might also attempt to block the block world interpretation. As Callender notes (2006, 3):

> In my opinion, by far the best way for the tenser to respond to Putnam et al. is to adopt the Lorentz 1915 interpretation of time dilation and Fitzgerald contraction. Lorentz attributed these effects (and hence the famous null results regarding an aether) to the Lorentz invariance of the dynamical laws governing matter and radiation, not to space-time structure. On this view, Lorentz invariance is not a space-time symmetry but a dynamical symmetry, and the special relativistic effects of dilation and contraction are not purely kinematical. The background space-time is Newtonian or neo-Newtonian, not Minkowskian. Both Newtonian and neo-Newtonian space-time include a global absolute simultaneity among their invariant structures (with Newtonian space-time singling out one of neo-Newtonian space-time's many preferred inertial frames as the rest frame). On this picture, there is no relativity of simultaneity and space-time is uniquely decomposable into space and time. Nonetheless, because matter and radiation transform between different frames via the Lorentz transformations, the theory is empirically adequate. Putnam's argument has no purchase here because Lorentz invariance has no repercussions for the structure of space and time. Moreover, the theory shouldn't be viewed as a desperate attempt to save absolute simultaneity in the face of the phenomena, but it should rather be viewed as a natural extension of the well-known Lorentz invariance of the free Maxwell equations. The reason why some tensers have sought all manner of strange replacements for special relativity when this comparatively elegant theory exists is baffling.

[1] See Tooley (1997) ch. 11, for one example.

The task we have set for ourselves in this paper is to take up the charge of Archimedean physics in a way far more radical than even time-symmetric quantum mechanics suggests. Our account is a hidden-variables statistical interpretation of a sort, but unlike Price and others we are not primarily motivated by saving locality. Rather we are motivated by seeing how far we can take Archimedean physics. What follows is a purely geometric (acausal and adynamical) account of NRQM. Our view defends the surprising thesis that the relativity of simultaneity plays an essential role in the space-time regime for which one can obtain the Heisenberg commutation relations of nonrelativistic quantum mechanics – the cornerstone of quantum theory. This point bears repeating. While it is widely appreciated that special relativity and quantum theory are not necessarily incompatible, what is not widely appreciated are a collection of formal results showing that quantum theory and the relativity of simultaneity are not only compatible, but in fact are intimately related. More specifically, in the present paper we will draw on these results and clearly show that it is precisely this "nonabsolute nature of simultaneity"[2] which survives the c $\rightarrow \infty$ limit of the Poincaré group and entails the canonical commutation relations of nonrelativistic quantum mechanics. These results lead us to formulate a new geometric account of NRQM that will be elucidated in later sections of the paper.

We will also show that this geometric interpretation of NRQM nicely resolves the standard conceptual problems with the theory: (i) prior to the invocation of any dynamical interpretation of quantum theory itself and (ii) *prior* to the issue of whether any interpretation of quantum theory – i.e. a *mechanics* of the quantum – can be rendered relativistically invariant/covariant. Namely, we will provide both a geometrical account of entanglement and "non-locality" free of tribulations, and a novel version of the statistical interpretation that deflates the measurement problem. Our geometrical NRQM has the further advantage that it does not lead to the aforementioned problems that some constructive accounts of NRQM face when relativity is brought into the picture, such as Bohmian mechanics and collapse accounts like the wave function interpretation of GRW. On the contrary, not only does our view require no preferred foliation but it also provides for a profound, though little-appreciated, unity between STR and NRQM *by way of the relativity of simultaneity* [3]. Our interpretation of NRQM can be characterized as follows:

1. Realism about M4 and the BW but not Hilbert space.
2. We adopt the view that NRQM is a geometric theory in the following respects:
   (a) It merely provides a probabilistic rule by which classical objects are related in space-time – i.e. we take NRQM to provide *constraints on the distribution of events in space-time*
   (b) It is not *fundamentally* a dynamical theory of the behavior of matter-in-motion. Our ontology does not accept matter-in-motion as *fundamental* (though such a view is phenomenologically/pragmatically useful)

[2] Kaiser (1981), p. 706.

[3] In this respect, our interpretation is close to that of Bohr and Ulfbeck. In their words, "quantal physics thus emerges as but an implication of relativistic invariance, liberated from a substance to be quantized and a formalism to be interpreted" (1995, 1).

(c) quantum "entities" and their characteristic properties such as entanglement and nonlocality are geometric features of the distribution of space-time relations
(d) Spatiotemporal relations are the means by which all physical phenomena (both quantum and classical) are modeled, allowing for a natural transition from quantum to classical mechanics as simply the transition from rarefied to dense collections of space-time relations

3. We adopt an explanatory strategy that is faithful to our methodological and ontological commitments: we take the view that the determination of events, properties, experimental outcomes, etc., in space-time is made with space-time symmetries both *globally* and *acausally/adynamically*. That is, we invoke an acausal, spatiotemorally global determination relation that respects neither past nor future common cause principles. Ultimately, we believe there exists a self-consistency relationship between classical objects (modeled as sources in quantum field theory) and space-time relations (modeled by the dynamical differential operators and their inverses, i.e. the propagator in quantum physics) *a la* Einstein's equations, which may be viewed as a self-consistency relation between the space-time metric and the distribution of matter/energy via the stress–energy tensor.

Many will assume that a geometric interpretation such as ours is impossible because quantum wave functions live in Hilbert space and contain much more information than can be represented in a classical space of three dimensions. The existence of entangled quantum systems provides one obvious example of the fact that more information is contained in the structure of quantum mechanics than can be represented completely in space-time. As Peter Lewis says, "[T]he inescapable conclusion for the wavefunction realist seems to be that the world has 3N dimensions; and the immediate problem this raises is explaining how this conclusion is consistent with our experience of a three-dimensional world" (2004, 717). On the contrary, the existence of the noncommutativity of quantum mechanics is deeply related to the structure of *space-time* itself, without having to invoke the geometry of Hilbert space. Surprisingly, as will be demonstrated in the following section, it is a space-time structure for which the relativity of simultaneity is upheld, and not challenged.

11.3 The Relativity of Simultaneity and Nonrelativistic Quantum Mechanics

Lorentz boosts (changes to moving frames of reference according to the Poincaré group of STR) do not commute with spatial translations, i.e. different results obtain when the order of these two operations is reversed. Specifically, this difference is a temporal displacement which is key to generating a BW. This is distinct from Newtonian mechanics whereby time and simultaneity are absolute per Galilean invariance. If space-time was Galilean invariant, observers would agree as to which events were simultaneous and presentism could be true. In such a space-time, it would not matter if you Galilean boosted then spatially translated, or spatially translated then Galilean boosted. *Prima facie*, one might suspect that *non-relativistic* quantum mechanics would be in accord with Galilean space-time. And indeed, the linear dynamics – the Schr*ö*dinger equation – is Galilean invariant (Brown and Holland 1999). However, as we will show, while it is indeed true that the Schr*ö*dinger dynamics is Galilean invariant, the appropriate space-time structure for which one can obtain the Heisenberg commutation relations is not a Galilean space-time! Surprisingly, it is a space-time structure "between" Galilean space-time and Minkowski space-time, but one for which the relativity of simultaneity is upheld in contrast to Galilean space-time.

We can see why this obtains, heuristically, by noting that the position of operator X in NRQM is proportional to $\frac{\partial}{\partial v}$ so it generates a boost in velocity v just as the momentum operator P, proportional to $\frac{\partial}{\partial x}$, generates a spatial translation in x. This boost does not reside in M4 because NRQM operators X, Y, and Z commute while boosts along these spatial directions in M4 do not commute. Neither does this boost reside in G4 since therein spatial translations commute with boosts (per absolute simultaneity), but X and P do not commute in NRQM. Thus, we expect the space-time of NRQM to reside "between" G4 and M4.

More precisely, if we *define* a commutator between position and momentum *in terms of* the generators of boosts and spatial translations respectively – and note that they *do not commute* when simultaneity is relative – it possible to show that one can arrive at the quantum-mechanical commutator *of position and momentum*, and have it *equal* the quantum mechanically well-known quantity $i\hbar$. This is equivalent to asking "what is the space-time structure such that, if simultaneity is non-absolute, the Heisenberg commutator can be deduced?"[4]

Quite surprisingly, it turns out that *because* boosts do not commute with spatial translations *given that simultaneity is relative*, one can indeed deduce the quantum mechanical Heisenberg commutator (in the appropriate "*weakly*" relativistic space-time regime). This shows that some interpretation exists for both non-relativistic quantum mechanics and any relativistic quantum mechanical theory, where there is a *single, unified, space-time arena* from which either theory can be obtained in the appropriate asymptotic limit. More specifically, what the formal results in the following sections will show is that classical mechanics "lives in" G4, NRQM "lives in" a space-time regime that is between G4 and M4 (we can call it K4 after Kaiser) and relativistic quantum field theory (RQFT) "lives in" M4. It will also become clear that NRQM is a small-scale, denumerable version of RQFT in that its oscillators are distributed denumerably through space, rather than continuously as in RQFT. All of this makes for a great deal more unity between space-time structures and quantum structures than is generally appreciated.

---

[4] Since quantum theory is already well-established empirically, we essentially know what needs to be derived, we just need to find the right space-time structure. This is, admittedly, flipping the order of discovery somewhat, and asking an entirely new question regarding the "origin" of quantum theory (looking to space-time structure, and not to the structure of matter *per se*, which is how the theory of the quantum was arrived at historically).

11.3.1 NRQM: Space-Time Structure for Commutation Relations

Kaiser[5] has shown that the noncommutivity of Lorentz boosts with spatial translations is *responsible for* the noncommutivity of the quantum mechanical position operator with the quantum mechanical momentum operator. He writes[6],

> For had we begun with Newtonian space-time, we would have the Galilean group instead of [the restricted Poincaré group]. Since Galilean boosts commute with spatial translations (time being absolute), the brackets between the corresponding generators vanish, hence no canonical commutation relations (CCR)! In the [c $\rightarrow \infty$ limit of the Poincaró algebra], *the CCR are a remnant of relativistic invariance where, due to the nonabsolute nature of simultaneity, spatial translations do not commute with pure Lorentz transformations*. [Italics is his].

Bohr and Ulfbeck[7] also realized that the "Galilean transformation in the weakly relativistic regime" is needed to construct a position operator for NRQM, and this transformation "includes the departure from simultaneity, which is part of relativistic invariance." Specifically, they note that the commutator between a "weakly relativistic" boost and a spatial translation results in "a time displacement," which is crucial to the relativity of simultaneity. Thus they write[8],

> For ourselves, an important point that had for long been an obstacle, was the realization that the position of a particle, which is a basic element of nonrelativistic quantum mechanics, requires the link between space and time of relativistic invariance.

So, *the essence of nonrelativistic quantum mechanics – its canonical commutation relations – is entailed by the relativity of simultaneity*. If the transformation equations entailed by some space-time structure necessitate a temporal displacement when boosting between frames, then the relativity of simultaneity is true of that space-time structure. Given this temporal displacement between boosted frames, and given that this implies the relativity of simultaneity, our arguments supplied above show that BW is true of this space-time structure. Furthermore, since the relativity of simultaneity, via the kind of temporal displacement necessitated by boosting between frames in this space-time regime, is essential to the Heisenberg or canonical commutation relations, we find a heretofore unappreciated deep unity between STR and *nonrelativistic* quantum mechanics.

To outline Kaiser's result, we take the limit c $\rightarrow \infty$ in the Lie algebra of the Poincaré group for which the nonzero brackets are:

$$[J_m, J_n] = iJ_k$$

$$[T_0, K_n] = iT_n$$

$$[K_m, K_n] = \frac{-i}{c^2} J_k$$

[5] Kaiser (1981 and 1990).

[6] Kaiser (1981), p. 706.

[7] Bohr and Ulfbeck (1995), section D of part IV, p. 28.

[8] *Ibid*., p. 24.

$$[J_m, K_n] = iK_k$$

$$[J_m, T_n] = iT_k$$

$$[T_m, K_n] = \frac{-i}{c^2}\delta_{\text{mn}}T_0$$

where expressions with subscripts m, n, and k denote 1, 2, and 3 cyclic, J_m are the generators of spatial rotations, T_0 is the generator of time translations, T_m are the generators of spatial translations, K_m are the boost generators, $i^2 = -1$, and c is the speed of light. We obtain

$$[J_m, J_n] = iJ_k$$

$$[M, K_n] = 0$$

$$[K_m, K_n] = 0$$

$$[J_m, K_n] = iK_k$$

$$[J_m, T_n] = iT_k$$

$$[T_m, K_n] = \frac{-i}{\hbar}\delta_{\text{mn}}M$$

where M is obtained from the mass-squared operator in the c $\rightarrow \infty$ limit since

$$c^{-2}\hbar T_0 = c^{-2}P_0$$

and

$$c^{-2}P_0 = \left(M^2 + c^{-2}P^2\right)^{1/2} = M + \frac{P^2}{2\text{Mc}^2} + O(c^{-4})$$

Thus, $c^{-2}T_0 \rightarrow \frac{M}{\hbar}$ in the limit c $\rightarrow \infty$. [M mI, where m is identified as "mass" by choice of "scaling factor" $\hbar$.] So,

$$P_s := \hbar T_s$$

and

$$Q_n := \frac{-\hbar}{m}K_n$$

give

$$[P_s, Q_n] = \frac{-\hbar}{m}[T_s, K_n] = \left(\frac{-\hbar^2}{m}\right)\left(\frac{i}{\hbar}\right)\delta_{\text{sn}}mI = -i\hbar\delta_{\text{sn}}I \tag{11.1}$$

Bohr and Ulfbeck (1995) point out that in this "weakly relativistic regime" the coordinate transformations now look like

$$\begin{aligned} X &= x - \text{vt} \\ T &= t - \frac{\text{vx}}{c^2} \end{aligned} \tag{11.2}$$

These transformations differ from Lorentz transformations because they lack the factor

$$\gamma = \left(1 - \frac{v^2}{c^2}\right)^{1/2}$$

which is responsible for time dilation and length contraction. And, these transformations differ from Galilean transformations by the temporal displacement $\frac{\mathrm{vx}}{c^2}$ which is responsible for the relativity of simultaneity, i.e. in a Galilean transformation time is absolute so $T = t$. Therefore, the space-time structure of Kaiser et al. lies between Galilean space-time and Minkowski space-time and we see that the Heisenberg commutation relations are not the result of Galilean invariance, where spatial translations commute with boosts, but rather they result from the relativity of simultaneity per Lorentz invariance.

11.3.2 Heterodoxy: NRQM Does not Live in Galilean Space-Time

The received view has it that Schrödinger's equation is Galilean invariant, so it is generally understood that NRQM resides in Galilean space-time and therefore respects absolute simultaneity[9]. However, as we have seen above, Kaiser (1981), Bohr and Ulfbeck (1995), and Anandan (2003) have shown independently that the Heisenberg commutation relations of NRQM follow from the relativity of simultaneity[10]. *Prima facie* these results seem incompatible with the received view, so to demonstrate that these results are indeed compatible, we now show that these results do not effect the Schrödinger dynamics[11].

Why is it that the dynamics of NRQM, given by the Schrödinger equation, are Galilean invariant? That is, why are the dynamics of NRQM unaffected by the relativity of simultaneity reflected in the geometry of Eq. (11.1)?

To answer this question we operate on $|\psi\rangle$ first with the spatial translation operator then the boost operator and compare that outcome to the reverse order of operations. The spatial translation (by a) and boost (by v) operators in x are:

$$U_T = e^{-\mathrm{iaT}_x} \text{ and } U_K = e^{-\mathrm{ivK}_x} \tag{11.3}$$

respectively. These yield

$$U_K U_T |\psi\rangle = U_T U_K e^{\mathrm{iavmI}/\hbar} |\psi\rangle \tag{11.4}$$

[9] See Brown and Holland (1999).

[10] Of course, all other commutation relations in NRQM follow from those of position and momentum with the exception of spin. Since, operationally, spin measurements are simply binary outcomes in space related to, for example, the spatial orientation of a Stern–Gerlach apparatus, our model encompasses such properties as spin to the extent that we model all outcomes in space and time as *irreducible relations* between the spatiotemporal regions corresponding to source and detector.

[11] See also Lepore (1960) who also realizes that this time-shift between frames is without effect on the dynamics of Schrödinger evolution.

Thus, we see that the geometric structure of Eq. (11.1) introduces a mere phase to $|\psi\rangle$ and is therefore, without consequence in the computation of expectation values. And in fact, this phase is consistent with that under which the Schrödinger equation is shown to be Galilean invariant[12].

Therefore, we realize that the space-time structure for NRQM, while not M4 in that it lacks time dilation and length contraction, nonetheless contains a "footprint of relativity"[13] due to the relativity of simultaneity. Thus, there is an unexpected and unexplored connection between the relativity of simultaneity and the noncommutativity of NRQM. In light of this result, it should be clear that there is no metaphysical tension between STR and NRQM. This formal result gives us motivation for believing that NRQM is intimately connected to the geometry of (a suitable) space-time.[14]

11.3.3 Philosophical Significance

One important point should be brought out, which reveals how we understand the relationship between space-time structure (given by relativity) and the theory of quantum mechanics (in a non-Minkowskian and non-Galilean space-time regime, i.e. K4). Most natural philosophers agree that STR just constrains the set of possible dynamical theories to those which satisfy the light and relativity postulates. It is often worried, as we have pointed out, that somehow quantum theory *violates* those constraints. The view we adopt here is importantly different in that we distinguish between:

1. The question of how to relate the *structures* of quantum theory and relativity
2. The question of the compatibility of constructive interpretations of quantum theory and whether they violate relativistic constraints

Per the collection of formal results due to Kaiser et al. outlined *supra*, we now understand that the space-time structure for which one can obtain the Heisenberg commutation relations is one where the relativity of simultaneity is *upheld*–a fact often not appreciated in most interpretations of quantum theory. Furthermore, with an ontology of space-time relations, we can construct a quantum density operator from the space-time symmetry group of any quantum experimental configuration, and use this to deduce and then explain the phenomenon of quantum interference – all by appealing to nothing more than a space-time structure for which one can obtain the Heisenberg commutator while obeying the relativity of simultaneity.

[12] See Eq. 6 in Brown and Holland (1999). A derivation of Eq. 11.1, assuming the acceptability of a phase difference such as that in Eq. 11.4, is in Ballentine (1990), p. 49–58.

[13] This phrase was used by Harvey Brown in a conversation with the authors while describing his work with Peter Holland (Brown and Holland, 1999).

[14] The Bohr et al. result of section 4 below shows how to relate this space-time geometry to nonrelativistic quantum mechanics by showing how a quantum-density operator can be constructed from the space-time symmetry group of the quantum mechanical experiment.

We take the deepest significance of the Kaiser et al. results to be that, given the asymptotic relationship between the space-time structure of special relativity and the "weakly relativistic" space-time structure of quantum theory, nonrelativistic quantum mechanics is something like a relativity theory in an "embryonic" stage. It is "embryonic" in that it is yet without the Lorentz-contraction factor γ that appears in the familiar Lorentz transformation equations of special relativity[15] and, as we will see, it assumes a denumerable distribution of oscillators where RQFT assumes a continuous distribution of oscillators.

Having identified the appropriate space-time structure for the Heisenberg commutation relations, and having discovered that this structure upholds the relativity of simultaneity, we have provided a geometric explanation for the quantum. A natural question now arises: what would the appropriate description of NRQM and quantum mechanical phenomena such as interference be like in light of the asymptotic relationship between relativity and quantum theory? Our "geometric" interpretation of NRQM elaborated in sections 4 and 5 is one answer to this question, an answer grounded in our fundamental ontology of space-time relations.

11.4 Density Matrix via Symmetry Group

Having found a *space-time* structure that is appropriate for the Heisenberg commutation relations (whose empirical manifestation is quantum interference), we now seek to address the question of how to model – *in space-time and not in Hilbert space* – any system which manifests quantum interference. That is, we are asking:

> *How can we describe a quantum system with nothing more than the geometry of space-time, where the relativity of simultaneity and the non-commutivity of position and momentum obtain?*

The following formal results provide us with an answer to this question.

11.4.1 Formalism

We present a pedagogical version of the appendix to Bohr, Mottelson, and Ulfbeck (2004a) wherein they show the density matrix can be derived using only the irreducible representations of the symmetry group elements, $g \in G$. We begin with two theorems from Georgi

> The matrix elements of the unitary, irreducible representations of G are a complete orthonormal set for the vector space of the regular representation, or alternatively, for functions of $g \in G$. (1999, 14)

[15] And given that it is the contraction/dilation phenomena, characteristic of relativity, that motivates the introduction of the "field" as a unifying structural device, non-relativistic quantum mechanics in light of this new space-time structure is simply relativity minus the "field."

> If a hermitian operator, H, commutes with all the elements, D(g), of a representation of the group G, then you can choose the eigenstates of H to transform according to irreducible representations of G. If an irreducible representation appears only once in the Hilbert space, every state in the irreducible representation is an eigenstate of H with the same eigenvalue. (*Ibid.*, p. 25)

What we mean by "the symmetry group" is precisely that group G with which some observable H commutes (although, these elements may be identified without actually constructing H). Thus, the mean value of our hermitian operator H can be calculated using the density matrix obtained wholly by $D(g)$ and $\langle D(g)\rangle$ for all g ϵ G. Observables such as H are simply 'along for the ride' so to speak.

To show how, in general, one may obtain the density matrix using only the irreducible representations[16] $D(g)$ and their averages $\langle D(g)\rangle$, we start with Eq. 1.68 of Georgi (*ibid.*, 18)

$$\sum_g \frac{n_a}{N}\left[D_a\left(g^{-1}\right)\right]_{\mathrm{kj}}[D_b(g)]_{\mathrm{lm}} = \delta_{\mathrm{ab}}\delta_{\mathrm{jl}}\delta_{\mathrm{km}}$$

where n_a is the dimensionality of the irrep, D_a, and N is the group order. If we consider but one particular irrep, D, this reduces to the orthogonality relation Eq. (11.5) of Bohr et al.

$$\sum_g \frac{n}{N}[D(g^{-1})]_{\mathrm{jk}}[D(g)]_{\mathrm{lm}} = \delta_{\mathrm{jl}}\delta_{\mathrm{km}} \tag{11.5}$$

where n is the dimension of the irrep. Now multiply by $[D(g')]_{\mathrm{jk}}$ and sum over k and j to obtain

$$\sum_j\sum_k\sum_g \frac{n}{N}\left[D(g^{-1})\right]_{\mathrm{jk}}[D(g)]_{\mathrm{lm}}\left[D(g')\right]_{\mathrm{jk}} = \sum_j\sum_k \delta_{\mathrm{jl}}\delta_{\mathrm{km}}\left[D\left(g'\right)\right]_{\mathrm{jk}} = \left[D\left(g'\right)\right]_{\mathrm{lm}}$$

The sum over j on the LHS gives

$$\sum_j [D(g^{-1})]_{\mathrm{kj}}[D(g')]_{\mathrm{jk}} = [D(g^{-1})D(g')]_{\mathrm{kk}}$$

The sum over k then gives the trace of $D(g^{-1})D(g')$, so we have

$$\frac{n}{N}\sum_g [D(g)]_{\mathrm{lm}}\mathrm{Tr}\left\{D(g^{-1})D(g')\right\} = [D(g')]_{\mathrm{lm}}$$

Dropping the subscripts we have Eq. (11.6) of Bohr et al.

$$\frac{n}{N}\sum_g D(g)\mathrm{Tr}\left\{D(g^{-1})D(g')\right\} = D(g') \tag{11.6}$$

If, in a particular experiment, we measure directly the click distributions associated with the various eigenvalues of a symmetry $D(g)$, we obtain its average outcome, $\langle D(g)\rangle$, i.e., eqn. 3 of Bohr et al.

---

[16] Hereafter, "irreps."

$$\langle D(g)\rangle = \sum_i \lambda_i p(\lambda_i) \tag{11.7}$$

where λ_i are the eigenvalues of $D(g)$ and p(λ_i) are the distribution frequencies for the observations of the various eigenvalues/outcomes.

In terms of averages, Bohr et al. Eq. (11.6) becomes

$$\frac{n}{N}\sum_g \langle D(g)\rangle \mathrm{Tr}\left\{D(g^{-1})D(g')\right\} = \left\langle D(g')\right\rangle \tag{11.8}$$

which they number Eq. (11.8). Since we want the density matrix to satisfy the standard relation (Bohr et al. Eq. (11.9))

$$\mathrm{Tr}\left\{\rho D(g')\right\} = \left\langle D(g')\right\rangle \tag{11.9}$$

it must be the case that (Bohr et al. Eq. (11.10))

$$\rho \equiv \frac{n}{N}\sum_g D(g^{-1})\langle D(g)\rangle \tag{11.10}$$

That this density operator is hermitian follows from the fact that the symmetry operators are unitary. That is, $D(g^{-1}) = D^{+}(g)$ implies $\left\langle D(g^{-1})\right\rangle = \langle D(g)\rangle^{*}$, thus

$$\rho^{+} = \frac{n}{N}\sum_g D^{+}(g^{-1})\langle D(g)\rangle^{*} = \frac{n}{N}\sum_g D(g)\left\langle D(g^{-1})\right\rangle = \frac{n}{N}\sum_g D(g^{-1})\langle D(g)\rangle = \rho$$

[The second-to-last equality holds because we are summing over all g and for each g there exists g^{-1}.] So, the density operator of Eq. (11.10) will be hermitian and, therefore, its eigenvalues are guaranteed to be real. This is not necessarily the case for $D(g)$, since we know only that they are unitary. However, we need only associate detector clicks with the eigenvalues of $D(g)$ and in this perspective *one does not attribute an eigenvalue of* $D(g)$ *to a property of some click-causing particle*. Therefore, whether or not the eigenvalues of any particular $D(g)$ are real or imaginary is of no ontological or empirical concern.

11.4.2 Philosophical Significance

With the above formal result in hand, we now provide a clear answer to the question posed at the beginning of this section:

> *The space-time symmetry group of the quantum mechanical experiment will yield the quantum mechanical density matrix.*

The methodological significance of the Bohr et al. formal result is that any NRQM system may be described with the appropriate space-time symmetry group. But the philosophical significance of this proof is more interesting, and one rooted in our ontological space-time relationalism. Our view is a form of ontological structural realism which holds that the features of our world picked out by STR and NRQM are structures as opposed to constructive entities; moreover, we think that the structures picked out by our most successful theories to date are geometrical structures. And those structures, if taken seriously, are, we posit, structures of space-time *relations*. Furthermore, we see the quantum theory as providing a *further structural constraint* on the distribution of space-time events. Isolated to an idealized model of "sources," "detectors," "mirrors," etc., our ontology is that each and every "click" or "measurement event" observed in the detector region is itself *evidence of space-time relations between the source and detector*. So, while the "click" itself maybe regarded as a classical object, it is not "caused by" a constructive entity such as a particle that is independent from the physical space-time geometry of this entire measurement process and experimental setup. Rather, the click itself is a manifestation of spatiotemporal relations between elements of the experimental setup. It is in this way, via our radical ontology of space-time relations[17], that the essential features of quantum systems with interference can be described with features of the space-time geometry *without appealing to features of the usual Hilbert space of quantum mechanical states*.[18]

Secondly, as will be demonstrated below, the Bohr et al. proof will allow us to show that the posit of a block world – the reality of all space-time events, and hence, in our ontology, of all space-time relations constituting those events–does real explanatory work. While one can imagine quite trivial explanations of EPR–Bell correlations invoking the blockworld, the Bohr et al. result will allow us to provide a nontrivial, geometric explanation for such quantum correlations.

Thirdly, as demonstrated below, the Bohr et al. result provides the foundation for our distinctly *geometrical* ontological structuralist[19] interpretation of NRQM. This ontology is an ontology of spatiotemporal relations which are the means by which all physical phenomena (including both quantum and classical "entities") are modeled. Our relationalism allows for a natural transition from quantum to classical mechanics (including the transition from quantum to classical probabilities) as simply the transition from rarefied to dense collections of space-time relations.[20]

11.5 Geometric Interpretation of NRQM

Given our geometrical interpretation of NRQM, it should be clear that we do not take detector events to be indicators of the trajectories of classical-like particles or wave functions, propagating from the source to the detector as in Bohm's mechanics

[17] Which, if you want to speak constructively, "constitute" the space-time geometry.

[18] A Hilbert space is not analogous to space-time geometry, but rather to phase-space geometry. Anandan (1991) for example, adopts the view that the geometry of Hilbert space is appropriate for a geometric interpretation of quantum theory.

[19] See French and Ladyman 2003 for an account of ontological strucuturalism in the context of quantum theory.

[20] Though a full explication and defense of this view is unfortunately beyond the scope of this paper.

or even, as it turns out, like disturbances in a field per RQFT. The transition amplitude in RQFT for a scalar field without scattering or sources is (Zee 2003, 18)

$$Z = \int D\varphi e^{i\int d^4x\left[\frac{1}{2}(\partial\varphi)^2 - V(\varphi)\right]} \tag{11.11}$$

According to Zee (2003), NRQM then obtains in $(0+1)$ dimensions. In Zee's derivation of Eq. (11.11) *from* NRQM, the field φ is obtained in the continuum limit of a discrete set of oscillators q_a distributed in a spatial lattice, where "a" denotes the oscillator's location in the spatial lattice. Any one of these q_a is supposed to replace φ in Eq. (11.11) in order that it yield NRQM. However, each q_a is fixed in space so the notion that we're integrating over all possible paths *in space* (the standard treatment) from a source to a detector when we compute Z is not ontologically consistent with the fact that we integrate over all values of q, but not over all values of the index "a" in q_a. We rather suggest that the method for producing NRQM from RQFT is to associate sources J(x) with elements in the experimental set up while assuming the q's are distributed discretely therein. Thus, we want to obtain NRQM from

$$Z = \int D\varphi e^{i\int d^4x\left[\frac{1}{2}(\partial\varphi)^2 - V(\varphi) + J(x)\varphi(x)\right]} \tag{11.12}$$

(Zee 2003, 84), so we must compute

$$Z = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \mathrm{dq}_1 \dots \mathrm{dq}_N e^{\frac{i}{2}q\cdot A\cdot q + \mathrm{iJ}\cdot q} \tag{11.13}$$

where A_{ij} is the discrete matrix counterpart to the dynamical differential operator associated with the equations of motion, e.g., the Klein–Gordon operator $-(\partial^2 + m^2)$ in the Gaussian theory, and J_n and q_m are the discrete vector versions of J(x,t) and q(x,t), denoting their location in the space-time lattice. The solution to Eq. (11.13) for each path from q_{initial} at the source to q_{final} at some point on the detector is

$$\left(\frac{(2\pi i)^N}{\det(A)}\right)^{1/2} e^{-\frac{i}{2}J\cdot A^{-1}\cdot J} \tag{11.14}$$

For the twin-slit experiment, which "has in it the heart of quantum mechanics. In reality, it contains the *only* mystery"(Feynman et al., 1965, italics theirs), we have four J's which must be taken into account to compute the amplitude between q_{initial} and q_{final}. The J's (used to denote sources/sinks for particles in RQFT) are located at the experimental source (label it J_1), at each slit in the screen (label them J_2 and J_4) and at some point on the detector (label it J_3). Since we have two paths from q_{initial} to q_{final}, i.e., $J_1 \to J_2 \to J_3$ and $J_1 \to J_4 \to J_3$ (see Fig. 11.1), the amplitude is given by

$$\psi \sim e^{-\frac{i}{2}\left(J_1A_{12}^{-1}J_2 + J_2A_{23}^{-1}J_3\right)} + e^{-\frac{i}{2}\left(J_1A_{14}^{-1}J_4 + J_4A_{43}^{-1}J_3\right)} \tag{11.15}$$

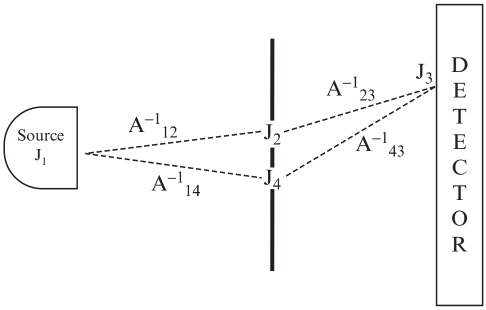

Fig. 11.1 Twin-slit experiment.

Typically, the source is equidistance from either slit so $J_1A_{12}^{-1}J_2 = J_1A_{14}^{-1}J_4$ and Eq. (11.15) reduces to the familiar form

$$\psi \sim e^{-\frac{i}{2}\left(J_2A_{23}^{-1}J_3\right)} + e^{-\frac{i}{2}\left(J_4A_{43}^{-1}J_3\right)} \tag{11.16}$$

upon removing the common phase factor which will not contribute to interference per the Born rule.

Since we are using the machinery of RQFT one might (erroneously) infer, in the parlance of particle physics, that we are creating a pair of particles at the source which propagate to each slit and then annihilate at the detector screen. Recall, however, that we have a discrete distribution of q's at the source, slits and detector per NRQM rather than the continuous field φ in the space between the source, screen and detector (again, this is a major difference between NRQM and RQFT). Thus, it should be clear per our leitmotif that there is no particle or wave propagating from the source through the slit(s) to the detector that causes a click thereupon. Rather the distribution of clicks at the detector simply evidences the space-time relations (given here by A_{ij}^{-1}) between the various elements of the experimental arrangement, i.e. between the only "things" with ontic status.

11.5.1 Interpretive Consequences of Our Geometrical NRQM

11.5.1.1 The Measurement Problem

According to the account developed here, we offer a deflation of the measurement problem with a novel form of a hidden-variables "statistical interpretation." The fundamental difference between our version of this view and the usual understanding of it is the following: whereas on the usual view the state description refers to an "ensemble", which is an ideal collection of similarly prepared quantum particles, "ensemble" according to our view is just an ideal collection of space-time regions D_i "prepared" with the same spatiotemporal boundary conditions per the experimental configuration itself. The union of the click events in each D_i, as i $\rightarrow \infty$, produces the characteristic Born distribution[21]. Accordingly, probability on our geometrical NRQM is interpreted per relative frequencies. It should be clear, also, that probabilities are understood as the likelihood that a particular relation between source-detector in space-time is realized, from among a set of all equally likely relations between source-detector.

On our view, the wave function description of a quantum system can be interpreted statistically because we now understand that, as far as measurement outcomes are concerned, the Born distribution has a basis in the space-time symmetries of experimental configurations. Each "click," which some would say corresponds to the impingement of a particle onto a measurement device and whose probability is computed from the wave-function (transition amplitude), corresponds to a space-time relation in the context of the experimental configuration. The measurement problem *exploits* the possibility of extending the wave function description from the quantum system to the whole measurement apparatus, whereas the space-time description according to our geometrical quantum mechanics *already includes* the apparatus via the space-time symmetries instantiated by the entire experimental configuration. The measurement problem is therefore a nonstarter on our view.

11.5.1.2 Entanglement and Nonlocality

On our geometric view of NRQM we explain entanglement as a feature of the space-time geometry[22] as follows. Each detection event, which evidences space-time relations, "selects" a trajectory from a family of possible Hamilton–Jacobi trajectories (one family per entangled "particle"). In the language of detection events *qua* relations, it follows that correlations are correlations between the members of the *families of trajectories* and these correlations are the result of the relevant space-time symmetries for the experimental configuration. And, since an experiment's space-time symmetries are manifested in the Hamilton–Jacobi *families* of trajectories throughout the relevant space-time region D, there is no reason to expect entanglement to diminish with distance from the source. Thus, the entanglement of families of trajectories is spatiotemporally global, i.e. nonlocal. That is, there is no reason to expect entanglement geometrically construed to respect any kind of common cause principle. Obviously, on our geometric interpretation there is no nonlocality in the odious sense we find in Bohm for example, that is, there are no instantaneous causal connections (construed dynamically or in terms of production – bringing new states of affairs into being) between spacelike separated events – no action at a distance. However our view is nonlocal in the sense that it violates the locality principle. The locality principle states: the result of a measurement is probabilistically independent of actions performed at spacelike separation from the

[21] There would be N first events in trials with N entangled particles, since each "particle" would correspond to a family of possible trajectories.

[22] Established in section 3 as one which is "weakly" relativistic in that it lacks the Lorentz contraction factor.

measurement. Keep in mind that in our BW setting, talk of "actions performed" gets only a purely logical-counterfactual meaning – the entire experimental EPR setup, its past, present and future and the space-time symmetries of that set-up are all just "there" – no one could really perform some "alternative" measurement on the other wing of the experiment without *changing the description of the measurement configuration as a spatiotemporal whole*.

We understand quantum facts to be facts about the spatiotemporal relations of a given physical system, not facts about the behavior of particles, or the interactions of measurement devices with wave functions, or the like. Entanglement and nonlocality are built, self-consistently, into the structure of space-time via relations A_{ij} and A_{ij}^{-1}, which *give rise to* classical objects in space-time regions where they are sufficiently dense. Correlations between spacelike separated events that violate Bell' inequalities are of no concern as long as space-time symmetries instantiated by the experimental apparatus warrant the correlated space-time relations. Since the nonlocal correlations derive from the spatiotemporal relations per the space-time symmetries of the experiment, satisfaction of any common-cause principle is superfluous. To sloganize: ours is a *purely* geometric/spacetime interpretation of nonrelativistic quantum mechanics.

That the density matrix may be obtained from the space-time symmetries of the Hamiltonian is consistent with the notion that $\psi^*\psi$ provides the distribution for detector events in single-event trials for each family of trajectories obtained via the Hamilton–Jacobi formalism. Our view exploits this correspondence to infer the existence of space-time relations between source and detector for each detector event. Subsequent detector events in close spatiotemporal proximity to the first tend to fall along a trajectory of the family consistent with the first event thereby allowing for the inference of a "particle" (even though, as shown above, there *isn't even a field* in the space-time region between these events). In this sense, what constitutes a "rarefied" distribution of space-time relations is but one relation per "particle," i.e. family of trajectories, since subsequent events tend to trace out classical trajectories (scattering and particle decay events aside). It is a collection of these single-event trials that will evidence quantum interference in, for example, the twin-slit experiment.

Our account provides a clear description, in terms of fundamental space-time relations, of quantum phenomena *that does not suggest the need for a "deeper" causal or dynamical explanation*. If explanation is simply determination, then our view explains the structure of quantum correlations by invoking what can be called *acausal, spatiotemporally global determination relations*. These global determination relations are given by the space-time symmetries which underlie a particular experimental setup. Not objects and dynamical laws, but rather acausal space-time relations per the relevant space-time symmetries do the fundamental explanatory work according to our version of geometrical quantum mechanics. We can invoke the entire space-time configuration of the experiment so as to predict, and explain, the EPR–Bell correlations. Indeed, it has been the purport of this paper that the space-time symmetries of the quantum experiment can be used to construct its quantum-density operator, that such a space-time is one for which simultaneity is relative, and that events in the detector regions evidence spatiotemporal relations.

This constitutes an acausal and non-dynamical characterization and explanation of entanglement. According to our view, the structure of EPR correlations are determined by the space-time relations instantiated by the experiment, understood as a spatiotemporal whole, i.e. *block-world*. This determination is obtained by systematically describing the spatiotemporal symmetry structure of the Hamiltonian for the experimental arrangement.[23] Since

1. The explanation lies in the space-time symmetries as evidenced, for example, in the family of trajectories per the Hamilton–Jacobi formalism
2. Each family of trajectories characterizes the distribution of space-time relations
3. We take those relations to be a timeless "block"
4. These relations collapse the matter-geometry dualism, therefore
5. Our geometrical quantum mechanics provides for an acausal, global, and non-dynamical understanding of quantum phenomena

11.6 Conclusion: NRQM Resides in a 4D Block World

Can one do justice to the noncommutative structure of NRQM without being a realist about Hilbert space? Our geometric interpretation constitutes an affirmative answer to this question. The trick is to appreciate that while everything "transpires" or rather resides in a 4D space-time and nowhere else, some phenomena, namely quantum phenomena, cannot be modeled with world lines if one is to do justice to its noncommutative structure. Thus while clicks in detectors are perfectly classical events, the clicks are not evidence of constructive entities such as particles with world lines, rather the clicks are manifestations of space-time relations between elements of the experimental configuration distributed per the space-time symmetries. Therefore, there is no "Dedukind cut" between the quantum and the classical as some versions of the Copenhagen interpretation would have it. After all, we can explain asymptotically the transition from the quantum to the classical in terms of the density of "events." And there is also no "Einstein separability" between the system being measured and the system doing the measuring on our interpretation. Our view respects the causal structure of Minkowski space-time in the sense that there are no faster than light "influences" or "productive" causes between spacelike separated events as there are in Bohm for example. So our view is not nonlocal in any robustly dynamical sense. However our view does violate Einstein separability and it does allow for "correlations" outside the light cone as determined acausally and globally by the space-time symmetries.

Such acausal global determination relations do not respect any common cause principle. This fact should not bother anyone who has truly transcended the idea

[23] The experimental apparatus itself providing the particular initial and final "boundary conditions" needed for a prediction unique to the apparatus.

that the dynamical or causal perspective is the most fundamental one. We are providing a model of an irreducibly relational block world, which is what realism about the quantum structure and the 4D space-time structure yields once one accepts the implication therein of Hilbert space anti-realism.

References

Albert, David in *Bohmian Mechanics And Quantum Theory: An Appraisal*, J.T. Cushing, A. Fine, and S. Goldstein (eds.), Kluwer, Dordrecht, 1996, p. 277–284.

Anandan, Jeeva (1991), A Geometric Approach to Quantum Mechanics, *Foundations of Physics* **21** (11): 1265–1284.

Anandan, Jeeva (2003), Laws, Symmetries, and Reality, *International Journal of Theoretical Physics* **42** (9): 1943–1955.

Barrett, Jeffrey (1999), *The Quantum Mechanics of Minds and Worlds*. Oxford: Oxford University Press.

Bohr, A. and Ulfbeck, O. (1995), Primary Manifestation of Symmetry. Origin of Quantum Indeterminacy," *Reviews of Modern Physics* **67** (1): 1–35.

Bohr, A., Mottelson, B.R., and Ulfbeck, O. (2004a), "The Principle Underlying Quantum Mechanics," *Foundations of Physics* **34** (3): 405–417.

Bohr, A., Mottelson, B.R., and Ulfbeck, O. (2004b), Quantum World Is Only Smoke and Mirrors, *Physics Today* **57** (10): 15–16.

Brown, H. and Holland, P.R. (1999), The Galilean Covariance of Quantum Mechanics in the Case of External Fields," *American Journal of Physics* **67** (3): 204–214.

Callender, C. (2006) in *Absolute Simultaneity*, W.L. Craig and Q. Smith (eds.), Oxford: Oxford University Press.

Feynman, R. P., Leighton, R.B. and Sands, M., (1965), *The Feynman Lectures on Physics, Vol. III, Quantum Mechanics*, Reading, MA: Addison-Wesley.

French, Steven, and Ladyman, James (2003), Remodelling Structural Realism: Quantum Physics and the Metaphysics of Structure,*Synthese* **136** (1): 31–56.

Georgi, Howard (1999), *Lie Algebras in Particle Physics,* 2nd Edn. New York: Perseus Books.

Holland, Peter R. (1993), *The Quantum Theory of Motion: An Account of the de Broglie-Bohm Causal Interpretation of Quantum Mechanics*. Cambridge: Cambridge University.

Kaiser, Gerald (1981), Phase-space approach to relativistic quantum mechanics. III. Quantization, relativity, localization and gauge freedom, *Journal of Mathematical Physics* **22** (4): 705–714.

Gerald Kaiser, (1990), *Quantum Mechanics, Relativity, and Complex Space-time: Towards a New Synthesis*. Amsterdam: North-Holland.

Lepore, J.V. (1960), "Commutation relations of quantum mechanics," *Physical Review* **119** (2): 821–826.

Lewis, Peter J. (2004), "Life in Configuration space," *British Journal for the Philosophy of Science* **55** (4): 713–729.

Price, H. (1996), *Time's Arrow and Archimedes Point.* Oxford: Oxford University Press.

Stuckey, W.M., Silberstein, M. and Cifone, M. to appear in *Physics Essays* **19** (2006; preprint quant-ph/0503065).

Tooley, Michael (1997), *Time, Tense and Causation*. New York: Clarendon Press.

Zee, A. (2003), *Quantum Field Theory in a Nutshell*. Princeton: Princeton University Press.

Chapter 12
Space-time: Arena or Reality?

H. I. Arcos and J. G. Pereira

12.1 Introduction

The concept of fundamental particle has been quite elusive along the history of physics. The term *fundamental* is commonly used as a synonymous of structureless particles. However, this assumption is clearly contradictory. For example, it is impossible to explain spin without assuming a structure for the particle. In fact, a point particle is by definition spherically symmetric, a symmetry violated by the presence of spin. This problem is usually circumvented by saying that spin is a purely quantum property, which cannot be explained by classical physics. This means to keep it as a mysterious property of nature.

If one assumes that a fundamental particle is a point-like object, several arguments against this idea show up immediately. First, as discussed above, a point-like object seems to be inconsistent with the existence of spin. Second, if we try to reconcile general relativity with point-particles, which are singular points in a pseudo-Riemannian space-time, unwanted features, like for example, ultraviolet divergences, will appear. A natural alternative would be to assume that a fundamental particle is a string-like object, a point of view adopted by string theory [1]. Similarly, one can introduce membranes as fundamental objects, or even extended objects with certain geometries. These models, however, are also plagued by problems. The membrane model has failed to generate a theory free of negatively normed states, or tachyons, and theories with extended objects have failed to explain the existence of supporting internal forces that avoid the model to collapse.

With the evolution of particle physics and gravitation, the idea that a fundamental particle should somehow be connected to space-time began to emerge. This is the case, for example, of Wheeler's approach, which was based on the concept of space-time foam. At the Planck scale, uncertainty in energy allows for large curvature values. At this energy, space-time can undergo deep transformations, which modify the small-scale topology of the continuum. This is where the "foam" notion becomes important. Small regions of space-time can join and/or separate giving rise

V. Petkov (ed.), *Relativity and the Dimensionality of the World*, 217–260.

to nontrivial topological structures. The simplest of these structures is the so called wormhole, a quite peculiar solution to Einstein's equation. It represents a topological structure that connects space-time points separated by an arbitrary spatial distance. An interesting property of the wormhole solution is that it can trap an electric field. Since, for an asymptotic observer, a trapped electric field is undistinguishable from a charge distribution, Wheeler introduced the concept of "charge without charge" [2]. However, as Wheeler himself stated, these Planckian wormholes could not be related to any particle model for several reasons: charge is not quantized, they are not stable, their mass/charge ratio is very different from that found in known particles, and half-integral spin cannot be defined for a simple wormhole solution. There was the option to interpret a particle as formed by a collective motion of wormholes, in the same way phonons behave as particles in a crystal lattice. None of these ideas were developed further.

The discovery of the Kerr–Newman (KN) solution [3–5] in the early 1960s opened the door for new attempts to explore space-time-rooted models for fundamental particles [6–9]. In particular, using the Hawking and Ellis extended interpretation of the KN solution [10], as well as the Wheeler's concept of "charge without charge," a new model has been put forward recently [11]. The purpose of this chapter is to present a glimpse on the characteristics of this model, as well as to analyze the consequences for the concept of space-time. We begin by reviewing, in the next section, the main properties and the topological structure of the KN solution.

12.2 Kerr–Newman Solution

12.2.1 The Kerr–Newman Metric

The stationary axially symmetric Kerr–Newman (KN) solution of Einstein's equations was found by performing a complex transformation on the tetrad field for the charged Schwarzschild (Reissner–Nordström) solution [3–5]. For $m^2 \geq a^2 + q^2$, it represents a black hole with mass m, angular momentum per unit mass a, and charge q (we use units in which $\hbar = c = 1$). In the so called Boyer-Lindquist coordinates r, θ, ϕ, the KN solution is written as

$$ds^2 = dt^2 - \frac{\rho^2}{\Delta}\, dr^2 - (r^2 + a^2)\sin^2\theta\, d\phi^2 - \rho^2\, d\theta^2 - \frac{Rr}{\rho^2}\,(dt - a\sin^2\theta\, d\phi)^2, \quad (12.1)$$

where

$$\rho^2 = r^2 + a^2\cos^2\theta, \quad \Delta = r^2 - Rr + a^2, \quad R = 2m - q^2/r.$$

This metric is invariant under the change $(t, a) \to (-t, -a)$. It is also invariant under $(m, r) \to (-m, -r)$ and $q \to -q$. This black hole is believed to be the final stage of a very general stellar collapse, where the star is rotating and its net charge is different from zero.

The structure of the KN solution changes deeply for $m^2 < a^2 + q^2$. Due to the absence of a horizon, it does not represent a black hole, but a circular naked singularity in space-time. The metric singularity at the ring that defines the border of the disk cannot be removed by any coordinate transformation. This means that there is a true singularity at the border. However, the metric singularity at the interior points of the disk can be removed by introducing a specific interpretation of the KN solution, as described by Hawking and Ellis [10]. In what follows we give a detailed description of the topological structure behind such interpretation.

12.2.2 The Hawking–Ellis Extended Interpretation

The lack of smoothness of the metric components across the enclosed disk can be remedied by considering the extended space-time interpretation of Hawking and Ellis [10]. The basic idea of this extension is to consider that our space-time is connected to another one through the interior points of the disk. This extended solution does not necessarily implies that the dimensionality of space-time is greater than four, but rather that the manifold volume is greater than expected. In other words, the disk surface (with the upper points considered different from the lower ones) is interpreted as a shared border between our space-time, denoted by $\mathbf{M}$, and another similar one, denoted by $\mathbf{M}'$. According to this construction, the KN metric components are no longer singular across the disk, making it possible to smoothly join the two space-times, giving rise to a single 4D space-time, denoted $\mathcal{M}$. This link can be seen in Fig. 12.1 as solid cylinders going from $\mathbf{M}$ to $\mathbf{M}'$. In this figure, to clearly distinguish the upper from the lower side, the disk was drawn as if it presented a finite thickness. In order to cross the disk, therefore, an electric field line that hits the surface A will forcibly emerge from surface D, in $\mathbf{M}'$. Then, it must go through surface C to finally emerge from surface B, in $\mathbf{M}$. This picture gives a clear idea of the topological structure underlying the KN solution.

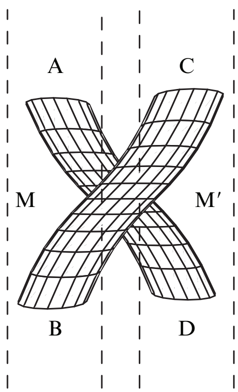

Fig. 12.1 To better visualize the intrinsic geometry of the KN manifold, the KN disk is drawn as if it presented a finite thickness, and consequently there is a space separation between the upper and lower surfaces of the disk. The left-hand side represents the upper and lower surfaces of the disk in $\mathbf{M}$, whereas the right-hand side represents the upper and lower surfaces of the disk in $\mathbf{M}'$.

Now, the singular disk is located at $\theta = \pi/2$ and $r = 0$. Therefore, if r is assumed to be positive in $\mathbf{M}$, it will be negative in $\mathbf{M}'$. Since the KN metric must be the same on both sides of the solution, the mass m will be negative in $\mathbf{M}'$. Furthermore, the magnitude of the electric charge q on both sides of the solution is, of course, the same. Taking into account that the source of the KN solution is represented by the electromagnetic potential

$$\boldsymbol{A} = -\frac{qr}{\rho^2}(dt - a\sin^2\theta d\phi), \tag{12.2}$$

which is clearly singular along the ring, and since r has different sign on different sides of the solution, we see from this expression that, if the charge is positive in one side, it must be negative in the other side.

12.2.3 Causality Versus Singularity

As already remarked, the above-extended interpretation does not eliminate the singularity at the rim of the disk. However, there are some arguments that can be used to circumvent this problem. First, it is important to observe that there is a torus-like region around the singular ring, in which the coordinate ϕ becomes timelike. Inside this region, defined by

$$r^2 + a^2 + \left(\frac{rR}{\rho^2}\right) a^2 \sin^2\theta < 0, \tag{12.3}$$

there will exist closed timelike curves [12]. In fact, when crossing the surface of this region, the signature of the metric changes from $(-,-,-,+)$ to $(-,-,+,+)$. This reduction in the number of spatial dimensions is a drawback of the solution.

Now, when the values of a, q and m are chosen to be those of the electron, the surface of the torus-like region is separated from the singular ring by a distance of the order of 10^{-34} cm, which coincides roughly with the Planck length. At this scale, as is well known, topology changes are expected to exist, and consequently changes in the connectedness of space-time topology are likely to occur. A solution to this problem is to excise the infinitesimal region around the singular ring on both the positive and negative r sides, and then glue back the manifold.[1] A simple drawing of the region to be excised can be seen in Fig. 12.2, where the direction of the gradient of r has been drawn at several points. As an example, note that the point A on the positive r side must be glued to the point A on the negative r side. If we glue all points of the torus border, we obtain a continuous path for the electric field lines that flow through the disk, even for those lines that would hit the disk at the singular ring. Furthermore, since the extrinsic curvature does not change sign

[1] This kind of singularity removal has already been explored by Punsly for the case of the Kerr solution [13].

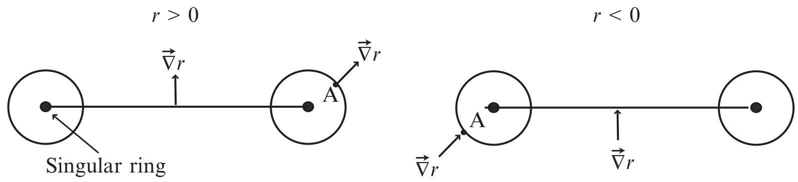

Fig. 12.2 Tubular-like regions around the singular ring, which is to be excised. Several ∇r directions are also depicted, which show how the borders in the positive and negative r sides can be continuously glued.

when crossing the hypersurface $g_{\phi\phi} = 0$, the above gluing process does not generate stress–energy [8].

An important point of the above structure is that, after removing the tubular region around the singular ring, the surface delimiting both space-times turns out to be defined by a reversed topological product between two 2-torus. As is well known, this is nothing, but the Klein bottle [14]. This is a crucial property because, as we are going to see, in order to present a spinorial behavior, any space-time topological structure must somehow involve the Klein bottle. And of course, in order to be used as a model for any fundamental particle, a topological structure must necessarily be a space-time spinorial structure.

12.3 The KN Solution as a Dirac Particle

12.3.1 Preliminaries

We are going now to explore the possibility of using the KN solution as a model for the electron. To begin with, let us observe that the total internal angular momentum L of the KN solution, on either side of $\mathcal{M}$, can be written as

$$L = ma. \tag{12.4}$$

If we take for a, m and q the experimentally known electron values, and considering that, for a spin $1/2$ particle $L = 1/2$, it is easy to see that the disk has a diameter equal to the Compton wavelength $\lambda/2\pi = 1/m$ of the electron. Consequently, the angular velocity ω of a point in the singular ring turns out to be

$$\omega = 2m, \tag{12.5}$$

which corresponds to the so called *Zitterbewegung* frequency [15, 16] for a point-like electron orbiting a ring of diameter equal to λ_e. This means that the KN solution has a gyromagnetic ratio $g = 2$ [4, 12]. Due to this property, several attempts to model the electron by using the KN solution have been made. In most of these models, however, the circular singularity was always surrounded by a massive ellipsoidal shell (bubble), so that it was actually unreachable. In other words, the singularity was considered to be non-physical in the sense that the presence of the massive bubble would preclude its formation.

Using the extended interpretation of Hawking and Ellis, a different model has been proposed recently [11]. Its main property is that, differently from older models, it is represented by an empty KN solution, that is, no surrounding massive bubble is supposed to exist around the singular ring.[2] Instead, we make use of the excision procedure to circumvent the problems related to the naked singularity and the non-causal regions. The fundamental property of this model is that Wheeler's idea of "charge without charge" and "mass without mass" can be extended to spin. As a consequence, it is able to provide a topological explanation for the concepts of charge, mass, and spin.

Charge can be interpreted as arising from the multi-connectedness of the spatial section of the KN solution. In other words, we can associate the electric charge of the KN solution with the net flux of a topologically trapped electric field. In fact, remember that, from the point of view of an asymptotic observer, a trapped electric field is indistinguishable from the presence of a charge distribution. Then, in analogy with the geometry of the wormhole solution, there must exist a continuous path for each electric field line going from one space to the other. Furthermore, the equality of magnetic moment on both sides of $\mathcal{M}$ implies that the magnetic field lines must also be continuous when passing through the disk enclosed by the singularity.

Mass can be associated with the degree of nonflatness of the KN solution. It is given by Komar's integral [18],

$$m = \int_{\partial\Sigma} \star d\xi, \tag{12.6}$$

which holds for any stationary, asymptotically flat space-time. In this expression, $\star$ denotes the Hodge dual operator, ξ is the stationary Killing one-form of the background metric, and $\partial\Sigma$ is a spacelike surface of the background metric. It should be noticed that the mass m is the *total* mass of the system, that is, the mass-energy contributed by the gravitational and the electromagnetic fields [19].

Finally, spin can be consistently interpreted as an internal rotational motion of the singular ring. Of course, after the excision process, it turns out to be interpreted as an internal rotation of the infinitesimally-sized Klein bottle. It is important to remark that the KN solution is a singular ring in space-time, not in the 3D space. In fact, if the singularity were, let us say, in the xy plane, the angular momentum would be just a component of the *orbital* angular momentum, for which the gyromagnetic factor is well known to be $g = 1$. Since the gyromagnetic factor of the KN solution is $g = 2$, the rotation plane must necessarily involve the time axis. In fact, we know from Noether's theorem that conservation of *spin* angular momentum is related to the invariance of the system under a rotation in a plane involving the time axis.

[2] A similar approach has been used by Burinskii; see [17], and references therein.

12.3.2 Wave–Particle Duality

If one tries to compute the size of the KN particle, a remarkable result is obtained. To see it, we write down the spatial metric of the KN solution, which is given by [20]

$$dl^2 = \rho^2 \left[\frac{1}{\Delta} dr^2 + d\theta^2 + \frac{\Delta \sin^2\theta}{\Delta - a^2 \sin^2\theta} d\phi^2 \right]. \tag{12.7}$$

If we use this metric to compute the spatial length $\mathcal{L}$ of the singular ring, we find it to be zero:

$$\mathcal{L} \equiv \int_0^{2\pi} dl = 0. \tag{12.8}$$

This result is consistent with previous analysis made by some authors [8, 12], who pointed out that an external observer is unable to "see" the KN solution as an extended object, but only as a point-like object. We can then say that the "particle" concept is validated in the sense that the nontrivial KN structure is seen, by all observers, as a point-like object. Although the spatial dimension of the disk is zero, its space-time dimension is of the order of the Compton wavelength for the particle, which for the electron is $\lambda = 10^{-11}$ cm.

It is important to remark that, since the KN object appears as a point-like object, observers in different Lorentz frames will see the same point-like particle. However, each observer will see the particle with a different Compton wavelength, that is, a wave-packet with a different length. In other words, each observer will see a wave-packet with a different frequency, as predicted by special relativity. This can be interpreted as a classical realization of the wave-particle duality.

12.3.3 Topological Structure

A simple analysis of the structure of the extended KN metric shows that it is possible to isolate *four* physically nonequivalent states on each side of $\mathcal{M}$, that is, on $\mathbf{M}$ and on $\mathbf{M}'$. These states can be labeled by the sense of rotation (a can be positive or negative), and by the sign of the electric charge (positive or negative). Each one of these nonequivalent states in $\mathbf{M}$ must be joined continuously through the KN disk to another one in $\mathbf{M}'$, but with opposite charge. Since we want a continuous joining of the metric components, this matching must take into account the sense of rotation of the rings. In Fig. 12.3, just as in Fig. 12.1, the tubular joining between $\mathbf{M}$ and $\mathbf{M}'$ are drawn for one specific value of the electric charge,[3] but taking into account the different spin directions in each disk, which are drawn as small arrows. The differences among the configurations are the orientation of the spin vector and the geometry of the tubes.

---

[3] Two signs for the electric charge q in $\mathbf{M}$ or $\mathbf{M}'$ are allowed since the KN metric depends quadratically on q.

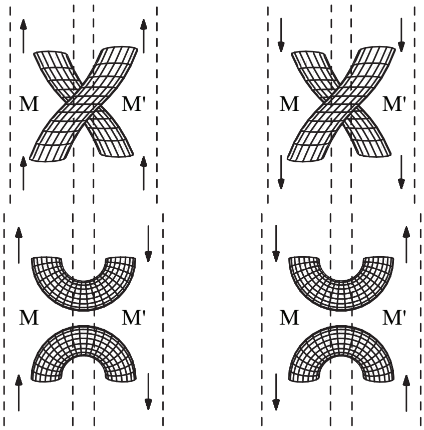

Fig. 12.3 The four possible geometric configurations of KN states for a specific value of the electric charge. The arrows indicate the sense of the spin vector.

It is important to remark that the model considers both sides of the solution, that is, $\mathbf{M}$ and $\mathbf{M}'$, as part of a single space-time.[4] The use of two space-times is just a mathematical necessity to describe the topological structure behind the KN solution. The question then arises on how to interpret the fact that the mass, and consequently the energy, acquires a negative value in $\mathbf{M}'$, if they are assumed to be positive in $\mathbf{M}$. The same happens with the sense of rotation, or equivalently, with the arrow of time. At this point it is possible to see the close analogy that exists between the topological structure of the KN solution and the structure of a Dirac spinor. In fact, the same questions on the interpretation of $\mathbf{M}$ and $\mathbf{M}'$ could be made on the interpretation of the two upper and the two lower components of the Dirac spinor. The answer to the latter question, as is well known, requires both special relativity and quantum mechanics, and consequently the notion of anti-particles to comply with negative energies [21]. We can then say that the necessity of two space-times to describe a spinorial structure in space-time is quite similar to the necessity of a four-component spinor to describe a spin-half particle.

12.3.4 Existence of Space-time Spinorial Structures

The excision process used to eliminate the noncausal region gives rise to highly nontrivial topological structure. Now, it is a well-known result that, in order to exhibit gravitational states with half-integral angular momentum, a 3-manifold must fulfill certain topological conditions. These conditions were stated by Friedman and Sorkin [22], whose results were obtained from a previous work by Hendricks [23] on the obstruction theory in 3D. Interesting enough, the KN solution can be show to satisfy these conditions, which means that it is actually a space-time spinorial structure [11].

[4] This is similar to the wormhole solution, which connects two points of the same space-time.

An alternative way to verify this result is to analyze the behavior of the KN topological structure under rotations. In general, when rotated by 2π, a classical object returns to its initial orientation. However, the topological structure of the KN solution presents a different behavior: it returns to its initial position only after a 4π rotation. This result can be understood from the topology of the 2D surface that is formed in the excision and gluing procedure. This surface, as we have already seen, is just a Klein bottle. A 2π rotation of the positive r side is equivalent to moving a point on the Klein bottle surface halfway from its initial position. Only after a 4π rotation it returns to its departure point. This is a well known property of Möbius strip, and consequently of the Klein bottle since the latter is obtained by a topological product of two Möbius strips.

12.3.5 Evolution Equation

As we have seen, the extended KN solution represents a space-time spinorial structure. It can, therefore, be naturally represented in terms of spinor variables of the Lorentz group SL(2, **C**). A crucial point towards this possibility is the fact that the KN solution presents four nonequivalent states, defined by the sense of rotation and by the sign of the electric charge. Since a Dirac spinor also has four independent components, it is not difficult to find an algebraic representation for the KN solution. Considering then an asymptotic observer in a Lorentz frame moving with a constant velocity, the evolution of the KN state vector is found to be governed by the Dirac equation [11]. Taking into account that the KN solution represents a space-time spinorial structures, we can say this is a natural and expected result.

12.4 Concluding Remarks

By using the extended space-time interpretation of Hawking and Ellis, together with Wheeler's idea of "charge without charge", the KN solution was shown to exhibit properties that are quite similar to those presented by an electron. Apart from the eventual importance of this result for particle physics, there is also deep consequence for the concept of space-time. At the early times of gravitation theory, space was considered simply an arena where all phenomena would take place. In other words, space was just a relation between the existing objects; without objects, there would be no space. Later on, the existence of an aether was considered, which in a sense would give some reality to the space. Since all experiments to detect such aether gave null results, space continued for some time to be this *mysterious nothing* in which we live in.

The advent of special relativity introduced the first important changes in our concept of space. Time lost its absolute character, and became just one more coordinate. Instead of living in a 3D space, we discovered that we actually live in a 4D space-time. The advent of general relativity introduced further and deeper conceptual changes in our notion of space-time. We discovered, for example, that space-time can storage energy. This means essentially that it could not anymore be interpreted as a simple arena because, if it can storage energy, it must have a concrete existence.

In addition to simple configurations, like a curved space-time, general relativity allows the existence of much more complex space-time structures. One example is the KN solution of Einstein's equation, which presents a very peculiar topological structure. Its main property is to be a spinorial space-time structure, which is revealed by the fact that only after a 4π rotation it returns to its initial position. The presence of the Klein bottle in the topological structure makes it easier to understand this property.

Now, if we consider that the topological structure is able trap an electric field, an asymptotic observer would see it as if the structure presented an electric charge. Furthermore, because the curved space-time associated to the topological structure has a non-vanishing energy, the same asymptotic observer would see it as if the structure presented a mass. When the experimental values for the electron charge and mass are used, the angular momentum of the KN solution is found to present a gyromagnetic factor $g = 2$. In addition to storage energy, therefore, space-time can also carry electric charge and spin angular momentum.

Due to the fact that it represents a space-time spinorial structure, the KN solution can be represented in terms of the spinor variables of the Lorentz group SL(2, **C**). Its space-time evolution is then naturally found to be governed by the Dirac equation. The KN structure, therefore, can be interpreted as a space-time-rooted electron model. Of course, it is not a finished model, and many points remain to be understood and clarified. For example, it is an open question whether it is applicable or not to other particles of nature. If, however, it shows to be a viable model, space-time will acquire a new and more important status. In fact, it will be not only the arena, but will also provide – through its highly nontrivial Planck-scale topological structures – the building blocks of all existing matter in the Universe, including ourselves.

Acknowledgments The authors would like to thank A. Burinskii, T. Nieuwenhuizen and G. 't Hooft for useful comments. They would like to thank also FAPESP-Brazil, CNPq-Brazil, and CAPES-Brazil for financial support.

References

1. M. B. Green, J. H. Schwarz, and E. Witten: *Superstring Theory* (Cambridge University Press, Cambridge, 1988).
2. J. A. Wheeler: *Geometrodynamics* (Academic Press, New York, 1962).
3. R. P. Kerr: Phys. Rev. Lett. **11**, 237 (1963).
4. E. T. Newman and A. I. Janis: J. Math. Phys. **6**, 915 (1965).
5. E. T. Newman et al: J. Math. Phys. **6**, 918 (1965).

6. C. A. Lopez: Phys. Rev. **D30**, 313 (1984); C. A. Lopez: Gen. Rel. Grav. **24**, 285 (1992).
7. M. Israelit and N. Rosen: Gen. Rel. Grav. **27**, 153 (1995).
8. W. Israel: Phys. Rev. **D2**, 641 (1970).
9. A. Burinskii: Sov. Phys. JETP **39**, 193 (1974).
10. S. W. Hawking and G. F. R. Ellis: *The Large Scale Structure of Space-Time* (Cambridge University Press, Cambridge, 1973) p. 161.
11. H. I. Arcos and J. G. Pereira: Gen. Rel. Grav. **36**, 2441 (2004) [hep-th/0210103]
12. B. Carter: Phys. Rev. **174**, 1559 (1968).
13. B. Punsly: J. Math. Phys. **28**, 859 (1987).
14. M. P. do Carmo: *Differential Geometry of Curves and Surfaces* (Prentice-Hall, New Jersey, 1976).
15. A. O. Barut and A. J. Bracken: Phys. Rev. **D23**, 2454 (1981).
16. A. O. Barut and W. Thacker: Phys. Rev. **D31**, 1386 (1985).
17. A. Burinskii: Phys. Rev. **D68**, 105004 (2003).
18. A. Komar: Phys. Rev. **113**, 934 (1959).
19. H. Ohanian and R. Ruffini: *Gravitation and Space-time* (Norton & Company, New York, 1994) p. 396.
20. L. D. Landau and E. M. Lifshitz: *The Classical Theory of Fields* (Pergamon, Oxford, 1975).
21. S. Weinberg: *Gravitation and Cosmology* (Wiley, New York, 1972), page 61.
22. J. L. Friedman and R. Sorkin: Phys. Rev. Lett. **44**, 1100 (1980).
23. H. Hendriks: Bull. Soc. Math. France Memoire **53**, 81 (1977); see section 4.3

Chapter 13
Dynamical Emergence of Instantaneous 3-Spaces in a Class of Models of General Relativity

Luca Lusanna and Massimo Pauri

Abstract The Hamiltonian structure of general relativity (GR), for both metric and tetrad gravity in a definite continuous family of space-times, is fully exploited in order to show that: (i) the *Hole Argument* can be bypassed by means of a specific *physical individuation* of point-events of the space-time manifold M^4 in terms of the *autonomous degrees of freedom* of the vacuum gravitational field (*Dirac observables*), while the *Leibniz equivalence* is reduced to differences in the *non-inertial appearances* (connected to *gauge* variables) of the same phenomena. (ii) The chrono-geometric structure of a solution of Einstein equations for given, gauge-fixed, initial data (a *3-geometry* satisfying the relevant constraints on the Cauchy surface), can be interpreted as an *unfolding* in mathematical global time of a sequence of *achronal 3-spaces* characterized by *dynamically determined conventions* about distant simultaneity. This result stands out as an important *conceptual difference* with respect to the standard chrono-geometrical view of special relativity (SR).

13.1 Introduction

The fact that, in the common sense view of ordinary life, phenomena tend to be intuited and described in a nonrelativistic 3D framework independent of any observer, is more or less justified on the basis of the conjunction of the relative smallness of ordinary velocities compared to the velocity of light and of the neurophysiological capabilities of our brain concerning temporal resolution.

Within the physical description of the world furnished by special relativity theory (SR) in terms of the mathematical representation of spatiotemporal phenomena in Minkowski space-time, one is immediately confronted with the problem of defining the 3D instantaneous space in which ordinary phenomena should be described.

As well known, the only possible resolution of this problem is based on the *relativization* of the description in relation to each observer, as ideally represented by a timelike worldline. This *ideal* observer chooses an arbitrary *convention* for the

V. Petkov (ed.), *Relativity and the Dimensionality of the World*, 229–260.

synchronization of distant clocks, namely an arbitrary foliation of space-time with spacelike 3-surfaces: the instantaneous 3-spaces identified by the convention. By exploiting any monotonically increasing function of the worldline proper time and by defining 3-coordinates having their origin on the worldline on each simultaneity 3-surfaces (i.e. a system of radar 4-coordinates), the observer builds either an inertial or a non-inertial frame, instantiating an observer- and frame-dependent notion of 3-space; see Ref. [1] for the contemporary treatment of synchronization and time comparisons in relativistic theories.

In the special case of inertial observers, the simplest way to define distant simultaneity (of a given event with respect to the observer) as well as to coordinatize space-time is to adopt the so-called Einstein convention and exploit two-way light signals with a single clock. Clearly, any two different observers must adopt the same convention in order to describe phenomena in a coherent way (essentially two different origins within the same frame).

The absolute chrono-geometrical structure of SR, together with the existence of the conformal structure of light cones (Lorentz signature) leads to the necessity of looking at Minkowski space-time as a whole 4D unit.

The way of dealing with such problems within general relativity (GR) has been considered until now as embodying a further level of complication because of the following facts:

(i) The universal nature of gravitational interaction
(ii) The fact that the whole inertio-gravitational and chrono-geometrical structure is jointly determined by the metric field tensor g
(iii) The fact that, unlike SR, in GR we have a system of partial differential equation for the dynamical determination of the chrono-geometrical structure of space-time
(iv) The fact that the symmetry group of the theory is no longer a Lie group like in SR but is the infinite group of diffeomorphisms in a pseudo-Riemannian 4D differentiable manifold M^4. This fact, which expresses the *general covariance* of the theory, concomitantly gives rise to the so-called *Hole phenomenology* (from the famous *Hole Argument*, formulated by Einstein in 1913, see Ref. [2]), which apparently (in Einstein's words, see Ref. [3]) "Takes away from space and time the last remnant of physical objectivity". Furthermore, it renders the Einstein Lagrangian *singular* with the consequence that Einstein's equations do not constitute a hyperbolic system of partial differential equations, a fact that makes the Cauchy initial value problem almost intractable in the configuration space M^4
(v) The absence of global inertial systems, which is a *global consequence* of the equivalence principle. Consequently, in topologically trivial Einstein's space-times, the only globally existing frames must be non-inertial
(vi) The necessity of selecting *globally hyperbolic* space-times in order to get a notion of *global mathematical time* (to be replaced by a physical clock in experimental practice) and avoid the so-called *problem of time* [4]

In this paper we will show that, at least for a definite continuous family of models of GR – analyzed within the Hamiltonian framework – it is possible to accomplish the following program:

(i) The metric field is naturally split into two distinct parts:

  (ia) An *epistemic* part, corresponding to the arbitrary constituents of the metric field (*gauge* variables) that must be completely fixed in order that the Hamilton–Einstein equations became a well-defined hyperbolic system. This complete gauge-fixing defines a *global non-inertial spatiotemporal frame* (called NIF) in which the true dynamics of the gravitational field must be described with all of the *generalized non-inertial effects* made explicit. Once the NIF is fixed, the standard passive 4-diffeomorphisms are subdivided in two classes: those adapted to the NIF, and those which are nonadapted: these latter modify only the 4-coordinates in a way that is not adapted to the NIF.

  (ib) An *ontic part*, corresponding to the *autonomous degrees of freedom* (2+2) of the vacuum gravitational field (*Dirac observables*, henceforth called DOs) expressed in that NIF.

(ii) A *physical individuation* of the point-events in M^4 can be obtained in terms of the DOs in that NIF, an individuation that downgrades the philosophical bearing of the *Hole Argument* (see later). In Ref. [5] we have shown that matter does contribute indirectly to the procedure of physical individuation, and we have suggested how this conceptual individuation could in principle be implemented with a well-defined empirical procedure, as a three-step experimental setup and protocol for positioning and orientation.

(iii) A careful reading of the Hamiltonian framework leads to the conclusion that the dynamical nature of the chrono-geometrical structure of every Einstein space-time (or *universe*) of the family considered entails the existence of a *dynamically determined convention* about distant simultaneity. This is tantamount to saying that every space-time of the family considered is *dynamically* generated in terms of a substructure of embedded *instantaneous 3-spaces* that foliates M^4 and defines an associated NIF (modulo gauge transformations, see later). More precisely, once the Cauchy data, i.e. a *3-geometry* satisfying the relevant constraints, are assigned in terms of the DOs on a *initial Cauchy surface*, *the solution of the Einstein–Hamilton equations embodies the unfolding in mathematical global time of a sequence of instantaneous 3-spaces identified by the Cauchy data chosen*. As a matter of fact, such 3-spaces are obtained, through the solution of an inverse problem, from the extrinsic curvature 3-tensor associated with the 4-metric tensor, solution of the equations in the whole space-time M^4.

Consequently, in a given Einstein space-time, there is *a preferred dynamical convention* for clock synchronization that should be used by every ideal observer (timelike world line). In practice, since the experimentalist is not aware of what

Einstein space-time he lives in, at the beginning the GR observers exploit arbitrary conventions like in SR. Actually, it should be noted that in GR, as in SR, the effective clock synchronization and the setting of a 3-space grid of coordinates is realized by *purely chrono-geometrical means*.[1] However, the preferred convention can be at least locally identified by making a measurement of g in the 4-coordinate system of the convention chosen and by solving the inverse problem. In this way such GR observers could identify the 3-spaces and resynchronize the clocks. Under the up-to-now confirmed assumption that the real space-time is an Einstein space, rather than a Weyl space, so that there is no *second clock effect* (see Ref. [6]), all of the clocks should maintain their synchronization on every instantaneous 3-space. Of course, a true physical coordinatization would require a dynamical treatment of realistic matter clocks.

Finally, it must be stressed that, unlike the situation of temporal ordering in SR, the unfolding of the 3-spaces constitutes here a unique *universal B-series ordering* of point-events.[2] Actually this holds true despite the fact that the stratification of M^4 in *achronal 3-spaces* is not gauge-independent. The point is that, *on-shell*, every dynamically admissible gauge transformation is the passive view of an active diffeomorphism within a *definite Einstein's universe*: it changes the NIF, the Hamiltonian (with the *tidal* and *inertial* effects), the world lines of material objects if present, and the same physical individuation of point-events (see later), in such a way that the *temporal order* of *any pair* of point-events and the *identification* of different material objects as to their relative *order* in space-time are not altered.

Let us conclude with a few philosophical remarks or, rather, specifications. In this paper we only deal with the *theoretical properties* of space-time or space + time, *as mathematically represented* in the three main space-time theories formulated in modern physics, namely, the Newtonian, the special and the general-relativistic. Accordingly, we do not take issue here with such themes as temporal *becoming* (absolute or relational), tensed or de-tensed *existence*, viewed as philosophical questions [7] connected in particular to special or general relativistic theories. Likewise, we exclude from the beginning any statement concerning cosmological issues, either Newtonian or general-relativistic. Above all, we are not concerned with the issue of an alleged *reification* of relativistic space-time as a *real* 4D continuum or a *reification* of the 3D Newtonian space as the *Raum* of our experience or phenomenal space in its various facets. We do believe that such purposes are philosophically misled and grounded on an untutored conflation of autonomous metaphysical issues with a literal interpretation of physical theories.

---

[1] Note that in GR one cannot ascertain in a simple way the usual relativistic effects concerning, e.g. rods contraction and time dilation. In order to show such effects one should have to exploit nonadapted coordinates restricted to a small worldtube in which SR could be considered as a good approximation.

[2] A B-series of temporal determinations concerning events is characterized by purely *relational statements* like "before than", "after than", and "simultaneous with". By contrast, an A-series is characterized by *monadic* attributes of single events, like "future", "present", and "past". A-series sentences and their truth-values depend upon the temporal perspective of the utterer, while B-series sentences have truth-values that are time-independent.

Certainly in Minkowski space-time there is no absolute fact of the matter as to which an event is *present*. Yet there is no absolute fact of the matter about presentness of events in Newtonian absolute time either, as there is no fact of the matter about presentness in *any* physical *theory* (though of course not in the physicists' practice !). Being valid by assumption at any time,[3] a physical theory cannot have the capacity of singling out a particular moment as "the present." Likewise no *physical experiment* can be devised having the capacity of telling whether a particular time signed by the hand of a clock is "the present" or not. This kind of knowledge requires the *conscious awareness* of a living subject, see Ref. [8]. However, no living observer can be forced within the Minkowski space-time (or any general-relativistic space-time model). No living observer can be there to collect in a *factually possible* (specifically *causal*) way the infinite amount of information spread on the space-like Cauchy surface which is necessary to solve the initial value problem according to well-known mathematical theorems and thereby defining the attributes of physical events. There is no living observer who can act to selectively generate a situation in their environment so that this situation, as a cause, will, according to their causal knowledge, give rise later with great probability to the effect which is desirable to them. Finally, there can be no living observer with the freedom to check the very empirical truth of the physical theory itself. Any alleged reification leads to a notion of the world which includes everything, in particular the object of an action and the agent of the same action. *This world, however, is a nonfactual world that nobody can "observe," study or control.*

These limitations, which are intrinsic to the nature of the scientific image, can be easily misunderstood. For example, the fact that Minkowski space-time must be considered as a whole 4D unit gave origin to the *prima facie* appealing but misleading notion of *block* universe, which seemed to entail that in Minkowski space-time there be no *temporal change*, on the grounds that there cannot be any *motion in time*. There is certainly no motion in time in the *ordinary sense of the term*, but *change* is there even if necessarily described in terms of the tenseless language of a (observer-dependent) *B-series*.

The issue of motion in our context can be seriously misinterpreted. The *ordinary sense* of the term – the one used in particular by the experimentalists in their laboratory – refers to the observation that there is some object "moving." Since this "moving" always takes place in "the present" of the experimentalist, as all other phenomena "taking place," one could be easily misled to believe that the *physical essence* of the concept of motion could be captured by the *experience* of an object as "moving," more or less on the same footing in which it is often asserted that there is a "moving now." Of course, there are moments of the motion history of the object that are *now* "past" and characterized as moments of a "remembered present." However, if the notion is no longer taking place "now," we can only describe the former motion as a purely B-series sequence of positions of the object at different times (of a clock), a description which expresses exactly the only *objective physical meaning* of "motion." Accordingly, since there is *no transiency* of the "now"

[3] Things seem a bit more complicated in cosmology, yet not substantially.

in Minkowski space-time, only *tenseless senses* of the words "becoming," "now," and even of "motion" are admitted in the relativistic idiom. Certainly, the ordinary sense of "motion" is perfectly legitimate in the *practice* of physics and, therefore, in the idiom of the experimentalist. This entails that *locally* and for a *limited length of time*, the experimentalist can project – so to speak – his practical view of the motion into Minkowski space-time as a mental aid for his intuition of the physical process. The experimentalist can consider, e.g. a world line of a point-mass endowed with a clock and, as time goes by, sign on the world line certain times indicated by his own synchronized physical clock in the laboratory, as a running chart. As long as the motion takes place, this can be a useful way of reasoning. It should be clear, however, that as soon as the limited allotted time has expired, the intuitive spatiotemporal representation of the motion given by the experimentalist is concluded and *nothing remains* in the Minkowski picture which is different by a section of a standard infinite world line with a B-series finite sequence of marked events.

On the other hand, this being said, could we suggest a specific case in connection with the so-called *endurantist/perdurantist* debate ? Briefly, the contrast can be roughly summarized as follows: the *endurantist* takes objects (including people) as lacking temporal extent and persisting by being *wholly present* at each moment of their history, while the *perdurantist* takes objects as persisting by being temporally extended and made up of different *temporal parts* at different times [9]. This opposition, which could appear *prima facie* as a mere matter of terminology at the level of the scientific image, is *formally* vindicated within the orthodox view of relativistic theories. Actually, if material objects were in any relevant sense 3D and persisted by occupying temporally unextended spatiotemporal regions, how could they fit in with the unavoidable 4D of relativistic space-time?

We believe, in general, that it is remarkably difficult, if not structurally unsound, to devise any conclusive argument from physics to metaphysical issues. It is true that 4D, as a philosophical stance, is sometimes used as a shield for perdurantism and 3D as a synonym for endurantism. Since, however, we are avowedly averse from any kind of *reification* of the relativistic models of space-time, we should specify the meaning, if any, of notions like *endurantism* or *perdurantism* once restricted to the idiom we deem admissible within the spatiotemporal scientific image. It is clear that, whatever meaning we are ready to allow for the notion of *object* as representable in Minkowski space-time, it cannot be *wholly present* in any sense. Things, however, are apparently different in GR, just in view of our results.

In our description of GR, the 4D spatiotemporal manifold is *dynamically foliated by gravitation* into global *achronal* 3-spaces at different global times. Therefore, the notion of a *wholly present material object* becomes *compatible* with an *endurantist* interpretation of temporal identity. Note that we are avowedly using the restricted adjective *compatible*, because we do not want to be committed to any specific philosophical stance about the issue of the *identity* of objects in general. We shall define a material object (say a *dust* filled sphere) as *wholly present* at a certain global time τ if all of the physical attributes of its constituent events can be obtained by physical information wholly contained in the structure of the 3-space at time τ. Note that such information contains in particular the 3-geometry of the leaf and all the relevant properties of the matter distribution that are also necessary for the formulation of the Cauchy problem of the theory. Having adopted this definition, we will return to the issue of the *endurantism/perdurantism* dispute, restricted to our formulation of GR, only at the end of the paper after expounding all the relevant theoretical features.

Finally, we should add an important remark: since, due to the universal nature of gravitation, SR should be carefully viewed as an approximation of GR rather than an *autonomous* theory, great part of the unending ontological debate about the issues of time, becoming, endurantism, perdurantism, etc. at the special relativistic level, should be reconsidered having in view the results we are going to discuss in the present paper.

In conclusion, here we are only interested in ascertaining whether and to what extent the notion of instantaneous 3-space is physically consistent and univocally definable. With this in view, while sketching the problem of distant simultaneity in the Newtonian and special-relativistic cases for the sake of argument, we will focus on the new and unexpected result concerning the natural, *dynamically ruled*, emergence of a notion of instantaneous 3-space in certain classes of models of GR.

13.2 Newton's Absolute Distant–Simultaneity

The absolute space of Newton is, by definition, an instantaneous 3-space at every value of absolute time. Leaving aside foundational problems of Newton and Galilei viewpoints, let us summarize the essential elements of the physicist's viewpoint about nonrelativistic mechanics.

The arena of Newton physics is Galilei space-time, in which both time and space have an absolute[4] status (its mathematical structure is the direct product $R^3 \times R$ and can be visualized as a foliation with base manifold the time axis and with Euclidean 3-spaces as fibers[5]). As a consequence, we have the absolute notions of simultaneity, instantaneous Euclidean 3-space and Euclidean spatial distance.

Space is a *container* of material bodies, i.e. objects endowed only with a (inertial) mass. The position of an object, unlike Newton tenets, is a relative frame-dependent notion. Furthermore, since there is an *absolute temporal distance* between events, while Newton's instantaneous 3-space is a *metric* space, Galilei's space-time has a *degenerate metric* structure.

Newton's first law, i.e. *Galilei law of inertia*, states that free objects move on straight lines, eliminating any intrinsic relevance of velocity.

Newton's second law, $\vec{F} = m\vec{a}$, identifies acceleration as the basic absolute quantity in the description of motion, where the force is intended to be measured statically.

---

[4] In various senses, the most important of which is the statement that time and space are entities independent of the dynamics.

[5] In physics it is always assumed that space, time, and space-time can be idealized as suitable mathematical manifolds possibly with additional structure.

Galilei relativity principle selects the inertial frames centered on inertial observers as the preferred ones due to the form-invariance of the second law under the Galilei transformations connecting inertial frames.

Gravity is described by an instantaneous action-at-a-distance interaction enjoying the special property of the equality of inertial and gravitational masses (*Galilei equivalence principle*).

In this absolutist point of view, the absolute existence of 3-space allows to develop a well-posed kinematics of isolated point-like N-body systems, which can be extended to rigid bodies (and then extended also to deformable ones, see for instance molecular physics). Given the positions $\vec{x}_i(t)$, $i-1,..,N$, of the bodies of mass m_i in a given inertial frame, we can uniquely define their *center of mass* $\vec{x}(t) = \sum_i m_i \vec{x}_i(t) / \sum_i m_i$, which describes a *decoupled pseudo-particle in inertial motion*, $\frac{d^2\, \vec{x}(t)}{dt^2} = 0$. All the dynamics is shifted to N-1 relative variables $\vec{r}_a(t)$, $a = 1,..,N-1$. At the Hamiltonian level, where the 3-velocities $\dot{\vec{x}}_i(t) = \frac{d\vec{x}_i(t)}{dt}$ are replaced by the momenta $\vec{p}_i(t) = m_i \dot{\vec{x}}_i(t)$, the separation of the center-of-mass conjugate variables ($\vec{x}(t)$, $\vec{p} = \sum_i \vec{p}_i(t) = const.$) from any set of conjugate relative variables $\vec{r}_a(t)$, $\vec{\pi}_a(t)$, is realized by a canonical transformation, which is a *point* transformation in the coordinates and the momenta separately.

The main property of the nonrelativistic notion of center of mass is that it can be determined locally in 3-space in the region occupied by the particles and does not depend on the complementary region of 3-space. Naively, one could say that if we eliminate the decoupled inertial pseudo-particle describing the center of mass, we shift to a *relational* description based on a set of *relative variables* $\vec{r}_a(t)$, $\vec{\pi}_a(t)$. However, as shown in Ref. [10], this is true only if the total barycentric angular momentum of the nonrelativistic universe is *zero*.[6] If the total angular momentum is different from zero, Newton relative motion in absolute 3-space satisfies equations of motion which are different from equations of motion having a purely relational structure [10]. This shows why the interpretations of the Newton rotating bucket are so different in the absolute and the relational descriptions.

Let us remark that, since inertial observers are idealizations, all realistic observers are (linearly and/or rotationally) accelerated, so that their spatial trajectories can be taken as the time axis of global (rigid or nonrigid) *non-inertial* frames. It turns out that in Newton's theory this leads only to the appearance of inertial forces proportional to the inertial mass of the accelerated body, rightly called fictitious (or apparent). Let us stress, on the other hand, that all realistic observers do experience

[6] This is connected with the fact that, while the 3-momentum Noether constants of motion satisfy an Abelian algebra at the Hamiltonian level, the angular momentum Noether constants of motion satisfy a non-Abelian one (only two of them, $\vec{J}^2$ and J_3, have vanishing Poisson brackets). As shown in Ref. [11] this explains why we can decouple the center of mass globally. On the contrary, there is no unique global way to separate 3 rotational degrees of freedom (see molecular dynamics [12]): this gives rise to *dynamical body frames* for deformable bodies (think of the diver and the falling cat). As well known and as shown in Ref. [13], in special relativity there is no notion of center of mass enjoying all the properties of the nonrelativistic ones. See Refs. [13, 14] for the types of nonrelativistic relative variables admitting a relativistic extension.

such forces and thereby have the problem of disentangling the real dynamical forces from the apparent ones.

13.3 Special Relativity: Conventional Distant Simultaneity

The arena of SR is Minkowski space-time, in which only the (3+1)-dimensional space-time is an absolute notion. Its Lorentz signature and its absolute chrono-geometrical structure allow to distinguish timelike, null, and spacelike intervals. However, given the world line of a timelike observer, in each point the observer can only identify the conformal structure of the incoming or outgoing rays of light (the past and future fixed light cone in that point): for the observer there is no intrinsic notion of simultaneous events, of instantaneous 3-space (to be used as a Cauchy surface for Maxwell equations), of spatial distance, or one-way velocity of light. The definition of instantaneous 3-space is then completely ruled by the conventions about distant simultaneity.

As well known, the starting point is constituted by the following basic postulates:

Two *light postulates* – The two-way (or round-trip) velocity of light (only one clock is involved in its definition) is (A) isotropic and (B) constant ($= c$).

The *relativity principle* (replacing the Galilei one) – It selects the relativistic inertial frames, centered on inertial timelike observers, and the Cartesian 4-coordinates x^μ, in which the line element is $ds^2 = \eta_{\mu\nu}\, dx^\mu\, dx^\nu$, $\eta_{\mu\nu} = \epsilon\,(+---)$, $\epsilon = \pm 1$ (according to particle physics or general relativity conventions, respectively).

The *law of inertia* states now that a test body moves along a flat timelike geodesics (a null one for a ray of light). The Poincaré group defines the transformations among the inertial frames. The preferred inertial frames are also selected by *Einstein's convention* for the synchronization of the clock of the inertial observer with any (in general accelerated) distant clock,[7] according to which the inertial instantaneous 3-spaces are the Euclidean space-like hyper-planes $x^o = ct = const.$ Only with this convention the one-way velocity of light between the inertial observer and any accelerated one coincides with the two-way velocity c.

The spatial distance between two simultaneous events in an inertial frame is the Euclidean distance along the connecting flat 3-geodesics.

[7] The inertial observer (γ) sends rays of light to another timelike observer γ_1, who reflects them back towards γ. Given the emission (τ_i) and adsorption (τ_f) times on γ, the point P of reflection on γ_1 is assumed to be simultaneous with the point Q on γ where $\tau_Q = \tau_i + \frac{1}{2}\,(\tau_f - \tau_i) = \frac{1}{2}\,(\tau_i + \tau_f) \stackrel{def}{=} \tau_P$. With this so-called *Einstein's $\frac{1}{2}$ convention for the synchronization of distant clocks*, the *instantaneous 3-space* is the spacelike hyper-plane $x^o = const.$ *orthogonal to* γ, the point Q is the midpoint between the emission and adsorption points and, since $\tau_P - \tau_i = \tau_f - \tau_P$, the *one-way* velocity of light between γ and every γ_1 is isotropic and equal to the round-trip velocity of light c. Note that in modern metrology clock synchronization is always performed by means of light signals and the velocity of light is assumed as a standard replacing the traditional unit of length.

Since inertial frames are still an idealization, we must consider the non-inertial ones centered on accelerated observers. As shown in Refs. [15] (see also Refs. [16, 17]) the standard 1+3 approach, trying to build non-inertial coordinates starting from the world line of the accelerated observers, meets coordinate singularities preventing their global definition.[8] Let us remember that the theory of measurements in non-inertial frames is based on the *locality principle* [18]: standard clocks and rods do not feel acceleration and at each instant the detectors of the instantaneously co-moving inertial observer give the correct data. Again this procedure fails in presence of electromagnetic fields when their wavelength is of the order of the acceleration radii [18] (the observer is not static enough during 5–10 cycles of such waves, so that their frequency cannot be measured).

The only known method to overcome these difficulties is to shift to the 3+1 point of view, in which, given the world line of the observer, one adds as an *independent structure* a 3+1 splitting of Minkowski space-time, which is nothing else than a clock synchronization convention. This allows to define a global non-inertial frame centered on the observer. This splitting foliates Minkowski space-time with spacelike hypersurfaces Σ_τ, which are the instantaneous (Riemannian) 3-spaces[9] associated to the given convention for clock synchronization (in general different from Einstein's). The leaves Σ_τ of the foliation are labeled by any scalar monotonically increasing function of the proper time of the observer. The intersection point of the observer world line with each Σ_τ is chosen as the origin of scalar curvilinear 3-coordinates. Such observer-dependent 4-coordinates $\sigma^A = (\tau; \sigma^r)$ are called *radar 4-coordinates*. Now the one-way velocity of light becomes, in general, both anisotropic and point-dependent, while the (Riemannian) spatial distance between two simultaneous points on Σ_τ is defined along the 3-geodesic joining them.

If $x^\mu = z^\mu(\tau, \vec{\sigma})$ describes the embedding of the associated simultaneity 3-surfaces Σ_τ into Minkowski space-time, so that the metric induced by the coordinate transformation $x^\mu \mapsto \sigma^A = (\tau, \vec{\sigma})$ is $g_{AB}(\tau, \vec{\sigma}) = \frac{\partial z^\mu(\sigma)}{\partial \sigma^A} \eta_{\mu\nu} \frac{\partial z^\nu(\sigma)}{\partial \sigma^B}$[10], the basic restrictions on the 3+1 splitting (leading to a nice foliation with space-like leaves) are the Møller conditions [15]

---

[8] Fermi coordinates, defined on hyper-planes orthogonal to the observer's 4-velocity become singular where the hyper-planes intersect, i.e. at distances from the world line of the order of the so-called linear and rotational *acceleration radii* ($\mathcal{L} = \frac{c^2}{a}$ for an observer with translational acceleration a; $\mathcal{L} = \frac{c}{\Omega}$ for an observer rotating with frequency Ω) [15, 18] (see also Ref. [19]). For rotating coordinates (rotating disk with the associated Sagnac effect) there is a coordinate singularity (the component g_{oo} of the associated 4-metric vanishes) at a distance from the rotation axis, where the tangential velocity becomes equal to c (the so-called horizon problem) [15].

[9] Let us stress that each instantaneous 3-space is a possible Cauchy surface for Maxwell equations. Namely, the added structure allows to have a well-posed initial value problem for these equations and to apply to them the theorem on the existence and uniqueness of the solutions of partial differential equations. The price to guarantee *predictability* is the necessity of giving Cauchy data on a noncompact spacelike 3-surface inside Minkowski space-time. This is the unavoidable element of *nonfactuality* which the 1+3 point of view would like to avoid.

[10] The 4-vectors $z^\mu_r(\tau, \vec{\sigma}) = \frac{\partial z^\mu(\tau, \vec{\sigma})}{\partial \sigma^r}$ are tangent to Σ_τ. If $l^\mu(\tau, \vec{\sigma})$ is the unit normal to Σ_τ (proportional to $\epsilon^\mu{}_{\alpha\beta\gamma}[z^\alpha_1 z^\beta_2 z^\gamma_3](\tau, \vec{\sigma})$), we have $z^\mu_\tau(\tau, \vec{\sigma}) = \frac{\partial z^\mu(\tau, \vec{\sigma})}{\partial \tau} = N(\tau, \vec{\sigma}) l^\mu(\tau, \vec{\sigma}) + N^r(\tau, \vec{\sigma}) z^\mu_r(\tau, \vec{\sigma})$, where $N(\tau, \vec{\sigma})$ and $N^r(\tau, \vec{\sigma})$ are the lapse and shift functions, respectively.

$$
\begin{aligned}
&\epsilon\, g_{\tau\tau}(\sigma) > 0,\\
&\epsilon\, g_{rr}(\sigma) < 0, \qquad \begin{vmatrix} g_{rr}(\sigma) & g_{rs}(\sigma) \\ g_{sr}(\sigma) & g_{ss}(\sigma) \end{vmatrix} > 0, \qquad \epsilon\, det\,[g_{rs}(\sigma)] < 0,\\
&\Rightarrow det\,[g_{AB}(\sigma)] < 0. \qquad (13.1)
\end{aligned}
$$

Furthermore, in order to avoid possible asymptotic degeneracies of the foliation, we must make the additional requirement that the simultaneity 3-surfaces Σ_τ must tend to spacelike hyper-planes at spatial infinity: $z^\mu(\tau,\vec{\sigma}) \rightarrow_{|\vec{\sigma}|\rightarrow\infty} x^\mu_s(\tau) + \epsilon^\mu_r\,\sigma^r$ and $g_{AB}(\tau,\vec{\sigma}) \rightarrow_{|\vec{\sigma}|\rightarrow\infty} \eta_{AB}$, with the ϵ^μ_r's being 3 unit spacelike 4-vectors tangent to the asymptotic hyperplane, whose unit normal is ϵ^μ_τ [the ϵ^μ_A form an asymptotic cotetrad, $\epsilon^\mu_A\,\eta^{AB}\,\epsilon^\nu_B = \eta^{\mu\nu}$].

As shown in Refs. [15, 16], Eqs. (13.1) forbid rigid rotations: *only differential rotations* are allowed (consistently with the modern description of rotating stars in astrophysics) and the simplest example is given by those 3+1 splittings whose simultaneity 3-surfaces are hyper-planes with rotating 3-coordinates described by the embeddings ($\sigma = |\vec{\sigma}|$)

$$
\begin{aligned}
&z^\mu(\tau,\vec{\sigma}) = x^\mu(\tau) + \epsilon^\mu_r\, R^r{}_s(\tau,\sigma)\,\sigma^s,\\
&R^r{}_s(\tau,\sigma) \rightarrow_{\sigma\rightarrow\infty} \delta^r_s, \qquad \partial_A\, R^r{}_s(\tau,\sigma) \rightarrow_{\sigma\rightarrow\infty} 0,\\
&R(\tau,\sigma) = \tilde{R}(\beta_a(\tau,\sigma)), \qquad \beta_a(\tau,\sigma) = F(\sigma)\,\tilde{\beta}_a(\tau), \quad a = 1,2,3,\\
&\frac{dF(\sigma)}{d\sigma} \neq 0, \qquad 0 < F(\sigma) < \frac{1}{A\,\sigma}. \qquad (13.2)
\end{aligned}
$$

Each $F(\sigma)$ satisfying the restrictions of the last line, coming from Eqs. (13.1), gives rise to a *global differentially rotating non-inertial frame*.

Since physical results in special relativity must not depend on the clock synchronization convention, a description including both standard inertial frames and admissible non-inertial ones is needed. This led to the discovery of *parametrized Minkowski theories*.

As shown in Refs. [20] (see also Refs. [15–17]), given the Lagrangian of every isolated system, one makes the coupling to an external gravitational field and then replaces the external metric with the $g_{AB}(\tau,\vec{\sigma})$ associated to a Møller-admissible 3+1 splitting. The resulting action principle $S = \int d\tau d^3\sigma\,\mathcal{L}(matter, g_{AB}(\tau,\vec{\sigma}))$ depends upon the system and the embedding $z^\mu(\tau,\vec{\sigma})$ and is invariant under frame-preserving diffeomorphisms: $\tau \mapsto \tau'(\tau,\vec{\sigma})$, $\sigma^r \mapsto \sigma'^r(\vec{\sigma})$. This *special-relativistic general covariance* implies the vanishing of the canonical Hamiltonian and the following 4 first-class constraints

$$
\begin{aligned}
\mathcal{H}_\mu(\tau,\vec{\sigma}) &= \rho_\mu(\tau,\vec{\sigma}) - \epsilon\, l_\mu(\tau,\vec{\sigma})\,\mathcal{M}(\tau,\vec{\sigma}) - \epsilon\, z_{r\mu}(\tau,\vec{\sigma})\, h^{rs}(\tau,\vec{\sigma})\,\mathcal{M}_s(\tau,\vec{\sigma}) \approx 0,\\
&\{\mathcal{H}_\mu(\tau,\vec{\sigma}), \mathcal{H}_\nu(\tau,\vec{\sigma}')\} = 0, \qquad (13.3)
\end{aligned}
$$

where $\rho_\mu(\tau,\vec{\sigma})$ is the momentum conjugate to $z^\mu(\tau,\vec{\sigma})$ and $[\sum_u h^{ru}\, g_{us}](\tau,\vec{\sigma}) = \delta^r_s$. $\mathcal{M}(\tau,\vec{\sigma}) = T_{\tau\tau}(\tau,\vec{\sigma})$ and $\mathcal{M}_r(\tau,\vec{\sigma}) = T_{\tau r}(\tau,\vec{\sigma})$ are the energy- and momentum-densities of the isolated system in Σ_τ-adapted coordinates [for N free particles we have $\mathcal{M}(\tau,\vec{\sigma}) = \sum_{i=1}^N \delta^3(\vec{\sigma} - \vec{\eta}_i(\tau)) \sqrt{m_i^2 + h^{rs}(\tau,\vec{\sigma})\, \kappa_{ir}(\tau)\, \kappa_{is}(\tau)}$, $\mathcal{M}_r(\tau,\vec{\sigma}) = \sum_{i=1}^N \delta^3(\vec{\sigma} - \vec{\eta}_i(\tau))\, \kappa_{ir}(\tau)$].

Since the matter variables have only Σ_τ-adapted Lorentz-scalar indices, the 10 constant of the motion corresponding to the generators of the external Poincaré algebra are

$$\begin{aligned} P^\mu &= \int d^3\sigma\, \rho^\mu(\tau,\vec{\sigma}), \\ J^{\mu\nu} &= \int d^3\sigma\, [z^\mu\, \rho^\nu - z^\nu\, \rho^\mu](\tau,\vec{\sigma}). \end{aligned} \tag{13.4}$$

The Hamiltonian gauge transformations generated by constraints of Eqs. (13.3) change the form and the coordinatization of the simultaneity 3-surfaces Σ_τ: as a consequence, the embeddings $z^\mu(\tau,\vec{\sigma})$ are *gauge variables*, so that in this framework the choice of the non-inertial frame and in particular of the convention for the synchronization of distant clocks [15, 16] is a *gauge choice*. All the inertial and non-inertial frames compatible with the Møller conditions of Eqs. (13.1) are *gauge equivalent* for the description of the dynamics of isolated systems.

A subclass of the embeddings $z^\mu(\tau,\vec{\sigma})$, in which the simultaneity leaves Σ_τ are equally spaced hyper-planes, describes the standard inertial frames if an inertial observer is chosen as origin of the 3-coordinates. Let us stress that every isolated system intrinsically identifies a special inertial frame, i.e. the rest frame. The use of radar coordinates in the rest frame leads to parametrize the dynamics according to the *Wigner-covariant rest-frame instant form of dynamics* developed in Refs. [20]. This instant form is a special case of *parametrized Minkowski theories* [20] [16],[11] in which the leaves of the 3+1 splitting of Minkowski space-time are inertial hyperplanes (simultaneity 3-surfaces called *Wigner hyper-planes*) orthogonal to the conserved 4-momentum P^μ of the isolated system.

The inertial rest-frame instant form is associated with the special gauge $z^\mu(\tau,\vec{\sigma}) = x_s^\mu(\tau) + \epsilon_r^\mu(u(P))\,\sigma^r$, $x_s^\mu(\tau) = Y_s^\mu(\tau) = u^\mu(P)\,\tau$, selecting the inertial rest frame of the isolated system centered on the Fokker–Pryce 4-center of inertia and having as instantaneous 3-spaces the Wigner hyperplanes.

Another particularly interesting family of 3+1 splittings of Minkowski space-time is defined by the embeddings

$$\begin{aligned} z^\mu(\tau,\vec{\sigma}) &= Y_s^\mu(\tau) + F^\mu(\tau,\vec{\sigma}) = u^\mu(P)\,\tau + F^\mu(\tau,\vec{\sigma}), \qquad F^\mu(\tau,\vec{0}) = 0, \\ &\to_{\sigma\to\infty} u^\mu(P)\,\tau + \epsilon_r^\mu(u(P))\,\sigma^r, \end{aligned} \tag{13.5}$$

with $F^\mu(\tau,\vec{\sigma})$ satisfying Eqs. (13.1).

[11] This approach was developed to give a formulation of the N-body problem on arbitrary simultaneity 3-surfaces. The change of clock synchronization convention may be formulated as a *gauge transformation* not altering the physics and there is no problem in introducing the electromagnetic field when the particles are charged (see Refs. [20–22]). The rest-frame instant form corresponds to the gauge choice of the 3+1 splitting whose simultaneity 3-surfaces are the intrinsic rest frame of the given configuration of the isolated system. See Ref. [23] for the Hamiltonian treatment of the relativistic center-of-mass problem and for the issue of reconstructing orbits in the 2-body case.

In this family the simultaneity 3-surfaces Σ_τ tend to Wigner hyper-planes at spatial infinity, where they are orthogonal to the conserved 4-momentum of the isolated system. Consequently, there are asymptotic inertial observers with world lines parallel to that of the Fokker–Pryce 4-center of inertia, namely there are the *rest-frame conditions* $p_r = \epsilon_r^\mu(u(P))\,P_\mu = 0$, so that the embeddings in Eqs. (13.5) define *global Møller-admissible non-inertial rest frames*.[12]

Since we are in non-inertial rest frames, the internal energy- and boost-densities contain the inertial potentials source of the relativistic inertial forces (see Ref. [24] for the quantization in non-inertial frames): more precisely, they are contained in the spatial components of the metric $g_{rs}(\tau,\vec{\sigma})$ associated to the embeddings of Eqs. (13.5).

In conclusion *the only notion of instantaneous 3-space* which can be introduced in special relativity is always observer $[X^\mu(\tau)]$- and frame $[\Sigma_\tau]$-dependent. The conceptual difficulties connected with the notion of relativistic center of mass also show that its definition (using the global Poincaré generators of the isolated system) *necessitates a whole instantaneous 3-space* Σ_τ. Therefore, even if we eventually get a decoupled pseudo-particle like in Newton theory,[13] we lose the possibility of treating disjoint, noninteracting, subsystems independently of one another. This is just due to the necessity of choosing a convention for the synchronization of distant clocks.

13.4 General Relativity: Dynamically Determined Distant Simultaneity

We will show that, contrary to a widespread opinion, general relativity – at least for a particular class of models[14] – contains in itself the capacity for a dynamical definition of instantaneous 3-spaces. Every model of GR in the given class, once completely specified in a precise sense that will be explained presently, entails that space-time *be* essentially the unfolding in time of a dynamically variable form of instantaneous 3-space.

In the years 1913–1916 Einstein developed general relativity by relying on the *equivalence principle* (equality of inertial and gravitational masses of nonspinning test bodies in free fall) and on the guiding principle of general covariance. Einstein's original view was that the principle had to express the impossibility of distinguishing a uniform gravitational field from the effects of a constant acceleration by means of local experiments in a sufficiently small region with negligible tidal forces. This led him to the *geometrization* of the gravitational interaction and to the replacement

[12] The only ones existing in tetrad gravity, due to the equivalence principle, in globally hyperbolic asymptotically flat space-times without super-translations as we shall see in the next section.

[13] However, the canonical transformations decoupling it from the relative variables are now non-point; only for free particles they remain point in the momenta, but not in the positions [23].

[14] Given the enormous variety of solutions of Einstein's equations, one cannot expect to find general answers to ontological questions.

of Minkowski space-time with a pseudo-Riemannian 4-manifold M^4 with non vanishing curvature Riemann tensor.

The *equivalence principle* entails the nonexistence of global inertial frames (SR relativity holds only in a small neighborhood of a body in free fall). The principle of *general covariance* (see Ref. [25] for a thorough review), which expresses the tensorial nature of Einstein's equations, has the following two consequences:

(i) The invariance of the Hilbert action under *passive* diffeomorphisms (the coordinate transformations in M^4), so that the second Noether theorem implies the existence of first-class constraints at the Hamiltonian level
(ii) The mapping of the solutions of Einstein's equations among themselves under the action of *active* diffeomorphisms of M^4 extended to the tensors over M^4 (*dynamical symmetries* of Einstein's equations)

The basic field of metric gravity is the 4-metric tensor with components ${}^4g_{\mu\nu}(x)$ in an arbitrary coordinate system of M^4. The peculiarity of gravity is that the 4-metric field, unlike the fields of electromagnetic, weak and strong interactions and the matter fields, has a *double role*:

(i) It is the mediator of the gravitational interaction (in analogy to all of the other gauge fields)
(ii) It determines *dynamically* the chrono-geometric structure of space-time M^4 through the line element $ds^2 = {}^4g_{\mu\nu}(x)\, dx^\mu\, dx^\nu$.

Consequently, the gravitational field *teaches relativistic causality* to all of the other fields: in particular it teaches to classical rays of light, photons, and gluons, which are the trajectories allowed for massless particles in each point of M^4.

Let us make a comment about the two main existing approaches for quantizing gravity.

(i) *Effective quantum field theory and string theory*. This approach contains the standard model of elementary particles and its extensions. However, since the quantization, namely, the definition of the Fock space, requires a background space-time for the definition of creation and annihilation operators, one must use the splitting ${}^4g_{\mu\nu} = {}^4\eta^{(B)}_{\mu\nu} + {}^4h_{\mu\nu}$ and quantize only the perturbation ${}^4h_{\mu\nu}$ of the background 4-metric $\eta^{(B)}_{\mu\nu}$ (usually B is either Minkowski or DeSitter space-time). In this way property (ii) is lost (one exploits the fixed non-dynamical chrono-geometrical structure of the background space-time), and gravity is replaced by a field of spin two over the background (and passive diffeomorphisms are replaced by a Lie group of gauge transformations acting in an "inner" space). The only difference between gravitons, photons, and gluons lies thereby in their quantum numbers.
(ii) *Loop quantum gravity*. This approach does not introduce a background space-time but, being inequivalent to a Fock space, has problems in incorporating particle physics. It exploits a fixed 3+1 splitting of the space-time M^4 and quantizes the associated instantaneous 3-spaces Σ_τ (quantum geometry). There is

no known way, however, to implement consistent unitary evolution (the problem of the super-Hamiltonian constraint). Furthermore, since the theory is usually formulated in spatially compact space-times without boundary, it admits no Poincaré symmetry group (and therefore no extra-dimensions as in string theory), and faces a serious problem concerning the definition of time: the so-called *frozen picture* without real evolution.

For outside points of view on loop quantum gravity and string theory see Ref. [26, 27], respectively.

Let us remark that all formulations of the theory of elementary particle and nuclear physics are a chapter of the theory of representations of the Poincaré group in the *inertial frames* of the spatially noncompact Minkowski space-time. As a consequence, if one looks at general relativity from the point of view of particle physics, the main problem to get a unified theory is that of conciliating the Poincaré group and the diffeomorphism group.

Let us now consider the ADM formulation of metric gravity [28] and its extension to tetrad gravity[15] obtained by replacing the 10 configurational 4-metric variables ${}^4g_{\mu\nu}(x)$ with the 16 cotetrad fields ${}^4E^{(\alpha)}_\mu(x)$ by means of the decomposition ${}^4g_{\mu\nu}(x) = {}^4E^{(\alpha)}_\mu(x)\,{}^4\eta_{(\alpha)(\beta)}\,{}^4E^{(\beta)}_\nu(x)$ [(α) are flat indices].

Then, after having restricted the model to globally hyperbolic, topologically trivial, spatially noncompact space-times (admitting a *global notion of time*), let us introduce a global 3+1 splitting of the space-time M^4 and choose the world line of a timelike observer. As in special relativity, let us make a coordinate transformation to observer-dependent *radar* 4-coordinates, $x^\mu \mapsto \sigma^A = (\tau, \sigma^r)$, adapted to the 3+1 splitting and using the observer world line as origin of the 3-coordinates. Again, the inverse transformation, $\sigma^A \mapsto x^\mu = z^\mu(\tau, \sigma^r)$, defines the embedding of the leaves Σ_τ into M^4. These leaves Σ_τ (assumed to be Riemannian 3-manifolds diffeomorphic to R^3, so that they admit global 3-coordinates σ^r and a unique 3-geodesic joining any pair of points in Σ_τ) are both Cauchy surfaces and *simultaneity surfaces* corresponding to a *convention for clock synchronization*. For the induced 4-metric we get

$$
\begin{aligned}
{}^4g_{AB}(\sigma) &= \frac{\partial z^\mu(\sigma)}{\partial \sigma^A}\,{}^4g_{\mu\nu}(x)\,\frac{\partial z^\nu(\sigma)}{\partial \sigma^B} = {}^4E^{(\alpha)}_A\,{}^4\eta_{(\alpha)(\beta)}\,{}^4E^{(\beta)}_B \\
&= \epsilon \begin{pmatrix} (N^2 - {}^3g_{rs}N^rN^s) & -{}^3g_{su}N^u \\ -{}^3g_{ru}N^u & -{}^3g_{rs} \end{pmatrix}(\sigma).
\end{aligned}
$$

Here ${}^4E^{(\alpha)}_A(\tau,\sigma^r)$ are adapted cotetrad fields, $N(\tau,\sigma^r)$ and $N^r(\tau,\sigma^r)$ the lapse and shift functions, and ${}^3g_{rs}(\tau,\sigma^r)$ the 3-metric on Σ_τ with signature $(+++)$. We see that, unlike in special relativity, in general relativity the quantities $z^\mu_A = \partial z^\mu/\partial\sigma^A$ are no more cotetrad fields on M^4. Here they are only transition functions

[15] This extension is needed to describe the coupling of gravity to fermions; it is a theory of timelike observers each one endowed with a tetrad field, whose timelike axis is the unit 4-velocity of the observer and whose spatial axes are associated with a choice of three gyroscopes.

between coordinate charts, so that the dynamical fields are now the real cotetrad fields ${}^4E^{(\alpha)}_A(\tau,\sigma^r)$ and not the embeddings $z^\mu(\tau,\sigma^r)$.

Let us try to identify a class of space-times and an associated family of admissible 3+1 splittings suitable to incorporate particle physics and provide a model for the solar system or our galaxy (and hopefully even allowing an extension to the cosmological context) with the following further requirements [29]:

(i) M^4 must be asymptotically flat at spatial infinity and the 4-metric must tend asymptotically to the Minkowski 4-metric there, in every coordinate system (this implies that the 4-diffeomorphisms must tend to the identity at spatial infinity). In such space-times, therefore, there is an *asymptotic background 4-metric* allowing to avoid the decomposition ${}^4g_{\mu\nu} = {}^4\eta_{\mu\nu} + {}^4h_{\mu\nu}$ in the bulk.

(ii) The boundary conditions on the fields on each leaf Σ_τ of the admissible 3+1 splittings must be such to reduce the *Spi* group of asymptotic symmetries (see Ref. [30]) to the ADM Poincaré group. This means that there should not be *super-translations* (direction-dependent quasi Killing vectors, obstruction to the definition of angular momentum in general relativity), namely, that all the fields must tend to their asymptotic limits in a direction-independent way (see Ref. [31]). This is possible only if the admissible 3+1 splittings have all the leaves Σ_τ tending to Minkowski space-like hyperplanes orthogonal to the ADM 4-momentum at spatial infinity [29]. In turn this implies that every Σ_τ is *the rest frame of the instantaneous 3-universe* and that there are asymptotic inertial observers to be identified with the *fixed stars*.[16] This requirement implies that the shift functions vanish at spatial infinity [$N^r(\tau,\sigma^r) \to O(1/|\sigma|^\epsilon)$, $\epsilon > 0$, $\sigma^r = |\sigma|\,\hat{u}^r$], where the lapse function tends to 1 [$N(\tau,\sigma^r) \to 1 + O(1/|\sigma|^\epsilon)$] and the 3-metric tends to the Euclidean 3-metric [${}^3g_{rs}(\tau,\sigma^u) \to \delta_{rs} + O(1/|\sigma|)$].

(iii) The admissible 3+1 splittings should have the leaves Σ_τ admitting a generalized Fourier transform (namely, they should be Lichnerowicz 3-manifolds [32] with involution). This would allow the definition of instantaneous Fock spaces in a future attempt of quantization.

(iv) All the fields on Σ_τ should belong to suitable weighted Sobolev spaces, so that M^4 has no Killing vectors and Yang–Mills fields on Σ_τ do not present Gribov ambiguities (due to the presence of gauge symmetries and gauge copies) [33].

In absence of matter the Christodoulou and Klainermann [34] space-times are good candidates: they are near Minkowski space-time in a norm sense, avoid singularity theorems by relaxing the requirement of conformal completability (so that it is possible to follow solutions of Einstein's equations on long times) and admit gravitational radiation at null infinity.

Since the simultaneity leaves Σ_τ are the rest frame of the instantaneous 3-universe, at the Hamiltonian level it is possible to define *the rest-frame instant form of metric and tetrad gravity* [29, 35]. If matter is present, the limit of this description for vanishing Newton constant will be reduced to the rest-frame instant form description of the same matter in the framework of parametrized Minkowski theories,

[16] In a final extension to the cosmological context they could be identified with the privileged observers at rest with respect to the background cosmic radiation.

while the ADM Poincaré generators will tend to the kinematical Poincaré generators of special relativity. In this way we get a model admitting *a deparametrization of general relativity to special relativity*. It is not known whether the rest-frame condition can be relaxed in general relativity without bringing in super-translations, since the answer to this question is connected with the nontrivial problem of boosts in general relativity.

Let us now come back to ADM tetrad gravity. The timelike vector ${}^4E^A_{(o)}(\tau,\sigma^r)$ of the tetrad field ${}^4E^A_{(\alpha)}(\tau,\sigma^r)$, dual to the cotetrad field ${}^4E^{(\alpha)}_A(\tau,\sigma^r)$, may be rotated to become the unit normal to Σ_τ in each point by means of a standard Wigner boost for timelike Poincaré orbits depending on three parameters $\varphi_{(a)}(\tau,\sigma^r)$, $a=1,2,3$: ${}^4E^A_{(o)}(\tau,\sigma^r) = L^A{}_B(\varphi_{(a)}(\tau,\sigma^r))\,{}^4\check{E}^B_{(o)}(\tau,\sigma^r)$. This allows to define the following cotetrads adapted to the 3+1 splitting (the so-called *Schwinger time gauge*) ${}^4\check{E}^{(o)}_A(\tau,\sigma^r) = \big(N(\tau,\sigma^r);0\big)$, ${}^4\check{E}^{(a)}_A(\tau,\sigma^r) = \big(N_{(a)}(\tau,\sigma^r);{}^3e_{(a)r}(\tau,\sigma^r)\big)$, where ${}^3e_{(a)r}(\tau,\sigma^r)$ are cotriads fields on Σ_τ (tending to $\delta_{(a)r}+O(1/|\sigma|)$ at spatial infinity) and $N_{(a)} = N^r\,{}^3e_{(a)r}$. As a consequence, the sixteen cotetrad fields may be replaced by the fields $\varphi_{(a)}(\tau,\sigma^r)$, $N(\tau,\sigma^r)$, $N_{(a)}(\tau,\sigma^r)$, ${}^3e_{(a)r}(\tau,\sigma^r)$, whose conjugate canonical momenta will be denoted as $\pi_N(\tau,\sigma^r)$, $\pi_{\vec{N}\,(a)}(\tau,\sigma^r)$, $\pi_{\vec{\varphi}\,(a)}(\tau,\sigma^r)$, ${}^3\pi^r_{(a)}(\tau,\sigma^r)$.

The local invariance of the ADM action entails the existence of 14 first-class constraints (10 primary and 4 secondary):

(i) $\pi_N(\tau,\sigma^r)\approx 0$ implying the secondary super-Hamiltonian constraint $\mathcal{H}(\tau,\sigma^r)\approx 0$;
(ii) $\pi_{\vec{N}\,(a)}(\tau,\sigma^r)\approx 0$ implying the secondary super-momentum constraints $\mathcal{H}_{(a)}(\tau,\sigma^r)\approx 0$;
(iii) $\pi_{\vec{\varphi}\,(a)}(\tau,\sigma^r)\approx 0$;
(iv) three constraints $M_{(a)}(\tau,\sigma^r)\approx 0$ generating rotations of the cotriads.

Consequently, there are 14 gauge variables, which, as shown in Refs. [5], describe the *generalized inertial effects* in the non-inertial frame defined by the chosen admissible 3+1 splitting of M^4 centered on an arbitrary timelike observer. The remaining independent "two + two" degrees of freedom are the gauge invariant DOs of the gravitational field describing *generalized tidal effects* (see Refs. [5]). The same degrees of freedom emerge in ADM metric gravity, where the configuration variables N, N^r, ${}^4g_{rs}$, with conjugate momenta π_N, $\pi_{\vec{N}\,r}$, ${}^3\Pi^{rs}$, are restricted by 8 first-class constraints ($\pi_N(\tau,\sigma^r)\approx 0 \to \mathcal{H}(\tau,\sigma^r)\approx 0$, $\pi_{\vec{N}\,r}(\tau,\sigma^r)\approx 0 \to \mathcal{H}^r(\tau,\sigma^r)\approx 0$).

As already said, the first-class constraints are the generators of the Hamiltonian gauge transformations, under which the ADM action is quasi-invariant (second Noether theorem):

(i) The gauge transformations generated by the four primary constraints $\pi_N(\tau,\sigma^r)\approx 0$, $\pi_{\vec{N}\,(a)}(\tau,\sigma^r)\approx 0$, modify the lapse and shift functions, namely how densely the simultaneity surfaces are packed in M^4 and which points have the same 3-coordinates on each Σ_τ.

(ii) Those generated by the three super-momentum constraints $\mathcal{H}_{(a)}(\tau,\sigma^r)\approx 0$ change the 3-coordinates on Σ_τ.
(iii) Those generated by the super-Hamiltonian constraint $\mathcal{H}(\tau,\sigma^r)\approx 0$ transform an admissible 3+1 splitting into another admissible one by realizing a normal deformation of the simultaneity surfaces Σ_τ [42]. As a consequence, *all the conventions about clock synchronization are gauge equivalent as in special relativity*.
(iv) Those generated by $\pi_{\vec{\varphi}\,(a)}(\tau,\sigma^r)\approx 0$, $M_{(a)}(\tau,\sigma^r)\approx 0$, change the cotetrad fields with local Lorentz transformations.

In the rest-frame instant form of tetrad gravity there are the three extra first-class constraints $P^r_{ADM}\approx 0$ (vanishing of the ADM 3-momentum as *rest-frame conditions*). They generate gauge transformations which change the timelike observer whose world line is used as origin of the 3-coordinates.

A fundamental technical point, which is of paramount importance for the physical interpretation, is the possibility of performing a separation of the gauge variables from the DOs by means of a Shanmugadhasan canonical transformation [36].

In Ref. [35] a Shanmugadhasan canonical transformation adapted to 13 first class constraints (not to the super-Hamiltonian one, because no one knows how to solve it except in the post-Newtonian approximation) has been introduced and exploited to clarify the interpretation. There are problems, however, when one introduces matter.

To avoid the above difficulties, a different Shanmugadhasan canonical transformation, adapted only to 10 constraints but allowing the addition of any kind of matter to the rest-frame instant form of tetrad gravity, has been recently found starting from a new parametrization of the 3-metric [37].

The basic idea is that the symmetric 3-metric tensor can be diagonalized with an orthogonal matrix depending on three Euler angles $\theta^i(\tau,\vec{\sigma})$. The three eigenvalues $\lambda_i(\tau,\vec{\sigma})$ are then replaced by the conformal factor $\phi(\tau,\vec{\sigma})$ of the 3-metric and by two *tidal* variables $R_{\bar{a}}(\tau,\vec{\sigma})$, $\bar{a}=1,2$. The defining equations and the resulting Shanmugadhasan canonical transformation are:

$$
\begin{aligned}
{}^3g_{rs} &= \sum_{uv} V_{ru}(\theta^n)\,\lambda_u\,\delta_{uv}\,V^T_{vs}(\theta^n)\\
&= \sum_a \Big(V_{ra}(\theta^n)\,\Lambda^a\Big)\Big(V_{sa}(\theta^n)\,\Lambda^a\Big) = \sum_a {}^3\bar{e}_{(a)r}\,{}^3\bar{e}_{(a)s} = \sum_a {}^3e_{(a)r}\,{}^3e_{(a)s},\\
\Lambda^a(\tau,\vec{\sigma}) &\stackrel{def}{=} \sum_u \delta_{au}\sqrt{\lambda_u(\tau,\vec{\sigma})} = \phi^2(\tau,\vec{\sigma})\,e^{\sum_{\bar{a}}\gamma_{\bar{a}b}R_{\bar{a}}(\tau,\vec{\sigma})}\\
&\quad \rightarrow_{r\rightarrow\infty}\ 1+\frac{M}{4r}+\frac{a^a}{r^{3/2}}+O(r^{-3}),\\
{}^3e_{(a)r} &= R_{(a)(b)}(\alpha_{(c)})\,{}^3\bar{e}_{(b)r},\\
{}^3\bar{e}_{(a)r} &= \sum_b {}^3e_{(b)r}\,R_{(b)(a)}(\alpha_{(e)}) = \sum_u \sqrt{\lambda_u}\,\delta_{u(a)}\,V^T_{ur}(\theta^n) = V_{ra}(\theta^n)\,\Lambda^a,
\end{aligned}
$$

$$
\begin{aligned}
{}^3\bar{e}^r_{(a)} &= \sum_u \frac{\delta_{u(a)}}{\sqrt{\lambda_u}}\, V_{ru} = \frac{V_{ra}(\theta^n)}{\Lambda^a},\\
\phi &= (det\, {}^3g)^{1/12} = ({}^3e)^{1/6} = {}^3\bar{e}^{1/6} = (\lambda_1\,\lambda_2\,\lambda_3)^{1/12} = (\Lambda^1\,\Lambda^2\,\Lambda^3)^{1/6}.
\end{aligned}
$$

$$
\begin{array}{|c|c|c|c|}
\hline
\varphi_{(a)} & n & n_{(a)} & {}^3e_{(a)r} \\
\hline
\approx 0 & \approx 0 & \approx 0 & {}^3\pi^r_{(a)} \\
\hline
\end{array}
\quad \longrightarrow \quad
\begin{array}{|cc|cc|c|}
\hline
\varphi_{(a)} & \alpha_{(a)} & n & \bar{n}_{(a)} & {}^3\bar{e}_{(a)r} \\
\hline
\approx 0 & \approx 0 & \approx 0 & \approx 0 & {}^3\tilde{\bar{\pi}}^r_{(a)} \\
\hline
\end{array}
$$

$$
\longrightarrow \quad
\begin{array}{|cc|cc|cc|}
\hline
\varphi_{(a)} & \alpha_{(a)} & n & \bar{n}_{(a)} & \theta^r & \Lambda^r \\
\hline
\approx 0 & \approx 0 & \approx 0 & \approx 0 & \pi^{(\theta)}_r & P_r \\
\hline
\end{array}\ ,
$$

$$
\begin{array}{|c|}
\hline
\Lambda^r \\
\hline
P_r \\
\hline
\end{array}
\quad \longrightarrow \quad
\begin{array}{|c|c|}
\hline
\phi & R_{\bar{a}} \\
\hline
\pi_\phi & \Pi_{\bar{a}} \\
\hline
\end{array}\ ,
$$

$$
\pi_\phi = -\epsilon\, \frac{c^3}{2\pi G}\, ({}^3\bar{e})^{5/6}\ {}^3K = 2\, \frac{\sum_b \Lambda^b P_b}{(\Lambda^1\Lambda^2\Lambda^3)^{1/6}}, \tag{13.6}
$$

where $\bar{n}_{(a)} = \sum_b n_{(b)}\, R_{(b)(a)}(\alpha_{(e)})$ are the shift functions at $\alpha_{(a)}(\tau,\vec{\sigma}) = 0$, $\alpha_{(a)}(\tau,\sigma^r)$ are three Euler angles and $\theta^r(\tau,\sigma^r)$ are three angles giving a coordinatization of the action of 3-diffeomorphisms on the cotriads ${}^3e_{(a)r}(\tau,\sigma^r)$. The configuration variable $\phi(\tau,\sigma^r) = \Big(det\, {}^3g(\tau,\sigma^r)\Big)^{1/12}$ is the conformal factor of the 3-metric: it can be shown that it is the unknown quantity in the super-Hamiltonian constraint (also named the Lichnerowicz equation). The gauge variables are n, $\bar{n}_{(a)}$, $\varphi_{(a)}$, $\alpha_{(a)}$, θ^r and π_ϕ, while $R_{\bar{a}}$, $\Pi_{\bar{a}}$, $\bar{a} = 1,2$, are the DOs of the gravitational field (in general they are not tensorial quantities).

This canonical transformation is the first explicit construction of a York map [38], in which the momentum conjugate to the conformal factor (*the gauge variable controlling the convention for clock synchronization*) is proportional to the trace ${}^3K(\tau,\vec{\sigma})$ of the extrinsic curvature of the simultaneity surfaces Σ_τ. Both the *tidal* and the *gauge* variables can be expressed in terms of the original variables. Moreover, in a family of completely fixed gauges differing with respect to the convention of clock synchronization, the deterministic Hamilton equations for the *tidal* variables and for *matter* variables contain *relativistic inertial forces* determined by ${}^3K(\tau,\vec{\sigma})$, which change from attractive to repulsive where the trace changes sign. These inertial forces do not have a nonrelativistic counterpart (the Newton 3-space is absolute) and could perhaps support the proposal of Ref. [39][17] according to which *dark matter* could be explained as an *inertial effect*. While in the MOND model [41] there is an arbitrary function on the acceleration side of Newton equations in the absolute

[17] The model proposed in Ref. [39] is too naive, as shown by the criticism in Ref. [40].

Euclidean 3-space, here we have the arbitrary gauge function ${}^3K(\tau,\vec{\sigma})$ on the force side of Hamilton equations.

Finally let us see which Dirac Hamiltonian H_D generates the τ-evolution in ADM canonical gravity. In *spatially compact space-times without boundary* H_D is a linear combination of the primary constraints plus the secondary super-Hamiltonian and super-momentum constraints multiplied by the lapse and shift functions, respectively (a consequence of the Legendre transform). Consequently, $H_D \approx 0$ and, in the reduced phase space, we get a vanishing Hamiltonian. This implies the so-called *frozen picture* and the problem of how to reintroduce a temporal evolution.[18] Usually one considers the normal (timelike) deformation of Σ_τ induced by the super-Hamiltonian constraint as evolution in a local time variable to be identified (the "multi-fingered" time point of view with a local, either extrinsic or intrinsic, time): this is the so-called *Wheeler–DeWitt interpretation*.[19]

On the contrary, *in spatially noncompact space-times* the definition of functional derivatives and the existence of a well-posed Hamiltonian action principle (with the possibility of a good control of the surface terms coming from integration by parts) require the addition of the *DeWitt surface term* [44] (living on the surface at spatial infinity) to the Hamiltonian. It can be shown [29] that in the rest-frame instant form this term, together with a surface term coming from the Legendre transformation of the ADM action, leads to the Dirac–Hamiltonian

$$H_D = \check{E}_{ADM} + (constraints) = E_{ADM} + (constraints) \approx E_{ADM}. \tag{13.7}$$

Here $\check{E}_{ADM}$ is the *strong ADM energy*, a surface term analogous to the one defining the electric charge as the flux of the electric field through the surface at spatial infinity in electromagnetism. Since we have $\check{E}_{ADM} = E_{ADM} + (constraints)$, we see that the nonvanishing part of the Dirac Hamiltonian is the *weak ADM energy* $E_{ADM} = \int d^3\sigma\, \mathcal{E}_{ADM}(\tau,\sigma^r)$, namely, the integral over Σ_τ of the ADM energy density (in electromagnetism this corresponds to the definition of the electric charge as the volume integral of matter charge density). Therefore there is no frozen picture but a consistent τ-evolution instead.

Note that the ADM energy density $\mathcal{E}_{ADM}(\tau,\sigma^r)$ is a *coordinate-dependent quantity*, because it depends on the gauge variables (namely upon the relativistic inertial effects present in the non-inertial frame): this is nothing else than the old *problem of energy* in general relativity. Let us remark that in most coordinate systems $\mathcal{E}_{ADM}(\tau,\sigma^r)$ does not agree with the pseudo-energy density defined in terms of the Landau–Lifschitz pseudo-tensor.

In order to get a deterministic evolution for the DOs[20] we must fix the gauge completely, that is we must add 14 gauge-fixing constraints satisfying an orbit con-

[18] See Ref. [4] for the problem of time in general relativity.

[19] Kuchar [43] says that the super-Hamiltonian constraint must not be interpreted as a generator of gauge transformations, but as an effective Hamiltonian.

[20] See Ref. [45] for the modern formulation of the Cauchy problem for Einstein equations, which mimics the steps of the Hamiltonian formalism.

dition and to pass to Dirac brackets. As already said, the correct way to do so is the following one:

(i) Add a gauge-fixing constraint to the secondary super-Hamiltonian constraint.[21] This gauge-fixing fixes the form of Σ_τ, i.e. the convention for the synchronization of clocks. The τ-constancy of this gauge-fixing constraint generates a gauge-fixing constraint to the primary constraint $\pi_N(\tau, \sigma^r) \approx 0$ for the determination of the lapse function.

(ii) Add three gauge-fixings to the secondary super-momentum constraints $\mathcal{H}_{(a)}(\tau, \sigma^r) \approx 0$. This fixes the 3-coordinates on each Σ_τ. The τ-constancy of these gauge fixings generates the three gauge fixings to the primary constraints $\pi_{\vec{N}\,(a)}(\tau, \sigma^r) \approx 0$ and leads to the determination of the shift functions (i.e. of the appearances of *gravito-magnetism*).

(iii) Add six gauge-fixing constraints to the primary constraints $\pi_{\vec{\varphi}\,(a)}(\tau, \sigma^r) \approx 0$, $M_{(a)}(\tau, \sigma^r) \approx 0$. This is a fixation of the cotetrad field which includes a convention on the choice and the transport of the three gyroscopes of every timelike observer of the two congruences associated with the chosen 3+1 splitting of M^4 (see Refs. [17, 18]).

(iv) In the rest-frame instant form we must also add three gauge fixings to the rest-frame conditions $P^r_{ADM} \approx 0$. The natural ones are obtained with the requirement that the three ADM boosts vanish. In this way we select a special timelike observer as origin of the 3-coordinates (like the Fokker–Pryce center of inertia in special relativity Refs. [14, 23]).

In this way all the gauge variables are fixed to be either numerical functions or well determined functions of the DOs. This complete gauge fixing is physically equivalent to a definition of the *global non-inertial frame centered on a timelike observer, carrying its pattern of inertial forces*, we have called NIF (see Ref. [5]). Note that in a NIF, the ADM energy density $\mathcal{E}_{ADM}(\tau, \sigma^r)$ becomes *a well defined function of the DOs only* and the Hamilton equations for them with E_{ADM} as Hamiltonian are a hyperbolic system of partial differential equations for their determination. For each choice of Cauchy data for the DOs on a Σ_τ, we obtain a solution of Einstein's equations (an Einstein universe) in the radar 4-coordinate system associated with the chosen 3+1 splitting of M^4.

Actually, the Cauchy data are the 3-geometry and matter variables on the Cauchy 3-surfaces of a kinematically possible NIF. Such data are restricted by the super-Hamiltonian and super-momentum constraints (which are four of Einstein's equations).

An Einstein space-time M^4 (a *4-geometry*) *is the equivalence class of all the completely fixed gauges (NIF) with gauge equivalent Cauchy data for the DOs on the associated Cauchy and simultaneity surfaces* Σ_τ.

Once a solution of the hyperbolic Hamilton equations (namely, the Einstein equations after a complete gauge fixing) has been found corresponding to a set of Cauchy data, in each NIF we know the DOs in that gauge (*the tidal effects*) and then the

---

[21] The special choice $\pi_\phi(\tau, \sigma^r) \approx 0$ implies that the DOs $R_{\bar{a}}$, $\Pi_{\bar{a}}$, remain canonical even if we do not know how to solve this constraint.

explicit form of the gauge variables (*the inertial effects*). Moreover, *the extrinsic curvature of the simultaneity surfaces Σ_τ is determined too*. Since the simultaneity surfaces are asymptotically flat, it is possible to determine their embeddings $z^\mu(\tau,\sigma^r)$ in M^4. As a consequence, *unlike special relativity*, the conventions for clock synchronization and the whole chrono-geometrical structure of M^4 (gravito-magnetism, 3-geodesic spatial distance on Σ_τ, trajectories of light rays in each point of M^4, one-way velocity of light) are *dynamically determined* .

Let us remark that, if we look at Minkowski space-time as a special solution of Einstein's equations with $R_{\bar{a}}(\tau,\sigma^r) = \Pi_{\bar{a}}(\tau,\sigma^r) = 0$ (zero Riemann tensor, no tidal effects, only inertial effects), we find [29] that the dynamically admissible 3+1 splittings (non-inertial frames) must have the simultaneity surfaces Σ_τ *3-conformally flat*, because the conditions $R_{\bar{a}}(\tau,\sigma^r) = \Pi_{\bar{a}}(\tau,\sigma^r) = 0$ imply the vanishing of the Cotton–York tensor of Σ_τ. Instead, in special relativity, considered as an *autonomous theory*, all the non-inertial frames compatible with the Møller conditions are admissible, so that there is much more freedom in the conventions for clock synchronization.

A first application of this formalism [46] has been the determination of post-Minkowskian background-independent gravitational waves in a completely fixed non-harmonic 3-orthogonal gauge with diagonal 3-metric. It can be shown that the requirements $R_{\bar{a}}(\tau,\sigma^r) << 1$, $\Pi_{\bar{a}}(\tau,\sigma^r) << 1$ lead to a weak field approximation based on a Hamiltonian linearization scheme:

(i) Linearize the Lichnerowicz equation, determine the conformal factor of the 3-metric, and then the lapse and shift functions
(ii) Find E_{ADM} in this gauge and disregard all the terms more than quadratic in the DOs
(iii) Solve the Hamilton equations for the DOs

In this way we get a solution of linearized Einstein's equations, in which the configurational DOs $R_{\bar{a}}(\tau,\sigma^r)$ play the role of the *two polarizations* of the gravitational wave and we can evaluate the embedding $z^\mu(\tau,\sigma^r)$ of the simultaneity surfaces of this gauge explicitly.

Let us conclude with some remarks about the interpretation of the space-time 4-manifold in general relativity.

In 1914 Einstein, during his researches for developing general relativity, faced the problem arising from the fact that the requirement of general covariance would involve a threat to the physical objectivity of the points of space-time M^4, which in classical field theories are usually assumed to have a *well-defined individuality*. This led him to formulate the *Hole Argument*. Assume that M^4 contains a *hole* $\mathcal{H}$, that is an open region where all the non-gravitational fields vanish. It is implicitly assumed that the Cauchy surface for Einstein's equations lies outside $\mathcal{H}$. Let us consider an active diffeomorphism A which remaps the points inside $\mathcal{H}$, but is the identity outside $\mathcal{H}$. For any point $p \in \mathcal{H}$ we have $p \mapsto D_A\, p \in \mathcal{H}$. The induced active diffeomorphism on the 4-metric tensor ${}^4g$, solution of Einstein's equations, will map it into another solution $D^*_A\, {}^4g$ (D^*_A is a dynamical symmetry of Einstein's equations) defined by $D^*_A\, {}^4g(D_A\, p) = {}^4g(p) \neq D^*_A\, {}^4g(p)$. Consequently, we get two solutions of

Einstein's equations with the same Cauchy data outside $\mathcal{H}$ and it is not clear how to save the identification of the mathematical points of M^4.

Einstein avoided the problem by means of the pragmatic *point-coincidence argument*: the only real-world occurrences are the (coordinate-independent) space-time coincidences (like the intersection of two world lines). However, the problem was rebirth by Stachel [47] and then by Earman and Norton [48], and this opened a rich philosophical debate that is still alive today.

We must face the following dilemma:

If we insist on the reality of space-time mathematical points independently of the presence of any physical field (the *substantivalist* point of view of philosophers), we are in trouble with predictability.

If we say that ${}^4g$ and $D_A^*\,{}^4g$ describe the same universe (the so-called *Leibniz equivalence*), we lose any physical objectivity of the space-time points (the *anti-substantivalist* point of view).

Stachel [47] suggested that a physical individuation of the point-events of M^4 could be made only by using *four individuating fields depending on the 4-metric on* M^4, namely that a tensor field on M^4 is needed to identify the points of M^4.

On the other hand, *coordinatization* is the only way to individuate the points *mathematically* since, as stressed by Hermann Weyl [49]: "There is no distinguishing objective property by which one could tell apart one point from all others in a homogeneous space: at this level, fixation of a point is possible only by a *demonstrative act* as indicated by terms like *this* and *there*."

To clarify the situation let us remember that Bergmann and Komar [50] gave *a passive reinterpretation of active diffeomorphisms as metric-dependent coordinate transformations* $x^\mu \mapsto y^\mu(x, {}^4g(x))$ restricted to the solutions of Einstein's equations (i.e. *on-shell*). It can be shown that *on-shell* ordinary passive diffeomorphisms and the *on-shell* Legendre pull-back of Hamiltonian gauge transformations are two (overlapping) dense subsets of this set of *on-shell* metric-dependent coordinate transformations. Since the Cauchy surface for the Hole Argument lies outside the hole (where the active diffeomorphism is the identity), it follows that the passive reinterpretation of the active diffeomorphism D_A^* must be an *on-shell Hamiltonian gauge transformation*, so that the *Leibniz equivalence* reduces to gauge equivalence in the sense of Dirac constraint theory (${}^4g$ and $D_A^*\,{}^4g$ belong to the same gauge orbit). In our language, *Leibniz equivalence* is then reduced to a change of NIF for the same Einstein *universe*.

What remains to be done is to implement Stachel's suggestion according to which the *intrinsic pseudo-coordinates* of Bergmann and Komar [51] should be used as *individuating fields*. These pseudo-coordinates for M^4 (at least when there are no Killing vectors) are four scalar functions $F^A[w_\lambda]$, $A, \lambda = 1,..,4$, of the four eigenvalues $w_\lambda({}^4g, \partial\,{}^4g)$ of the spatial part of the Weyl tensor. Since these eigenvalues can be shown to be in general functions of the 3-metric, of its conjugate canonical momentum (namely, of the extrinsic curvature of Σ_τ) and of the lapse and shift functions, the pseudo-coordinates are well defined in phase space and can be used as a label for the points of M^4.

The final step [5] is to implement the individuation of point-events by considering an arbitrary kinematically admissible 3+1 splitting of M^4 with a given time-like observer and the associated radar 4-coordinates σ^A (a NIF), and imposing the following gauge fixings to the secondary super-Hamiltonian and super-momentum constraints (the only restriction on the functions F^A is the orbit condition):

$$\chi^A(\tau,\sigma^r) = \sigma^A - F^A[w_\lambda] \approx 0. \tag{13.8}$$

In this way we break general covariance completely and we determine the gauge variables θ^r and π_ϕ. Then the τ-constancy of these gauge fixings will produce the gauge fixings determining the lapse and shift functions. After having fixed the Lorentz gauge freedom of the cotetrads, we arrive at a completely fixed gauge in which, after the transition to Dirac brackets, we get $\sigma^A \equiv \tilde{F}^A[r_{\bar{a}}(\sigma),\pi_{\bar{a}}(\sigma)]$, namely the conclusion that the *radar* 4-coordinates of a point in M^4_{3+1}, the copy of M^4 coordinatized with the chosen non-inertial frame, are determined *off-shell* by the four DOs of that gauge: in other words *the* individuating fields *are nothing else than the genuine* tidal *effects of the gravitational field*. By varying the functions F^A we can make an analogous off-shell identification in every other admissible non-inertial frame. The procedure is consistent, because the DOs are functionals of the metric and the extrinsic curvature on a whole 3-space Σ_τ but in fact know the whole 3+1 splitting M^4_{3+1} of M^4.

Some consequences of this identification of the point-events of M^4 are:

(i) *The physical space-time M^4 and the vacuum gravitational field are essentially the same entity*. The presence of matter modifies the solutions of Einstein equations, i.e. M^4, but plays only an indirect role in this identification (see Ref. [5]). On the other hand, matter is fundamental in establishing a (still lacking) dynamical theory of measurement exploiting nontest objects. Consequently, instead of the dichotomy substantivalism/relationism, it seems that this analysis – as a case study limited to the class of space-times dealt with – may offer a new more articulated point of view, which can be named *point structuralism* (see Ref. [52]).

(ii) The *reduced phase space* of this model of general relativity is the space of *abstract* DOs (pure *tidal* effects without *inertial* effects), which can be thought of as four fields residing on an abstract space-time $\tilde{M}^4$ defined as the equivalence class of all the admissible, non-inertial frames M^4_{3+1} containing the associated inertial effects.

(iii) Each *radar* 4-coordinate system of an admissible non-inertial frame M^4_{3+1} has an associated *noncommutative structure*, determined by the Dirac brackets of the functions $\tilde{F}^A[r_{\bar{a}}(\sigma),\pi_{\bar{a}}(\sigma)]$ determining the gauge, a fact that could play a role in the quantization of the theory.

As a final remark, let us note that these results on the identification of point-events are *model dependent*. In spatially compact space-times without boundary, the DOs are *constants of the motion* due to the frozen picture. As a consequence, the gauge fixings $\chi^A(\tau,\sigma^r) \approx 0$ (in particular χ^τ) cannot be used to rebuild the temporal dimension: probably only the instantaneous 3-space of a 3+1 splitting can be individuated in this way.

13.5 Conclusions

Our everyday experience of macroscopic objects and processes is scientifically described in terms of Newtonian physics with its separate notions of time (and simultaneity) and (Euclidean instantaneous) 3-space. A huge amount of philosophical literature has been devoted to the analysis of the consequences that follow from the empirical fact that light velocity, as well as that of any causal propagation, has finite magnitude. Our macroscopic experience is dominated by Maxwell equations even for the fact that, from the physical and neurophysiological point of view, all the information that reaches our brain is of electromagnetic origin. Therefore, the consequences of the finite magnitude of the causal propagation of energy and information has a direct bearing on our phenomenological experience. On the other hand, the conventional nature of the definition of distant simultaneity that follows from the analysis of the basic structure of causal influences in SR seems to conflict with every possible notion of *3-dimensional reality* of objects and processes which stands at the basis of our phenomenological experience since it entails that no observer- and frame-independent notions of simultaneity and instantaneous 3-space be possible. Even if – from the technical point of view – the question of the conventionality of simultaneity can be rephrased as a gauge problem, it lasted as source of an unending debate involving old fundamental issues concerning the philosophy of time, like that of the nature of now-ness, becoming, reality or unreality of time, past and future, with all possible ramifications and varieties of philosophical distinctions.

It should not be undervalued that relativistic thinking unifies the *physical* notions of space and time in a 4D structure, whilst space and time maintain a substantial ontological diversity in our phenomenological experience. While time is experienced as "flowing," space is not. Furthermore time, even more than space, plays a fundamental constitutive role for our "being in the world" and for *subjectivity* in general, which manifests itself in living beings with various gradations. There is, therefore, a deep contrast between the formal intersubjective unification of space and time in the scientific relativistic image, on the one hand, and the ontological diversity of time and space within the subjectivity of experience, on the other. This appears to be the most important and difficult question that physics raises to contemporary philosophy, since it reveals the core of the relation between *reality* of *experience* and *reality–objectivity* of *knowledge*. Dismissing this contrast by a literal adoption of the scientific image is not as much a painless and obvious operation as rather an implicit adoption of a strong physicalist philosophical position that should be argued for itself.

This said, we have faced the question to investigate a possible contribution of the inclusion of gravity (which, as well-known, is a universal interaction that cannot be shielded) to the clarification of the problem of relativistic distant simultaneity. This has been done having in view certainly not a *technical resolution* of the above philosophical contrast, rather as the achievement of a notion of distant simultaneity within the scientific image which be at least *compatible* with our deep experience of what Whitehead called the “cosmic unison”.

As a matter of fact, *we have shown that the inclusion of gravity deeply changes the state of affairs about relativistic simultaneity*.

In brief, we have identified a class of curved pseudo-Riemannian space-times in which the following results holds:

(i) Outside the solutions of Einstein’s equations (i.e. *off-shell*), these space-times admit 3+1 splittings, which can be interpreted as *kinematically possible global non-inertial laboratories* (kinematically possible NIFs) centered on *arbitrary accelerated observers*. The viewpoint following from this concept leads to a frame-dependent notion of *instantaneous 3-space,* which is concomitantly a *clock synchronization convention*. As in SR, all these conventions are gauge equivalent, so that there is *no Wheeler–DeWitt interpretation* of the gauge transformations generated by the super-Hamiltonian constraint.

(ii) The *off-shell* Hamiltonian separation of the *tidal* degrees of freedom of the gravitational field from the *gauge variables* implies the interpretation of the latter as *relativistic inertial effects* which are shown in the chosen kinematical NIF. Since in this class of space-times the Hamiltonian is the weak ADM energy plus a combination of the first-class constraints, in every completely fixed gauge (a well-defined kinematical NIF) it follows deterministic evolution of the *tidal degrees of freedom* in mathematical time (to be replaced by a physical clock when eventually possible) governed by the *tidal forces* and the *inertial forces* of that NIF (note that unlike the Newtonian physics, such forces are in general functions of the tidal degrees of freedom too).[22]

(iii) The solution of Hamilton equations in a completely fixed gauge with given Cauchy data for the *tidal* degrees of freedom (and matter if present) determines a solution of Einstein’s equations in a well-defined system of 4-coordinates, which *on-shell* are reinterpretable as coordinates adapted to a *dynamically determined* NIF (one of its leaves is the Cauchy surface on which the Cauchy data have been assigned).

(iv) Given any solution of Einstein’s equations in a given 4-coordinate system, we can determine the dynamical 3+1 splitting (a *dynamical* NIF) of Einstein’s space-time, one of whose simultaneity 3-surfaces is just the Cauchy surface of the solution. Consequently, there is a *dynamical emergence of the instantaneous 3-spaces*, leaves of the dynamical NIF, for each solution of Einstein’s equations in a given 4-coordinate system (adapted *on-shell* to the dynamical NIF). Moreover, all the chrono-geometrical structure of Einstein’s space-time ($ds^2 = {}^4g_{\mu\nu}(x)\, dx^\mu\, dx^\nu$) is *dynamically determined.*

(v) These results and a revisitation of the Hole Argument imply that *space-time and vacuum gravitational field are two faces of the same reality*, and we get

[22] In these globally hyperbolic space-times there is no *frozen picture of dynamics*.

a *new kind of structuralism* (with elements of both the *substantivalist* and *relationist* points of view) implying a 4D holism (see Ref. [5]) resulting from a foliation with 3D instantaneous 3-spaces.[23]

In conclusion what in Newton's theory was an absolute Euclidean instantaneous 3-space reappears in GR as a *dynamicadlly emergent* Riemannian time-varying instantaneous 3-space, which is a simultaneity leaf of a *dynamical* NIF *uniquely associated to a solution of Einstein's equations in 4-coordinates adapted to the NIF itself.* The NIF is centered on a timelike (in general accelerated) observer, whose world line can be made to coincide with the Fokker–Pryce center of inertia by means of a suitable gauge fixing to the rest-frame conditions. In the post-Newtonian approximation around the Earth we describe the situation in a *quasi-inertial frame* with harmonic 4-coordinates, as those considered in the IAU conventions for the geocentric celestial reference frame [54].

Admittedly, all the physical implications of this viewpoint must still be worked out (for instance the determination of non-inertial frames in which the Riemannian distance from the Earth to a galaxy equals the galaxy luminosity distance and the implications for dark matter and dark energy of the dynamical instantaneous 3-spaces).

Let us remark that in SR (and in GR too before identifying the preferred dynamical convention of clock synchronization), an ideal observer has the following freedom in the description of the phenomena around him:

(i) The arbitrary choice of the clock synchronization convention, i.e. of the instantaneous 3-spaces
(ii) The choice of the 4-coordinate system.

After these choices, the observer has a description of the other world lines and/or worldtubes simulating the phenomena with Hamiltonian evolution in the chosen time parameter. All these descriptions have been shown to be gauge-equivalent in the previous sections. Every other ideal observer has the same type of freedom in the description of the phenomena.

Each solution of Einstein's equations, i.e. each Einstein *universe*, in our class of models, is an equivalence class of well-defined *dynamical* NIFs (the *epistemic part* of the metric field describing the *generalized relativistic inertial effects*) with their *dynamical clock synchronization conventions*, their *dynamical instantaneous 3-spaces* and their *dynamical individuation of point-events*.[24] The NIFs selected by one solution are *different* from the NIFs selected by a *different solution*. Let us stress, however, that given a solution, the set of the associated NIFs is a *substantially smaller* set than that of the a-priori *kinematically* possible NIFs both in GR and in SR, since the only restrictions at the kinematical level are given by the Møller

[23] In spatially compact space-times without boundary, where there is a frozen picture of dynamics and only a local time-evolution according to the Wheeler–DeWitt interpretation, only 3-space, but not the time direction, can be determined from the gravitational field.

[24] Maybe even realized by means of 4-coordinates not adapted to the NIF; we simplified the exposition by formulating the NIFs with adapted *radar* 4-coordinates

conditions.[25] Given an Einstein *universe*, all the associated NIFs in the equivalence class are connected by *on-shell Hamiltonian gauge transformations* (containing adapted passive diffeomorphisms) so that they know – as it were – the Cauchy data of the solution. Note moreover that they also contain the freedom of changing the timelike observer's origin of the 3-coordinates on the instantaneous 3-spaces, and the freedom of making an arbitrary (tensorial) passive diffeomorphism leading to nonadapted 4-coordinates.

From the point of view of the 4D picture with the freedom of passive diffeomorphisms one could be led to adopt the misleading notion of "block universe" even in GR. However, one should not forget that this "block universe" is the equivalence class of *dynamical* NIFs (stratification or 3+1 splittings with dynamical generated instantaneous 3-spaces). Once the epistemic framework of a NIF is chosen, a *well-defined B-series is established*, since the notion of "earlier than," "later than," and "simultaneous with" is globally defined in terms of the mathematical time of the globally hyperbolic space-time (to be then replaced with a physical clock monotonically increasing in the mathematical time, at least for a finite interval). Unlike in SR, where each observer has their own B-series, we have disclosed the relevant fact that, at least for a specific class of models, GR is characterized by a *universal B-series*. As a matter of fact, a Hamiltonian gauge transformation changes the NIF, the Hamiltonian, the matter distribution, the Cauchy surface and the form in which the Cauchy data are given on it, in such a way that the B-series relations between *any pair of events* is left invariant. The same happens with the freedom of changing the time-like observer that defines the origin of the coordinates on the instantaneous 3-spaces.

Nothing are we willing to say about the relevance of A-series determinations within the scientific image, as already stressed in the Introduction. We maintain that any tensed determination is wholly foreign to any kind of physical description of the world. Clearly, unlike the case of SR, we here have Hamiltonian evolution of the 3-space too, determined by the ADM energy (which depends on the *ontic* tidal effects and on the matter). This evolution, however – as to its temporal characterization – is not substantially different from the Newtonian evolution of matter in absolute time.

It remains to be clarified – as anticipated in the Introduction – the import (if any) of our technical results on the dichotomy *endurantism/perdurantism*. Having already stated our definition of *wholly presentness* of a *physical object* in the Introduction, we must now only ascertain which of the relevant physical features of a general-relativistic space-time with matter we have enumerated so far, tend to support a reasonable notion of *endurantism* or *perdurantism* at the level of the scientific image.

It seems clear that all the attributes which are necessary to define a spatially-extended *physical object* belonging to a dynamically generated 3-space Σ_τ at a certain time τ, can be obtained by the chrono-geometrical features intrinsic to Σ_τ and matter distribution on it. Note – remarkably – that among such attributes there is

[25] If Minkowski space-time without matter is considered as a special solution of Einstein's equations, its dynamical NIFs have the simultaneity leaves 3-conformally flat [29, 35].

the 3-geometry of Σ_τ. We can conclude, therefore, that, in this limited sense, our results support an *endurantist* view of physical objects. It is interesting to note, on the other hand, that our analysis does not support a likewise simple *endurantist* view of the space-time structure itself. We have established that the *reality* of the vacuum space-time of GR is ontologically equivalent to the *reality* of the autonomous degrees of freedom of the gravitational field as described by the DOs (viz., the *ontic part* of the metric field). At this point we should look at space-time itself as at something sharing the attributes of a peculiar *physical object*. We could ask accordingly whether and to what extent the mathematical structure of the DOss allows an *endurantist* interpretation of such a peculiar object. The answer is simple: as already said the DOs, though being local fields indexed by the *radar* coordinates σ^A, when considered in relation to the 4-metric field g, are highly nonlocal functionals of the 3-metric field and of the extrinsic curvature 3-tensor on the whole 3-space Σ_τ[26]. Due to the extrinsic curvature, the structure of the DOs involves therefore an infinitesimal τ-continuum of 3-spaces around Σ_τ. The individuation procedure involves moreover a *temporal gauge* (A $=0$ in Eq. (13.8)). In conclusion, the *physical individuation* of the space-time *point-events*, defined by Eq. (13.8), *cannot* be considered as an attribute depending upon information wholly contained in the 3-space considered at time τ. This conclusion, however, should not be viewed as an unexpected and unsatisfactory result, given the double role of the metric field in GR.

Let us close by stressing a general fundamental feature of GR. Though Einstein's partial differential equations are defined in a 4D framework, this framework must be considered as an unfolding of 3D substructures because of the nature of the Cauchy problem. Consequently, the *models* of GR are subdivided into two disjoint classes:

(i) The 4D ones (with spatially compact space-times), having the *problem of time*, the *frozen picture*, and a lacking *physical individuation* of point-events
(ii) The asymptotically flat space-times, with their 3+1 splitting and *dynamical emergence* of *achronal* 3-spaces, a *nontrivial temporal evolution* and a *physical individuation* of point-events.

References

1. G. Petit and P. Wolf, *Relativistic Theory for Time Comparisons: a Review*, METROLOGIA, **42**, S138–S144, (2005).
2. A. Einstein, *Die formale Grundlage der allgemeinen Relativitätstheorie*, in *Preuss. Akad. der Wiss. Sitz.*, pp. 1030–1085: (1914).
3. A. Einstein, *Die Grundlage der allgemeinen Relativitätstheorie*, *Annalen der Physik* **49**, 769–822, (1916); translation by W. Perrett and G. B. Jeffrey, *The foundation of the general theory of relativity*, in *The Principle of Relativity*, pp. 117–118 Dover, New York (1952).
4. C.J. Isham, *Canonical Quantum Gravity and the Problem of Time*, in *Integrable Systems, Quantum Groups and Quantum Field Theories*, (eds.) L.A. Ibort and M.A. Rodriguez (Salamanca, Kluwer, London, 1993).

[26] Let us recall that this 3-tensor describes the *embedding* of the instantaneous 3-space in the Einstein's 4-manifold through its dependence upon the *τ-derivative* of the 3-metric and the *lapse* and *shift* functions.

K. Kuchar, *Time and Interpretations of Quantum Gravity*, in Proc. 4th Canadian Conf. on *General Relativity and Relativistic Astrophysics*, (eds.) G. Kunstatter, D. Vincent, and J. Williams (World Scientific, Singapore, 1992).

5. L. Lusanna and M. Pauri, *Explaining Leibniz Equivalence as Difference of Non-Inertial Appearances: Dis-solution of the Hole Argument and Physical Individuation of Point-Events*, Talk at the Oxford Conference on Spacetime Theory (2004), Studies in History and Philosophy of Modern Physics, **37(4)**, pp. 692–725 (2006) (gr-qc/0604087); *The Physical Role of Gravitational and Gauge Degrees of Freedom in General Relativity. I: Dynamical Synchronization and Generalized Inertial Effects; II: Dirac versus Bergmann Observables and the Objectivity of Space-Time*, Gen. Rel. Grav. **38**, 187 and 229 (2006) (gr-qc/0403081 and 0407007).
6. V. Perlick, *Characterization of Standard Clocks in General Relativity*, in *Semantical Aspects of Spacetime Theories*, U. Mayer and H.–J. Schmidt (eds.) B.I. Wissenshaftverlag, pp. 169–179 (Mamheim, 1994).
7. *Being and Becoming in Modern Physics*, Stanford Encyclopedia of Philosophy, (http://plato.stanford.edu/entries/spacetime-bebecome/); *Philosophy of Space and Time*, Wikipedia, the free encyclopedia (http://en.wikipedia.org/wiki/$Philosophy_o f_s pace_a nd_t ime$).
8. A. Grünbaum, *The Meaning of Time*, in *Essays in Honor of Carl G. Hempel*, A Tribute in Occasion of his Sixty-Fifth Birthday. pp. 147–177 (D. Reidel, Dordrecht, 1972).
9. *Temporal Parts*, Stanford Encyclopedia of Philosophy: (http://plato.stanford.edu/entries/temporal-parts/).
10. D. Lynden-Bell, *A Relative Newtonian Mechanics*, *Mach's principle. From Newton's bucket to quantum gravity*, (eds.) J. Barbour and H. Pfister pp. 172–178 (Birkhauser, Boston, 1995).
11. D. Alba, L. Lusanna, and M. Pauri, *Dynamical Body Frames, Orientation-Shape Variables and Canonical Spin Bases for the Nonrelativistic N-Body Problem*, J. Math. Phys. **43**, 373 (2002) (hep-th/0011014).
12. R.G. Littlejohn and M. Reinsch, *Gauge Fields in the Separation of Rotations and Internal Motions in the n-Body Problem*, Rev. Mod. Phys. **69**, 213 (1997).
13. D. Alba, L. Lusanna, and M. Pauri, *Centers of Mass and Rotational Kinematics for the Relativistic N-Body Problem in the Rest-Frame Instant Form* J. Math. Phys. **43**, 1677–1727 (2002) (hep-th/0102087).
14. D. Alba, L. Lusanna, and M. Pauri, *New Directions in Non-Relativistic and Relativistic Rotational and Multipole Kinematics for N-Body and Continuous Systems* (2005), in *Atomic and Molecular Clusters: New Research* (ed.) Y.L. Ping (Nova Science, New York, 2006) (hep-th/0505005).
15. D. Alba and L. Lusanna, *Simultaneity, Radar 4-Coordinates and the 3+1 Point of View About Accelerated Observers in Special Relativity* (2003) (gr-qc/0311058); *Generalized Radar 4-Coordinates and Equal-Time Cauchy Surfaces for Arbitrary Accelerated Observers* (2005), to appear in Int. J. Mod. Phys. D (gr-qc/0501090).
16. L. Lusanna, *The Chrono-geometrical Structure of Special and General Relativity: a Re-Visitation of Canonical Geometrodynamics*, Lectures given at the 42nd Karpacz Winter School of Theoretical Physics, Current Mathematical Topics in Gravitation and Cosmology (Ladek, Poland, 6–11 February 2006) Int. J. Geom. Methods in Mod. Phys., **4**, 79 (2007) (gr-qc/0604120).
17. L. Lusanna, *The Chronogeometrical Structure of Special and General Relativity: Towards a Background-Independent Description of the Gravitational Field and Elementary Particles* (2004), in *General Relativity Research Trends*, Horizon in World Physics vol. 249 (ed.) A. Reiner (Nova Science, New York, 2005) (gr-qc/0404122).
    L. Lusanna, *Towards a Unified Description of the Four Interactions in Terms of Dirac-Bergmann Observables*, in *Quantum Field Theory: A 20th Century Profile of the Indian National Science Academy*, (ed.) A.N. Mitra, foreward F.J. Dyson (Hindustan Book Agency, New Delhi, 2000) (hep-th/9907081).
18. B. Mashhoon, *The Hypothesis of Locality and its Limitations*, (gr-qc/0303029); *Limitations of Spacetime Measurements*, Phys. Lett. **A143**, 176 (1990). *The Hypothesis of Locality in Relativistic Physics*, Phys. Lett. **A145**, 147 (1990). *Measurement Theory and*

General Relativity, in *Black Holes: Theory and Observation*, Lecture Notes in Physics 514 (ed.) F. W. Hehl, C. Kiefer, p. 269 and R.J.K. Metzler (Springer, Heidelberg, 1998), *Acceleration-Induced Nonlocality*, in *Advances in General Relativity and Cosmology*, (ed.) G. Ferrarese (Pitagora, Bologna, 2003) (gr-qc/0301065).
B. Mashhoon and U. Muench, *Length Measurement in Accelerated Systems*, Ann. Phys. (Leipzig) **11**, 532 (2002).
19. M. Pauri and M. Vallisneri, *Maerzke-Wheeler Coordinates for Accelerated Observers in Special Relativity*, Found. Phys. Lett. **13**, 401 (2000).
20. L. Lusanna, *The N- and 1-Time Classical Descriptions of N-Body Relativistic Kinematics and the Electro-Magnetic Interaction*, Int. J. Mod. Phys. **A12**, 645 (1997).
21. H. Crater and L. Lusanna, *The Rest-Frame Darwin Potential from the Lienard-Wiechert Solution in the Radiation Gauge*, Ann. Phys. (N.Y.) **289**, 87–177 (2001) (hep-th/0001046).
22. D. Alba, H. Crater, and L. Lusanna, *The Semiclassical Relativistic Darwin Potential for Spinning Particles in the Rest-Frame Instant Form: Two-Body Bound States with Spin 1/2 Constituents*, Int. J. Mod. Phys. **A16**, 3365–3477 (2001) (hep-th/0103109).
23. D. Alba, H. Crater and L. Lusanna, *Hamiltonian Relativistic Two-Body Problem: Center of Mass and Orbit Reconstruction*, J. Phys. **A40**, 9585 (2007) (hep-th/0610200).
24. D. Alba and L. Lusanna, *Quantum Mechanics in Non-Inertial Frames with a Multi-Temporal Quantization Scheme: I) Relativistic Particles*, Int. J. Mod. Phys. **A21**, 2781 (2006) (hep-th/0502060).
D. Alba, *Quantum Mechanics in Non-Inertial Frames with a Multi-Temporal Quantization Scheme: II) Non-Relativistic Particles*, Int. J. Mod. Phys. **A21**, 3917 (2006) (hep-th/0504060).
25. J. Norton, *General Covariance and the Foundations of General Relativity: Eight Decades of Dispute*, Rep. Prog. Phys. **56**, 791–858 (1993).
26. H. Nicolai, K. Peeters, and M. Zamaklar, *Loop Quantum Gravity: an Outside View* (hep-th/0501114).
27. L. Smolin, *How Far are we from Quantum Gravity?* (hep-th/0303185).
28. R. Arnowitt, S. Deser, and C.W. Misner, *The Dynamics of General Relativity*, in *Gravitation: an Introduction to Current Research*, (pp. 227–265) (ed.) (Wiley, New York, 1962).
29. L. Lusanna, *The Rest-Frame Instant Form of Metric Gravity*, Gen. Rel. Grav. **33**, 1579–1696 (2001) (gr-qc/0101048).
30. R.M. Wald, *General Relativity* (University of Chicago Press, Chicago, 1984).
31. T. Regge and C. Teitelboim, *Role of Surface Integrals in the Hamiltonian Formulation of General Relativity*, Ann. Phys. (N.Y.) **88**, 286–318 (1974).
R. Beig and N.O'Murchadha, *The Poincaré Group as the Simmetry Group of Canonical General Relativity*, Ann. Phys. (N.Y.) **174**, 463–498 (1987).
32. A. Lichnerowicz, *Propagateurs, Commutateurs et Anticommutateurs en Relativite Generale*, in Les Houches 1963, *Relativity, Groups and Topology* (eds.) C. DeWitt and B. DeWitt (Gordon & Breach, New York, 1964).
C. Moreno, *On the Spaces of Positive and Negative Frequency Solutions of the Klein-Gordon Equation in Curved Space-Times*, Rep. Math. Phys. **17**, 333 (1980).
33. L. Lusanna, *Classical Yang-Mills Theory with Fermions. I. General Properties of a System with Constraints*, Int. J. Mod. Phys. **A10**, 3531–3579 (1995); *II. Dirac's Observables*, Int. J. Mod. Phys. **A10**, 3675–3757 (1995).
34. D. Christodoulou and S. Klainerman, *The Global Nonlinear Stability of the Minkowski Space*. (Princeton University Press, Princeton, 1993).
35. L. Lusanna and S. Russo, *A New Parametrization for Tetrad Gravity*, Gen. Rel. Grav. **34**, 189–242 (2002) (gr-qc/0102074).
R. De Pietri, L. Lusanna, L. Martucci, and S. Russo, *Dirac's Observables for the Rest-Frame Instant Form of Tetrad Gravity in a Completely Fixed 3-Orthogonal Gauge*, Gen. Rel. Grav. **34**, 877–1033 (2002) (gr-qc/0105084).
36. S. Shanmugadhasan, *Canonical Formalism for Degenerate Lagrangians,* J. Math. Phys. **14**, 677 (1973); L. Lusanna, *Canonical Transformations, Function Groups and the Extended Second Noether Theorem*, Int. J. Mod. Phys. **A8**, 4193 (1993).

37. D. Alba and L. Lusanna, *The York Map as a Shanmugadhasan Canonical Transformation in Tetrad Gravity and the Role of Non-Inertial Frames in the Geometrical View of the Gravitational Field* (gr-qc/0604086) submitted to Gen. Rel. Grav.
38. J. Isemberg and J.E. Marsden, *The York map is a Canonical Transformation* , J. Geom. Phys. **1**,85 (1984).
39. F.I. Cooperstock and S. Tieu, *General Relativity resolves Galactic Rotation Curves without Exotic Dark Matter* (astro-ph/0507619) and *Perspectives on Galactic Dynamics via General Relativity* Int. J. Mod. Phys. **A13**, 2293 (2007) (astro-ph/0512048).
40. M. Korzynski, *Singular Disk of Matter in the Cooperstock-Tieu Galaxy Model*, (astro-ph/0508377).
    D. Vogt, *Presence of Exotic Matter in the Cooperstock and Tieu Galaxy Model* (astro-ph/0510750).
    D. Garfinkle, *The Need for Dark Matter in Galaxies*, (astro-ph/0511082).
    D.J. Cross, *Comments on the Cooperstock-Tieu Galaxy Model*, (astro-ph/0601191).
41. M. Milgrom, *A Modification of the Newtonian Dynamics as a Possible Alternative to the Hidden Mass Hypothesis*, Astrophys. J. **270**, 365 (1983); *MOND as Modified Inertia* (astro-ph/0510117).
42. C. Teitelboim, *The Hamiltonian Structure of Spacetime*, in *General Relativity and Gravitation*, (ed.) A. Held, vol. I (Plenum, New York, 1980).
43. K. Kuchar, *Canonical Quantum Gravity*, in G. Kunstatter, D.E. Vincent, and J.G. Williams (eds.), *Cordoba 1992, General Relativity and Gravitation* pp.119–150 (gr-qc/9304012) (IOP, Bristol, 1993).
44. B. DeWitt, *Quantum Theory of Gravity, I) The Canonical Theory*, Phys. Rev. **160**, 1113–1148 (1967); *II) The Manifestly Covariant Theory*, **162**, 1195–1239 (1967).
45. A. Rendall, *Local and Global Existence Theorems for the Einstein Equations*, Online. Living Rev. Relativity **1**, n. 4; *ibid.* (2000) **3**, n. 1; (gr-qc/0001008).
    H. Friedrich and A. Rendall, *The Cauchy Problem for Einstein Equations*, in *Einstein's Field Equations and their Physical Interpretation* (ed.) B.G. Schmidt, (Berlin, Springer, 2000) (gr-qc/0002074).
    M. Sa-nchez, *Cauchy Hypersurfaces and Global Lorentzian Geometry*, (math.DG/0604265).
46. J. Agresti, R. De Pietri, L. Lusanna, and L. Martucci, *Hamiltonian Linearization of the Rest-Frame Instant Form of Tetrad Gravity in a Completely Fixed 3-Orthogonal Gauge: a Radiation Gauge for Background-Independent Gravitational Waves in a Post-Minkowskian Einstein Space-Time*, Gen. Rel. Grav. **36**, 1055–1134 (2004) (gr-qc/0302084).
47. J. Stachel, *Einstein's Search for General Covariance, 1912–1915.* Ninth International Conference on General Relativity and Gravitation, Jena (1980), (ed.) E. Schmutzer (Cambridge University Press, Cambridge, 1983).
48. J. Earman and J. Norton, *What Price Spacetime Substantivalism? The Hole Story*, Br. J. Philos. **38**, 515–525 (1987).
49. H. Weyl, *Groups, Klein's Erlangen Program.* Quantities, ch.I, sec.4 of *The Classical Groups, their Invariants and Representations*, 2nd edn. pp. 16 (Princeton University Press, Princeton, 1946).
50. P.G. Bergmann and A. Komar, *The Coordinate Group Symmetries of General Relativity*, Int. Theor. Phys., **5**, 15–28 (1972).
51. P.G. Bergmann and A. Komar, *Poisson Brackets Between Locally Defined Observables in General Relativity*, Phys. Rev. Lett. **4**, 432–433 (1960).
    P.G.Bergmann, *The General Theory of Relativity*, in *Handbuch derPhysik*, Vol. IV, *Principles of Electrodynamics and Relativity*, pp. 247–272 (ed.) S.Flugge, (Springer, Berlin, 1962).
52. M. Dorato and M. Pauri, *Holism and Structuralism in Classical and Quantum General Relativity*, Pittsburgh-Archive, ID code 1606, *The Structural Foundations of Quantum Gravity*, D. Rickles, S. French, and J. Saatsi, Ch.V, pp. 121–151, Clarendon Press, Oxford (2006).
53. P.G. Bergmann, *Observables in General Relativity*, Rev. Mod. Phys. **33**, 510–514 (1961).
54. M. Soffel *et al*, *The IAU 2000 Resolutions for Astrometry, Celestial Mechanics and Metrology in the Relativistic Framework: Explanatory Supplement*, Astron. J. **126**, 2687–2706 (2003); (Astroph/0303376).

Chapter 14
Lorentzian Space-Times from Parabolic and Elliptic Systems of PDEs

Carlos Barceló

Abstract In this essay we try to convey two main ideas. On the one hand, the 4D space-time structure (Lorentzian geometry) with which we are used to describe the world, might be quite a generic emergent feature of sufficiently complex systems (much in the same way as hydrodynamics emerges as the macroscopic behaviour of different many-particle systems). On the other hand, the description of these systems in regimes even just beyond the Lorentzian geometric regime (still macroscopic), could be so different as to require us to change our basic relativistic notions of time, space and causality. To illustrate this point we will take the scalar field equation over a curved space-time – it is a Hyperbolic partial differential equation (PDE) – as the simplest example of a relativistic equation. Then, we will see that this equation can be approximately recovered starting from quite generic systems of PDEs of a completely different character, that is, parabolic and elliptic systems of PDEs. We will discuss different consequences of this view in regard to our usual description of the world as a 4D Lorentzian space-time.

14.1 Introduction

Before scientists were settled down with the theory of relativity the world was typically considered to be a 3D Euclidean space filled with 3D objects of different kinds. In addition, there was a single and absolute time parameter labelling the changes taking place in this world. With the advent of relativity theory this picture of reality underwent a big change. Among other things, within this theory one has to give up the notion of an absolute and simultaneous present. Time is no longer a parameter external to the world, but an inseparable part of it. The classical extended objects populating this, now 4D, world had to be themselves 4D: The invariant (or absolute) notions within relativity can only be defined by making use of relations between space and time. In relativity the three spacial dimensions and time are treated on equal footing and we are used to talk about the world as a 4D space-time. However,

V. Petkov (ed.), *Relativity and the Dimensionality of the World*, 261–269.

let us recall that time and space are not strictly equivalent even in relativity theory. The relativistic description of the world retains a sense of evolution in "time" so that the 4D continuum has always special directions: The directions of causal propagation. The events happening in a region of space-time have influence in the causal future of this region. In this essay I will use the word relativity to mean descriptions of the world based on curved space-time geometries of Lorentzian character (i.e. with special directions).

This change of conceptual paradigm has a mirror in the way one constructs mathematical models of reality. The classical perception of a 3D world evolving in time prior to relativity, was accommodated in the mathematical models of reality in the form of systems of ordinary differential equations, or when considering fields, in the form of systems of partial differential equations (PDEs) with an elliptic or parabolic character (see for example [1]). An example of elliptic PDE is the Poisson equation for a Newtonian distribution of masses:

$$\nabla^2 \phi(\mathbf{x}) = \left(\partial_x^2 + \partial_y^2 + \partial_z^2\right) \phi(\mathbf{x}) = \rho(\mathbf{x}). \tag{14.1}$$

This PDE relates the Newtonian potential ϕ with the matter density ρ. An example of Parabolic PDE is the heat diffusion equation,

$$\left(\partial_t - \kappa \nabla^2\right) T(\mathbf{x},t) = 0, \tag{14.2}$$

where $T(\mathbf{x},t)$ represents the temperature and κ the thermal diffusivity. Poisson equation does not involve any notion of causal evolution in time, while Heat diffusion equation does incorporate causal evolution in time, but at an infinite speed, as opposed to relativity.

Since the advent of relativity, the mathematical models of reality changed towards descriptions based on hyperbolic PDEs. These systems of equations do incorporate the notion of causal relation through the exchange of finite speed signals in specific directions, in accordance with relativity. A consequence of relativity is that every sensible basic theory of reality has to contain wave-like solutions. For example, Einstein equations of general relativity can be understood as the hyperbolic (or wave-like) generalization of Newtonian gravitational theory.

We will expand a bit on the differences between elliptic, parabolic and hyperbolic systems of PDEs at the end of this essay, but now let me add a new ingredient to our discussion.

14.2 Analogue Models of General Relativity

By the time relativity theory came to existence, different systems of hyperbolic PDEs (and their general theory) had already been profusely studied. One of the reasons for that is the knowledge of the existence of wave phenomena (of hyperbolic character), since at least the Greek civilization. Waves in the surface of a liquid, sound waves through the air or solid bodies, and Maxwell description of electromagnetic phenomena are examples of waves in nature.

One of the main outputs of relativity is the necessity of describing every fundamental phenomena in nature as a generalized wave equation. Moreover, the world is modelled to be a curved 4D geometry controlling the behaviour of every such wave. The geometrical properties of the curved 4D world are described by a metric tensor field $g_{\mu\nu}(\mathbf{x},t)$ (where μ,ν are indices that take values t,x,y,z labelling the 4D) of Lorentzian character (that is, in which from the four independent directions at each point there is always one different from the other three; in mathematical terms a Lorentzian metric tensor has signature $-+++$). It is this Lorentzian requirement, imposed to the world geometry, that insures the Hyperbolic character of the basic equations of mathematical physics.

The simplest and more paradigmatic example of relativistic behaviour that one can think of is described by scalar field equation over a curved space-time:

$$\frac{1}{\sqrt{-g}}\partial_\mu\sqrt{-g}g^{\mu\nu}\partial_\nu\phi = 0. \tag{14.3}$$

In this expression we use the sum convention for repeated indexes. In addition, $\phi(\mathbf{x},t)$ represents a scalar field, $g^{\mu\nu}$ is the inverse of the metric tensor and g its determinant. In the case of a Minkowskian (flat) metric, this equation simply reads

$$\left(-\partial_t^2 + c^2\partial_x^2 + c^2\partial_y^2 + c^2\partial_z^2\right)\phi = 0, \tag{14.4}$$

that is, it is exactly the wave equation.

At this point of the discussion, it is not difficult to understand why the same relativistic equations used in the modern fundamental description of the world, can be found in the analysis of wave (or oscillatory) phenomena in many day-to-day systems. This is the main observation underneath an area of research come to known as "Analogue models of general relativity" [2] or "Analogue gravity" [3]. Many relativistic phenomena of difficult observability in nature could be reproduced in analogue system of condensed matter (as their behaviour is formally equivalent), allowing to experiment with them. Here by condensed matter systems I understand systems of many interacting particles that can, however, be described in a macroscopic-average fashion.

From a reversed perspective, it has been seen that in many condensed matter systems, the behaviour of the perturbations of the system with respect to a given background configuration (that can, in many circumstances, be the lowest energy state of the system, i.e. the vacuum state [4]), can be described as if the background configuration had acquired the form of an effective curved Lorentzian geometry (see, for example [3]). It is very suggesting to realize that Lorentzian geometries can easily appear in this process of describing the holistic state of the system as separated into a macroscopic state plus small fluctuations around it.

Imagine that we take a condensed-matter-like system (we can think of it in abstract terms) which is appropriately described in an averaged fashion by a hyperbolic system of PDEs (in general non-linear). One can then linearize around a fixed, but arbitrary, background configuration and see whether the linearized equations have or have not the form of typical relativistic equations. Of course, for a completely general Hyperbolic system of PDEs the answer is no; the existence of limiting speeds for the propagation of signals is a necessary but not a sufficient requirement to recover a relativistic behaviour [5]. The important fact, however, is that there is a very large set of systems from which the answer is yes.[1] This is for example what happens with the equations of hydrodynamics for an ideal fluid. From them, it is not difficult to prove that sound waves in a moving fluid behave as a scalar field over a curved effective space-time [6].

However, what we want to highlight in this essay is that in many circumstances it is possible to obtain an effective Lorentzian description for the perturbations of a system even though one did not start from a hyperbolic system of PDEs, but from a parabolic or an elliptic system of PDEs.

14.3 A Simple Parabolic Model

Some particularly interesting systems in which the previous idea gets clearly implemented are the weakly-interacting Bose–Einstein condensates (BECs) [8–10]. An appropriate description of these quantum gases is provided by the Gross–Pitaevskii equation:

$$\mathrm{i}\hbar \frac{\partial}{\partial t}\psi(t,\mathbf{x}) = \left(-\frac{\hbar^2}{2m}\nabla^2 + V_{\mathrm{ext}}(\mathbf{x}) + g\,|\psi|^2\right)\psi(t,\mathbf{x}). \tag{14.5}$$

Here $\psi(t,\mathbf{x})$ represents the macroscopic wave function of the system, $V_{\mathrm{ext}}(\mathbf{x})$ is the external potential applied to the system and g is a constant controlling the interaction strength between the different particles in the gas (for a review article on BECs see, for example [7]). This equation has a Parabolic character as one can easily see by comparing it with the Heat diffusion equation.

Adopting the Madelung representation for the macroscopic wave function

$$\psi = \sqrt{n}e^{\mathrm{i}\theta/\hbar}e^{-\mathrm{i}\mu t/\hbar} \tag{14.6}$$

(here n is the condensate density, μ the chemical potential and θ a phase factor which is related to the velocity potential), and substituting in Eq. (14.5) we arrive at the two real equations

$$\partial_t n = -\frac{1}{m}\nabla\cdot(n\nabla\theta), \tag{14.7}$$

$$\partial_t \theta = -\frac{1}{2m}(\nabla\theta)^2 - g\,n - V_{\mathrm{ext}} - \mu - V_{\mathrm{quantum}}, \tag{14.8}$$

[1] The restriction to be imposed into a general hyperbolic system of PDEs to give place to a relativistic description are somewhat parallel to the restriction that the equivalence principle imposed at the time of constructing theories of gravity [5].

where the "quantum potential" is defined as

$$V_{\text{quantum}} = -\frac{\hbar^2}{2m}\frac{\nabla^2\sqrt{n}}{\sqrt{n}}. \tag{14.9}$$

By linearizing these equations, that is, by writing

$$n(\mathbf{x},t) = n_0(\mathbf{x}) + g^{-1}\widetilde{n}_1(\mathbf{x},t), \tag{14.10}$$
$$\theta(\mathbf{x},t) = \theta_0(\mathbf{x}) + \theta_1(\mathbf{x},t), \tag{14.11}$$

with $\widetilde{n}_1$ and θ_1 representing small perturbations of the density and phase of the BEC (the extra normalization factor g^{-1} is not important, we just put it because it simplifies future writing), we arrive at the following linear PDEs for the perturbations:

$$\partial_t\widetilde{n}_1 = -\nabla\cdot\left(\widetilde{n}_1\mathbf{v} + c^2\nabla\theta_1\right), \tag{14.12}$$
$$\partial_t\theta_1 = -\mathbf{v}\cdot\nabla\theta_1 - \widetilde{n}_1 + \frac{1}{4}\xi^2\nabla\cdot\left[c^2\nabla\left(\frac{\widetilde{n}_1}{c^2}\right)\right]. \tag{14.13}$$

In the process of arriving to these expressions we have defined $\mathbf{v} \equiv \nabla\theta_0/m$, $c^2 \equiv gn_0/m$ and $\xi \equiv \hbar/mc$. These magnitudes, $\mathbf{v}$, c and ξ, represent respectively, the local velocity of the gas flow, the local velocity of sound in the gas, and the so-called healing length of the BEC.

As we have mentioned, these linear system of PDEs is parabolic (in fact a non-linear system of PDEs is Parabolic if and only if all its linearizations are Parabolic). However, for phenomena occurring in length scales much larger than ξ, the only term containing this parameter – the term coming from the quantum potential – can be neglected, as it becomes much smaller that the rest. Then, mixing the two equations it is not difficult to see that θ_1 satisfies

$$\frac{1}{\sqrt{-g}}\partial_\mu\sqrt{-g}g^{\mu\nu}\partial_\nu\theta_1 = 0, \tag{14.14}$$

with

$$(g_{\mu\nu}) = \frac{m}{g}c\left(\begin{array}{c|c} v^2-c^2 & -\mathbf{v}^{\mathrm{T}} \\ \hline -\mathbf{v} & \mathbf{I} \end{array}\right). \tag{14.15}$$

That is, in this approximation the θ_1 field satisfies precisely the scalar wave equation in a curved effective space-time: This is the paradigm of hyperbolic PDE in relativistic physics. (This propagation features are also shared by the field $\widehat{n}_1$).

Therefore, we see that although within the previous description of BECs information in one region can travel towards other regions, strictly speaking, at arbitrarily high velocities (infinite propagation speed for signal is a characteristic of parabolic PDEs), transmission restricted to low-energy channels (involving modification of the system at length scales larger than the healing length) have effectively a maximum speed for propagation, as in relativity. Then, the velocity of sound in BECs plays the role of the velocity of light in relativity.

It is suggestive to think that Planck's length in general relativity (Planck length is defined as $L_P \equiv \sqrt{G\hbar/c^3} \simeq 10^{-35}$m, where G is Newton constant, $\hbar$ Planck constant and c is the speed of light) could be playing a role similar to the healing length in BECs. In comparison with Planck scale, all experiments performed up to now by humans can be considered "low energy" experiments. Therefore, the possibility is open that when using "high energy" channels for transmission of information, we find transfer speeds higher than c. (However, for example in [11] the reader can find a description of observations putting already strong constraints to high-energy modifications of c.)

14.4 A Simple Elliptic Model

Let us now give a very simple example of elliptic PDE that however can behave approximately as hyperbolic. Take the following single PDE

$$\frac{\xi^2}{c^2}\left(\partial_t^4 + c^4\partial_x^4 + c^4\partial_y^4 + c^4\partial_z^4\right)\phi(\mathbf{x},t)$$
$$+\frac{1}{\sqrt{-g}}\partial_\mu\sqrt{-g}g^{\mu\nu}\partial_\nu\,\phi(\mathbf{x},t) = 0. \tag{14.16}$$

where $g_{\mu\nu}$ corresponds to a well-defined Lorentzian metric tensor. In principle this Elliptic PDE (see next section) could be found to be appropriate for the study of the continuum limit of a 4D net of relations between some set of infinitesimal constituents. Here we consider this PDE as given, without discussing any further its possible deeper origin.

Strictly speaking this PDE does not incorporate any notion of evolution in time. We have used the t label to represents the fourth variable in the PDE but there is not distinction between the coordinates x, y, z and t. However, as happened in the previous parabolic example, when considering solutions with length scales much larger that ξ and time scales much longer than ξ/c, one can approximate the elliptic PDE by a hyperbolic PDE, in this example, one again recovers the wave equation over a curved Lorentzian space-time:

$$\frac{1}{\sqrt{-g}}\partial_\mu\sqrt{-g}g^{\mu\nu}\partial_\nu\,\phi(\mathbf{x},t) = 0. \tag{14.17}$$

We could have used much more complicated examples of PDEs but I think this is enough to illustrate the following point: In a system in which, strictly speaking, there is no transmission of signals (as we typically understand this notion), there are "low energy" phenomena perfectly describable in terms of causal exchange of signals among different regions of "space" (a notion that has effectively shown up thanks to the approximation).

14.5 Characteristic Surfaces

Let me now expand a bit on the differences between elliptic, hyperbolic and parabolic systems of PDEs. Elliptic and hyperbolic (or parabolic) systems of PDEs differ in the existence or not of the so-called characteristic surfaces. Let us take, for simplicity, a system of k linear PDEs of order m in four independent variables denoted generically by $\{\mathfrak{x}\}$

$$A_{ij}^{\mu_1\cdots\mu_m}(\mathfrak{x})\partial_{\mu_1}\cdots\partial_{\mu_m}\phi^j(\mathfrak{x}) + [\text{Lower order derivatives}]_{ij}\ \phi^j(\mathfrak{x}) = 0. \quad (14.18)$$

Given a surface $S(\mathfrak{x}) = 0$ with $\nabla S(\mathfrak{x}) \neq 0$ it is said to be characteristic if

$$\det\left[A_{ij}^{\mu_1\cdots\mu_m}(\mathfrak{x})\partial_{\mu_1}S(\mathfrak{x})\cdots\partial_{\mu_m}S(\mathfrak{x})\right] = 0. \quad (14.19)$$

If a surface is not characteristic it is said to be free. In a free surface by knowing the values of ϕ_j and of their derivatives in the direction orthogonal to the surface up to order $m-1$

$$\left(\delta^{\nu\mu}\partial_\nu S\partial_\mu\right)^p\phi_j \quad p = \{1,2,\cdots,m-1\} \quad (14.20)$$

one can uniquely find the values of $\left(\delta^{\nu\mu}\partial_\nu S\partial_\mu\right)^m\phi_j$ in that surface. In other words, the data set

$$\left\{\phi_j, \left(\delta^{\nu\mu}\partial_\nu S\partial_\mu\right)\phi_j, \left(\delta^{\nu\mu}\partial_\nu S\partial_\mu\right)^2\phi_j, \cdots, \left(\delta^{\nu\mu}\partial_\nu S\partial_\mu\right)^{m-1}\phi_j\right\} \quad (14.21)$$

in a free surface is freely specifiable and from it, one can extrapolate the values of the fields in other surfaces close by. A system of PDEs is called elliptic if any given surface happens to be free. On the other hand, hyperbolic and parabolic systems of PDEs do have characteristic surfaces. In these surfaces the previous data set cannot be freely specified. This implements mathematically the notion of causality extracted from our experience of the world: The description of the future is not independent of the description of its causal past. Characteristic surfaces can be thought of as surfaces through which causality is being transferred. A light wavefront in Minkowski space-time is an example of characteristic surface.

The difference between parabolic and hyperbolic PDEs is technically more subtle. It is said that a system of k PDEs of order m is hyperbolic if there exists a system of coordinates in which (i) One can give arbitrary values to $\partial_{\mathfrak{x}_1}S(\mathfrak{x}), \partial_{\mathfrak{x}_2}S(\mathfrak{x}), \partial_{\mathfrak{x}_3}S(\mathfrak{x})$ and (ii) The characteristic Eq. (14.19) has then $k\times m$ real non-zero solutions for $\partial_{\mathfrak{x}_4}S(\mathfrak{x})$. (If these $k\times m$ real non-zero solutions are all different, the system is called totally hyperbolic). When some of the solutions coming from the characteristic equation are zero, the system is defined as parabolic. In this case one can say that there are directions in space (and in field space) for which the velocity of transmission of signals become infinite.

14.6 Discussion

The way in which one thinks about a system of PDEs that is hyperbolic or elliptic (or of an intermediate parabolic character) is very different. In the former case we have a notion of causality, in the latter we do not. The central idea that this paper is trying to convey is that it is not at all difficult to find relativistic geometrical structures (i.e. Lorentzian space-times) emerging from sufficiently complex systems, and that, independently of the ultimate structural character of such systems. For example, and only as a provocative thought, the world could be really flat and 3-D as in Newtonian physics and still we, as internal beings with only low-energy perceptions, could find it more adequate to describe phenomena in a 4D relativistic manner. In this case, the world and the objects that populate it, as described by external observers to the system or internal observers capable of exploring arbitrarily-high-energy behaviours, would be three dimensional. At each instant of time (an external coordinate to the world), these observers could give, in principle, a precise and absolute description of this 3D world. However, typical internal observers (trying to understand low-energy phenomena) would associate time and space coordinates differently. space-time would be an internal and emergent property of the system. For these internal observers there would be a maximum velocity for the transmission of signals and, therefore, they would tend to develop a relativistic world view. For them, the world would be inseparably 4D though retaining a sense of evolution in time; this time would be now an internal and operational characteristic of the system, not an absolute concept. Our example based on BEC illustrates this possibility.

Another possibility, even more radical, is that the ultimate structural character of the world system were elliptic, with four independent variables, as in the example we described above.[2] We have shown that in this case the situation is quite similar to that in the parabolic case: low-energy internal observers could find it more appropriate to describe the world in a relativistic manner. However, strictly speaking, one could say that causal future and past do not exist except in our partial and approximate description of the world.

To end this essay I would like to remark that although the conceptual framework of theories of emergent relativity tends to rest importance to the details of the underlying structures, I still believe on the enormous strength that fundamental approaches to the description of the world might have. Such approaches might not provide the ultimate answer to our questions, but long lasting and insightful scientific paradigms.

Acknowledgments I would like to thank Luis Garay for very valuable comments.

[2] In this essay we restrict ourselves to systems with a well-defined total dimensionality; for instance, the total number of dimensions is taken to be always four. However, we would like to mention that within emergent theories of relativity it might be sensible to think of the dimensionality of the world as being also an emergent notion.

References

1. R. Courant and D. Hilbert, *Methods of Mathematical Physics,* vol II, (Wiley, 1991).
2. M. Novello, M. Visser and G. Volovik (eds.), *Artificial Black Holes*, (World Scientific, Singapore, 2002).
3. C. Barceló, S. Liberati and M. Visser, Analogue Gravity, Living Rev. Relativity **8**, (2005), 12. URL (cited on 15-05-06): http://www.livingreviews.org/lrr-2005-12 [arXiv:gr-qc/0505065].
4. G. E. Volovik, *The Universe in a Helium Droplet,* (Oxford University Press, 2003).
5. C. Barcelo, S. Liberati and M. Visser, Refringence, field theory, and normal modes, Class. Quant. Grav. **19**, 2961 (2002) [arXiv:gr-qc/0111059].
6. M. Visser, Acoustic black holes: horizons, ergospheres, and Hawking radiation," Class. Quant. Grav. **15**, 1767 (1998) [arXiv:gr-qc/9712010].
7. F. Dalfovo, S. Giorgini, L. Pitaevskii and S. Stringari, Theory of Bose-Einstein condensation in trapped gases, Rev. Mod. Phys. **71**, 463 (1999) [arXiv:cond-mat/9806038].
8. L. J. Garay, J. R. Anglin, J. I. Cirac and P. Zoller, Black holes in Bose-Einstein condensates, Phys. Rev. Lett. **85**, 4643 (1999) [arXiv:gr-qc/0002015].
9. L. J. Garay, J. R. Anglin, J. I. Cirac and P. Zoller, "Sonic black holes in dilute Bose-Einstein condensates," Phys. Rev. A **63**, 023611 (2001) [arXiv:gr-qc/0005131].
10. C. Barceló, S. Liberati and M. Visser, Analog gravity from Bose-Einstein condensates, Class. Quant. Grav. **18**, 1137 (2001) [arXiv:gr-qc/0011026].
11. T. Jacobson, S. Liberati and D. Mattingly, Lorentz violation at high energy: Concepts, phenomena and astrophysical constraints, Ann. Phys. **321**, 150 (2006) [arXiv:astro-ph/0505267].

Fundamental Theories of Physics

Series Editor: Alwyn van der Merwe, University of Denver, USA

1. M. Sachs: *General Relativity and Matter*. A Spinor Field Theory from Fermis to Light-Years. With a Foreword by C. Kilmister. 1982 ISBN 90-277-1381-2
2. G.H. Duffey: *A Development of Quantum Mechanics*. Based on Symmetry Considerations. 1985 ISBN 90-277-1587-4
3. S. Diner, D. Fargue, G. Lochak and F. Selleri (eds.): *The Wave-Particle Dualism*. A Tribute to Louis de Broglie on his 90th Birthday. 1984 ISBN 90-277-1664-1
4. E. Prugovečki: *Stochastic Quantum Mechanics and Quantum Spacetime.* A Consistent Unification of Relativity and Quantum Theory based on Stochastic Spaces. 1984; 2nd printing 1986 ISBN 90-277-1617-X
5. D. Hestenes and G. Sobczyk: *Clifford Algebra to Geometric Calculus*. A Unified Language for Mathematics and Physics. 1984 ISBN 90-277-1673-0; Pb (1987) 90-277-2561-6
6. P. Exner: *Open Quantum Systems and Feynman Integrals*. 1985 ISBN 90-277-1678-1
7. L. Mayants: *The Enigma of Probability and Physics*. 1984 ISBN 90-277-1674-9
8. E. Tocaci: *Relativistic Mechanics, Time and Inertia.* Translated from Romanian. Edited and with a Foreword by C.W. Kilmister. 1985 ISBN 90-277-1769-9
9. B. Bertotti, F. de Felice and A. Pascolini (eds.): *General Relativity and Gravitation*. Proceedings of the 10th International Conference (Padova, Italy, 1983). 1984 ISBN 90-277-1819-9
10. G. Tarozzi and A. van der Merwe (eds.): *Open Questions in Quantum Physics*. 1985 ISBN 90-277-1853-9
11. J.V. Narlikar and T. Padmanabhan: *Gravity, Gauge Theories and Quantum Cosmology*. 1986 ISBN 90-277-1948-9
12. G.S. Asanov: *Finsler Geometry, Relativity and Gauge Theories*. 1985 ISBN 90-277-1960-8
13. K. Namsrai: *Nonlocal Quantum Field Theory and Stochastic Quantum Mechanics*. 1986 ISBN 90-277-2001-0
14. C. Ray Smith and W.T. Grandy, Jr. (eds.): *Maximum-Entropy and Bayesian Methods in Inverse Problems*. Proceedings of the 1st and 2nd International Workshop (Laramie, Wyoming, USA). 1985 ISBN 90-277-2074-6
15. D. Hestenes: *New Foundations for Classical Mechanics.* 1986 ISBN 90-277-2090-8; Pb (1987) 90-277-2526-8
16. S.J. Prokhovnik: *Light in Einstein's Universe.* The Role of Energy in Cosmology and Relativity. 1985 ISBN 90-277-2093-2
17. Y.S. Kim and M.E. Noz: *Theory and Applications of the Poincaré Group*. 1986 ISBN 90-277-2141-6
18. M. Sachs: *Quantum Mechanics from General Relativity.* An Approximation for a Theory of Inertia. 1986 ISBN 90-277-2247-1
19. W.T. Grandy, Jr.: *Foundations of Statistical Mechanics.* Vol. I: *Equilibrium Theory*. 1987 ISBN 90-277-2489-X
20. H.-H von Borzeszkowski and H.-J. Treder: *The Meaning of Quantum Gravity*. 1988 ISBN 90-277-2518-7
21. C. Ray Smith and G.J. Erickson (eds.): *Maximum-Entropy and Bayesian Spectral Analysis and Estimation Problems.* Proceedings of the 3rd International Workshop (Laramie, Wyoming, USA, 1983). 1987 ISBN 90-277-2579-9
22. A.O. Barut and A. van der Merwe (eds.): *Selected Scientific Papers of Alfred Landé.* [*1888-1975*]. 1988 ISBN 90-277-2594-2

Fundamental Theories of Physics

23. W.T. Grandy, Jr.: *Foundations of Statistical Mechanics.* Vol. II: *Nonequilibrium Phenomena.* 1988 ISBN 90-277-2649-3
24. E.I. Bitsakis and C.A. Nicolaides (eds.): *The Concept of Probability.* Proceedings of the Delphi Conference (Delphi, Greece, 1987). 1989 ISBN 90-277-2679-5
25. A. van der Merwe, F. Selleri and G. Tarozzi (eds.): *Microphysical Reality and Quantum Formalism, Vol. 1.* Proceedings of the International Conference (Urbino, Italy, 1985). 1988 ISBN 90-277-2683-3
26. A. van der Merwe, F. Selleri and G. Tarozzi (eds.): *Microphysical Reality and Quantum Formalism, Vol. 2.* Proceedings of the International Conference (Urbino, Italy, 1985). 1988 ISBN 90-277-2684-1
27. I.D. Novikov and V.P. Frolov: *Physics of Black Holes.* 1989 ISBN 90-277-2685-X
28. G. Tarozzi and A. van der Merwe (eds.): *The Nature of Quantum Paradoxes.* Italian Studies in the Foundations and Philosophy of Modern Physics. 1988 ISBN 90-277-2703-1
29. B.R. Iyer, N. Mukunda and C.V. Vishveshwara (eds.): *Gravitation, Gauge Theories and the Early Universe.* 1989 ISBN 90-277-2710-4
30. H. Mark and L. Wood (eds.): *Energy in Physics, War and Peace.* A Festschrift celebrating Edward Teller's 80th Birthday. 1988 ISBN 90-277-2775-9
31. G.J. Erickson and C.R. Smith (eds.): *Maximum-Entropy and Bayesian Methods in Science and Engineering.* Vol. I: *Foundations*. 1988 ISBN 90-277-2793-7
32. G.J. Erickson and C.R. Smith (eds.): *Maximum-Entropy and Bayesian Methods in Science and Engineering.* Vol. II: *Applications*. 1988 ISBN 90-277-2794-5
33. M.E. Noz and Y.S. Kim (eds.): *Special Relativity and Quantum Theory.* A Collection of Papers on the Poincaré Group. 1988 ISBN 90-277-2799-6
34. I.Yu. Kobzarev and Yu.I. Manin: *Elementary Particles. Mathematics, Physics and Philosophy*. 1989 ISBN 0-7923-0098-X
35. F. Selleri: *Quantum Paradoxes and Physical Reality.* 1990 ISBN 0-7923-0253-2
36. J. Skilling (ed.): *Maximum-Entropy and Bayesian Methods.* Proceedings of the 8th International Workshop (Cambridge, UK, 1988). 1989 ISBN 0-7923-0224-9
37. M. Kafatos (ed.): *Bell's Theorem, Quantum Theory and Conceptions of the Universe*. 1989 ISBN 0-7923-0496-9
38. Yu.A. Izyumov and V.N. Syromyatnikov: *Phase Transitions and Crystal Symmetry*. 1990 ISBN 0-7923-0542-6
39. P.F. Fougère (ed.): *Maximum-Entropy and Bayesian Methods.* Proceedings of the 9th International Workshop (Dartmouth, Massachusetts, USA, 1989). 1990 ISBN 0-7923-0928-6
40. L. de Broglie: *Heisenberg's Uncertainties and the Probabilistic Interpretation of Wave Mechanics.* With Critical Notes of the Author. 1990 ISBN 0-7923-0929-4
41. W.T. Grandy, Jr.: *Relativistic Quantum Mechanics of Leptons and Fields*. 1991 ISBN 0-7923-1049-7
42. Yu.L. Klimontovich: *Turbulent Motion and the Structure of Chaos*. A New Approach to the Statistical Theory of Open Systems. 1991 ISBN 0-7923-1114-0
43. W.T. Grandy, Jr. and L.H. Schick (eds.): *Maximum-Entropy and Bayesian Methods*. Proceedings of the 10th International Workshop (Laramie, Wyoming, USA, 1990). 1991 ISBN 0-7923-1140-X
44. P. Pták and S. Pulmannová: *Orthomodular Structures as Quantum Logics*. Intrinsic Properties, State Space and Probabilistic Topics. 1991 ISBN 0-7923-1207-4
45. D. Hestenes and A. Weingartshofer (eds.): *The Electron.* New Theory and Experiment. 1991 ISBN 0-7923-1356-9

Fundamental Theories of Physics

46. P.P.J.M. Schram: *Kinetic Theory of Gases and Plasmas.* 1991 ISBN 0-7923-1392-5
47. A. Micali, R. Boudet and J. Helmstetter (eds.): *Clifford Algebras and their Applications in Mathematical Physics.* 1992 ISBN 0-7923-1623-1
48. E. Prugovečki: *Quantum Geometry.* A Framework for Quantum General Relativity. 1992 ISBN 0-7923-1640-1
49. M.H. Mac Gregor: *The Enigmatic Electron.* 1992 ISBN 0-7923-1982-6
50. C.R. Smith, G.J. Erickson and P.O. Neudorfer (eds.): *Maximum Entropy and Bayesian Methods.* Proceedings of the 11th International Workshop (Seattle, 1991). 1993 ISBN 0-7923-2031-X
51. D.J. Hoekzema: *The Quantum Labyrinth.* 1993 ISBN 0-7923-2066-2
52. Z. Oziewicz, B. Jancewicz and A. Borowiec (eds.): *Spinors, Twistors, Clifford Algebras and Quantum Deformations.* Proceedings of the Second Max Born Symposium (Wrocław, Poland, 1992). 1993 ISBN 0-7923-2251-7
53. A. Mohammad-Djafari and G. Demoment (eds.): *Maximum Entropy and Bayesian Methods.* Proceedings of the 12th International Workshop (Paris, France, 1992). 1993 ISBN 0-7923-2280-0
54. M. Riesz: *Clifford Numbers and Spinors* with Riesz' Private Lectures to E. Folke Bolinder and a Historical Review by Pertti Lounesto. E.F. Bolinder and P. Lounesto (eds.). 1993 ISBN 0-7923-2299-1
55. F. Brackx, R. Delanghe and H. Serras (eds.): *Clifford Algebras and their Applications in Mathematical Physics.* Proceedings of the Third Conference (Deinze, 1993) 1993 ISBN 0-7923-2347-5
56. J.R. Fanchi: *Parametrized Relativistic Quantum Theory.* 1993 ISBN 0-7923-2376-9
57. A. Peres: *Quantum Theory: Concepts and Methods.* 1993 ISBN 0-7923-2549-4
58. P.L. Antonelli, R.S. Ingarden and M. Matsumoto: *The Theory of Sprays and Finsler Spaces with Applications in Physics and Biology*. 1993 ISBN 0-7923-2577-X
59. R. Miron and M. Anastasiei: *The Geometry of Lagrange Spaces: Theory and Applications*. 1994 ISBN 0-7923-2591-5
60. G. Adomian: *Solving Frontier Problems of Physics: The Decomposition Method.* 1994 ISBN 0-7923-2644-X
61. B.S. Kerner and V.V. Osipov: *Autosolitons.* A New Approach to Problems of Self-Organization and Turbulence. 1994 ISBN 0-7923-2816-7
62. G.R. Heidbreder (ed.): *Maximum Entropy and Bayesian Methods.* Proceedings of the 13th International Workshop (Santa Barbara, USA, 1993) 1996 ISBN 0-7923-2851-5
63. J. Peřina, Z. Hradil and B. Jurčo: *Quantum Optics and Fundamentals of Physics*. 1994 ISBN 0-7923-3000-5
64. M. Evans and J.-P. Vigier: *The Enigmatic Photon.* Volume 1: The Field $\boldsymbol{B}^{(3)}$. 1994 ISBN 0-7923-3049-8
65. C.K. Raju: *Time: Towards a Constistent Theory*. 1994 ISBN 0-7923-3103-6
66. A.K.T. Assis: *Weber's Electrodynamics*. 1994 ISBN 0-7923-3137-0
67. Yu. L. Klimontovich: *Statistical Theory of Open Systems.* Volume 1: A Unified Approach to Kinetic Description of Processes in Active Systems. 1995 ISBN 0-7923-3199-0; Pb: ISBN 0-7923-3242-3
68. M. Evans and J.-P. Vigier: *The Enigmatic Photon.* Volume 2: Non-Abelian Electrodynamics. 1995 ISBN 0-7923-3288-1
69. G. Esposito: *Complex General Relativity.* 1995 ISBN 0-7923-3340-3

Fundamental Theories of Physics

70. J. Skilling and S. Sibisi (eds.): *Maximum Entropy and Bayesian Methods.* Proceedings of the Fourteenth International Workshop on Maximum Entropy and Bayesian Methods. 1996 ISBN 0-7923-3452-3
71. C. Garola and A. Rossi (eds.): *The Foundations of Quantum Mechanics Historical Analysis and Open Questions.* 1995 ISBN 0-7923-3480-9
72. A. Peres: *Quantum Theory: Concepts and Methods.* 1995 (see for hardback edition, Vol. 57) ISBN Pb 0-7923-3632-1
73. M. Ferrero and A. van der Merwe (eds.): *Fundamental Problems in Quantum Physics*. 1995 ISBN 0-7923-3670-4
74. F.E. Schroeck, Jr.: *Quantum Mechanics on Phase Space.* 1996 ISBN 0-7923-3794-8
75. L. de la Peña and A.M. Cetto: *The Quantum Dice.* An Introduction to Stochastic Electrodynamics. 1996 ISBN 0-7923-3818-9
76. P.L. Antonelli and R. Miron (eds.): *Lagrange and Finsler Geometry.* Applications to Physics and Biology. 1996 ISBN 0-7923-3873-1
77. M.W. Evans, J.-P. Vigier, S. Roy and S. Jeffers: *The Enigmatic Photon.* Volume 3: Theory and Practice of the $\boldsymbol{B}^{(3)}$ Field. 1996 ISBN 0-7923-4044-2
78. W.G.V. Rosser: *Interpretation of Classical Electromagnetism.* 1996 ISBN 0-7923-4187-2
79. K.M. Hanson and R.N. Silver (eds.): *Maximum Entropy and Bayesian Methods.* 1996 ISBN 0-7923-4311-5
80. S. Jeffers, S. Roy, J.-P. Vigier and G. Hunter (eds.): *The Present Status of the Quantum Theory of Light.* Proceedings of a Symposium in Honour of Jean-Pierre Vigier. 1997 ISBN 0-7923-4337-9
81. M. Ferrero and A. van der Merwe (eds.): *New Developments on Fundamental Problems in Quantum Physics.* 1997 ISBN 0-7923-4374-3
82. R. Miron: *The Geometry of Higher-Order Lagrange Spaces.* Applications to Mechanics and Physics. 1997 ISBN 0-7923-4393-X
83. T. Hakioğlu and A.S. Shumovsky (eds.): *Quantum Optics and the Spectroscopy of Solids.* Concepts and Advances. 1997 ISBN 0-7923-4414-6
84. A. Sitenko and V. Tartakovskii: *Theory of Nucleus.* Nuclear Structure and Nuclear Interaction. 1997 ISBN 0-7923-4423-5
85. G. Esposito, A.Yu. Kamenshchik and G. Pollifrone: *Euclidean Quantum Gravity on Manifolds with Boundary.* 1997 ISBN 0-7923-4472-3
86. R.S. Ingarden, A. Kossakowski and M. Ohya: *Information Dynamics and Open Systems.* Classical and Quantum Approach. 1997 ISBN 0-7923-4473-1
87. K. Nakamura: *Quantum versus Chaos*. Questions Emerging from Mesoscopic Cosmos. 1997 ISBN 0-7923-4557-6
88. B.R. Iyer and C.V. Vishveshwara (eds.): *Geometry, Fields and Cosmology.* Techniques and Applications. 1997 ISBN 0-7923-4725-0
89. G.A. Martynov: *Classical Statistical Mechanics.* 1997 ISBN 0-7923-4774-9
90. M.W. Evans, J.-P. Vigier, S. Roy and G. Hunter (eds.): *The Enigmatic Photon.* Volume 4: New Directions. 1998 ISBN 0-7923-4826-5
91. M. Rédei: *Quantum Logic in Algebraic Approach.* 1998 ISBN 0-7923-4903-2
92. S. Roy: *Statistical Geometry and Applications to Microphysics and Cosmology*. 1998 ISBN 0-7923-4907-5
93. B.C. Eu: *Nonequilibrium Statistical Mechanics.* Ensembled Method. 1998 ISBN 0-7923-4980-6

Fundamental Theories of Physics

94. V. Dietrich, K. Habetha and G. Jank (eds.): *Clifford Algebras and Their Application in Mathematical Physics.* Aachen 1996. 1998 ISBN 0-7923-5037-5
95. J.P. Blaizot, X. Campi and M. Ploszajczak (eds.): *Nuclear Matter in Different Phases and Transitions.* 1999 ISBN 0-7923-5660-8
96. V.P. Frolov and I.D. Novikov: *Black Hole Physics*. Basic Concepts and New Developments. 1998 ISBN 0-7923-5145-2; Pb 0-7923-5146
97. G. Hunter, S. Jeffers and J-P. Vigier (eds.): *Causality and Locality in Modern Physics.* 1998 ISBN 0-7923-5227-0
98. G.J. Erickson, J.T. Rychert and C.R. Smith (eds.): *Maximum Entropy and Bayesian Methods.* 1998 ISBN 0-7923-5047-2
99. D. Hestenes: *New Foundations for Classical Mechanics (Second Edition)*. 1999 ISBN 0-7923-5302-1; Pb ISBN 0-7923-5514-8
100. B.R. Iyer and B. Bhawal (eds.): *Black Holes, Gravitational Radiation and the Universe*. Essays in Honor of C. V. Vishveshwara. 1999 ISBN 0-7923-5308-0
101. P.L. Antonelli and T.J. Zastawniak: *Fundamentals of Finslerian Diffusion with Applications.* 1998 ISBN 0-7923-5511-3
102. H. Atmanspacher, A. Amann and U. Müller-Herold: *On Quanta, Mind and Matter Hans Primas in Context.* 1999 ISBN 0-7923-5696-9
103. M.A. Trump and W.C. Schieve: *Classical Relativistic Many-Body Dynamics.* 1999 ISBN 0-7923-5737-X
104. A.I. Maimistov and A.M. Basharov: *Nonlinear Optical Waves.* 1999 ISBN 0-7923-5752-3
105. W. von der Linden, V. Dose, R. Fischer and R. Preuss (eds.): *Maximum Entropy and Bayesian Methods Garching, Germany 1998.* 1999 ISBN 0-7923-5766-3
106. M.W. Evans: *The Enigmatic Photon Volume 5: O(3) Electrodynamics.* 1999 ISBN 0-7923-5792-2
107. G.N. Afanasiev: *Topological Effects in Quantum Mecvhanics.* 1999 ISBN 0-7923-5800-7
108. V. Devanathan: *Angular Momentum Techniques in Quantum Mechanics.* 1999 ISBN 0-7923-5866-X
109. P.L. Antonelli (ed.): *Finslerian Geometries A Meeting of Minds.* 1999 ISBN 0-7923-6115-6
110. M.B. Mensky: *Quantum Measurements and Decoherence Models and Phenomenology.* 2000 ISBN 0-7923-6227-6
111. B. Coecke, D. Moore and A. Wilce (eds.): *Current Research in Operation Quantum Logic*. Algebras, Categories, Languages. 2000 ISBN 0-7923-6258-6
112. G. Jumarie: *Maximum Entropy, Information Without Probability and Complex Fractals.* Classical and Quantum Approach. 2000 ISBN 0-7923-6330-2
113. B. Fain: *Irreversibilities in Quantum Mechanics*. 2000 ISBN 0-7923-6581-X
114. T. Borne, G. Lochak and H. Stumpf: *Nonperturbative Quantum Field Theory and the Structure of Matter*. 2001 ISBN 0-7923-6803-7
115. J. Keller: *Theory of the Electron*. A Theory of Matter from START. 2001 ISBN 0-7923-6819-3
116. M. Rivas: *Kinematical Theory of Spinning Particles*. Classical and Quantum Mechanical Formalism of Elementary Particles. 2001 ISBN 0-7923-6824-X
117. A.A. Ungar: *Beyond the Einstein Addition Law and its Gyroscopic Thomas Precession*. The Theory of Gyrogroups and Gyrovector Spaces. 2001 ISBN 0-7923-6909-2
118. R. Miron, D. Hrimiuc, H. Shimada and S.V. Sabau: *The Geometry of Hamilton and Lagrange Spaces*. 2001 ISBN 0-7923-6926-2

Fundamental Theories of Physics

119. M. Pavšič: *The Landscape of Theoretical Physics: A Global View*. From Point Particles to the Brane World and Beyond in Search of a Unifying Principle. 2001 ISBN 0-7923-7006-6
120. R.M. Santilli: *Foundations of Hadronic Chemistry*. With Applications to New Clean Energies and Fuels. 2001 ISBN 1-4020-0087-1
121. S. Fujita and S. Godoy: *Theory of High Temperature Superconductivity*. 2001 ISBN 1-4020-0149-5
122. R. Luzzi, A.R. Vasconcellos and J. Galvão Ramos: *Predictive Statitical Mechanics*. A Nonequilibrium Ensemble Formalism. 2002 ISBN 1-4020-0482-6
123. V.V. Kulish: *Hierarchical Methods*. Hierarchy and Hierarchical Asymptotic Methods in Electrodynamics, Volume 1. 2002 ISBN 1-4020-0757-4; Set: 1-4020-0758-2
124. B.C. Eu: *Generalized Thermodynamics*. Thermodynamics of Irreversible Processes and Generalized Hydrodynamics. 2002 ISBN 1-4020-0788-4
125. A. Mourachkine: *High-Temperature Superconductivity in Cuprates*. The Nonlinear Mechanism and Tunneling Measurements. 2002 ISBN 1-4020-0810-4
126. R.L. Amoroso, G. Hunter, M. Kafatos and J.-P. Vigier (eds.): *Gravitation and Cosmology: From the Hubble Radius to the Planck Scale*. Proceedings of a Symposium in Honour of the 80th Birthday of Jean-Pierre Vigier. 2002 ISBN 1-4020-0885-6
127. W.M. de Muynck: *Foundations of Quantum Mechanics, an Empiricist Approach*. 2002 ISBN 1-4020-0932-1
128. V.V. Kulish: *Hierarchical Methods*. Undulative Electrodynamical Systems, Volume 2. 2002 ISBN 1-4020-0968-2; Set: 1-4020-0758-2
129. M. Mugur-Schächter and A. van der Merwe (eds.): *Quantum Mechanics, Mathematics, Cognition and Action*. Proposals for a Formalized Epistemology. 2002 ISBN 1-4020-1120-2
130. P. Bandyopadhyay: *Geometry, Topology and Quantum Field Theory*. 2003 ISBN 1-4020-1414-7
131. V. Garzó and A. Santos: *Kinetic Theory of Gases in Shear Flows*. Nonlinear Transport. 2003 ISBN 1-4020-1436-8
132. R. Miron: *The Geometry of Higher-Order Hamilton Spaces*. Applications to Hamiltonian Mechanics. 2003 ISBN 1-4020-1574-7
133. S. Esposito, E. Majorana Jr., A. van der Merwe and E. Recami (eds.): *Ettore Majorana: Notes on Theoretical Physics*. 2003 ISBN 1-4020-1649-2
134. J. Hamhalter. *Quantum Measure Theory*. 2003 ISBN 1-4020-1714-6
135. G. Rizzi and M.L. Ruggiero: *Relativity in Rotating Frames*. Relativistic Physics in Rotating Reference Frames. 2004 ISBN 1-4020-1805-3
136. L. Kantorovich: *Quantum Theory of the Solid State: an Introduction*. 2004 ISBN 1-4020-1821-5
137. A. Ghatak and S. Lokanathan: *Quantum Mechanics: Theory and Applications*. 2004 ISBN 1-4020-1850-9
138. A. Khrennikov: *Information Dynamics in Cognitive, Psychological, Social, and Anomalous Phenomena*. 2004 ISBN 1-4020-1868-1
139. V. Faraoni: *Cosmology in Scalar-Tensor Gravity*. 2004 ISBN 1-4020-1988-2
140. P.P. Teodorescu and N.-A. P. Nicorovici: *Applications of the Theory of Groups in Mechanics and Physics*. 2004 ISBN 1-4020-2046-5
141. G. Munteanu: *Complex Spaces in Finsler, Lagrange and Hamilton Geometries*. 2004 ISBN 1-4020-2205-0

Fundamental Theories of Physics

142. G.N. Afanasiev: *Vavilov-Cherenkov and Synchrotron Radiation*. Foundations and Applications. 2004 ISBN 1-4020-2410-X
143. L. Munteanu and S. Donescu: *Introduction to Soliton Theory: Applications to Mechanics*. 2004 ISBN 1-4020-2576-9
144. M.Yu. Khlopov and S.G. Rubin: *Cosmological Pattern of Microphysics in the Inflationary Universe*. 2004 ISBN 1-4020-2649-8
145. J. Vanderlinde: *Classical Electromagnetic Theory*. 2004 ISBN 1-4020-2699-4
146. V. Čápek and D.P. Sheehan: *Challenges to the Second Law of Thermodynamics.* Theory and Experiment. 2005 ISBN 1-4020-3015-0
147. B.G. Sidharth: *The Universe of Fluctuations*. The Architecture of Spacetime and the Universe. 2005 ISBN 1-4020-3785-6
148. R.W. Carroll: *Fluctuations, Information, Gravity and the Quantum Potential*. 2005 ISBN 1-4020-4003-2
149. B.G. Sidharth: *A Century of Ideas.* Personal Perspectives from a Selection of the Greatest Minds of the Twentieth Century. 2007 ISBN 1-4020-4359-7
150. S.-H. Dong: *Factorization Method in Quantum Mechanics.* 2007 ISBN 1-4020-5795-4
151. R.M. Santilli: *Isodual Theory of Antimatter with applications to Antigravity, Grand Unification and Cosmology.* 2006 ISBN 1-4020-4517-4
152. A. Plotnitsky: *Reading Bohr: Physics and Philosophy.* 2006 ISBN 1-4020-5253-7
153. V. Petkov: *Relativity and the Dimensionality of the World.* 2007 ISBN 978-1-4020-6317-6
154. H.O. Cordes: *Precisely Predictable Dirac Observables.* 2006 ISBN 1-4020-5168-9
155. C.F. von Weizsäcker: *The Structure of Physics.* Edited, revised and enlarged by Thomas Görnitz and Holger Lyre. 2006 ISBN 1-4020-5234-0
156. S.V. Adamenko, F. Selleri and A. van der Merwe (eds.): *Controlled Nucleosynthesis*. Breakthroughs in Experiment and Theory. 2007 ISBN 978-1-4020-5873-8